Mathematical Statistics: Exercises and Solutions

Jun Shao

Springer

Mathematical Statistics: Exercises and Solutions

Jun Shao

Mathematical Statistics: Exercises and Solutions

Springer

Jun Shao
Department of Statistics
University of Wisconsin
Madison, WI 52706
USA
shao@stat.wisc.edu

Library of Congress Control Number: 2005923578

ISBN-10: 0-387-24970-2 Printed on acid-free paper.
ISBN-13: 978-0387-24970-4

Printed in the United States of America. (EB)

9 8 7 6 5 4 3 2 1

springeronline.com

To My Parents

Preface

Since the publication of my book *Mathematical Statistics* (Shao, 2003), I have been asked many times for a solution manual to the exercises in my book. Without doubt, exercises form an important part of a textbook on mathematical statistics, not only in training students for their research ability in mathematical statistics but also in presenting many additional results as complementary material to the main text. Written solutions to these exercises are important for students who initially do not have the skills in solving these exercises completely and are very helpful for instructors of a mathematical statistics course (whether or not my book *Mathematical Statistics* is used as the textbook) in providing answers to students as well as finding additional examples to the main text. Motivated by this and encouraged by some of my colleagues and Springer-Verlag editor John Kimmel, I have completed this book, *Mathematical Statistics: Exercises and Solutions*.

This book consists of solutions to 400 exercises, over 95% of which are in my book *Mathematical Statistics*. Many of them are standard exercises that also appear in other textbooks listed in the references. It is only a partial solution manual to *Mathematical Statistics* (which contains over 900 exercises). However, the types of exercise in *Mathematical Statistics* not selected in the current book are (1) exercises that are routine (each exercise selected in this book has a certain degree of difficulty), (2) exercises similar to one or several exercises selected in the current book, and (3) exercises for advanced materials that are often not included in a mathematical statistics course for first-year Ph.D. students in statistics (e.g., Edgeworth expansions and second-order accuracy of confidence sets, empirical likelihoods, statistical functionals, generalized linear models, nonparametric tests, and theory for the bootstrap and jackknife, etc.). On the other hand, this is a stand-alone book, since exercises and solutions are comprehensible independently of their source for likely readers. To help readers not using this book together with *Mathematical Statistics*, lists of notation, terminology, and some probability distributions are given in the front of the book.

All notational conventions are the same as or very similar to those in *Mathematical Statistics* and so is the mathematical level of this book. Readers are assumed to have a good knowledge in advanced calculus. A course in real analysis or measure theory is highly recommended. If this book is used with a statistics textbook that does not include probability theory, then knowledge in measure-theoretic probability theory is required.

The exercises are grouped into seven chapters with titles matching those in *Mathematical Statistics*. A few errors in the exercises from *Mathematical Statistics* were detected during the preparation of their solutions and the corrected versions are given in this book. Although exercises are numbered independently of their source, the corresponding number in *Mathematical Statistics* is accompanied with each exercise number for convenience of instructors and readers who also use *Mathematical Statistics* as the main text. For example, Exercise 8 (#2.19) means that Exercise 8 in the current book is also Exercise 19 in Chapter 2 of *Mathematical Statistics*.

A note to students/readers who have a need for exercises accompanied by solutions is that they should not be completely driven by the solutions. Students/readers are encouraged to try each exercise first without reading its solution. If an exercise is solved with the help of a solution, they are encouraged to provide solutions to similar exercises as well as to think about whether there is an alternative solution to the one given in this book. A few exercises in this book are accompanied by two solutions and/or notes of brief discussions.

I would like to thank my teaching assistants, Dr. Hansheng Wang, Dr. Bin Cheng, and Mr. Fang Fang, who provided valuable help in preparing some solutions. Any errors are my own responsibility, and a correction of them can be found on my web page http://www.stat.wisc.edu/~ shao.

Madison, Wisconsin Jun Shao
April 2005

Contents

Notation

$\mathcal{R}$: The real line.

$\mathcal{R}^k$: The k-dimensional Euclidean space.

$c = (c_1, ..., c_k)$: A vector (element) in $\mathcal{R}^k$ with jth component $c_j \in \mathcal{R}$; c is considered as a $k \times 1$ matrix (column vector) when matrix algebra is involved.

c^τ: The transpose of a vector $c \in \mathcal{R}^k$ considered as a $1 \times k$ matrix (row vector) when matrix algebra is involved.

$\|c\|$: The Euclidean norm of a vector $c \in \mathcal{R}^k$, $\|c\|^2 = c^\tau c$.

$|c|$: The absolute value of $c \in \mathcal{R}$.

A^τ: The transpose of a matrix A.

$\mathrm{Det}(A)$ or $|A|$: The determinant of a matrix A.

$\mathrm{tr}(A)$: The trace of a matrix A.

$\|A\|$: The norm of a matrix A defined as $\|A\|^2 = \mathrm{tr}(A^\tau A)$.

A^{-1}: The inverse of a matrix A.

A^-: The generalized inverse of a matrix A.

$A^{1/2}$: The square root of a nonnegative definite matrix A defined by $A^{1/2}A^{1/2} = A$.

$A^{-1/2}$: The inverse of $A^{1/2}$.

$\mathcal{R}(A)$: The linear space generated by rows of a matrix A.

I_k: The $k \times k$ identity matrix.

J_k: The k-dimensional vector of 1's.

$\emptyset$: The empty set.

(a, b): The open interval from a to b.

$[a, b]$: The closed interval from a to b.

$(a, b]$: The interval from a to b including b but not a.

$[a, b)$: The interval from a to b including a but not b.

$\{a, b, c\}$: The set consisting of the elements a, b, and c.

$A_1 \times \cdots \times A_k$: The Cartesian product of sets $A_1, ..., A_k$, $A_1 \times \cdots \times A_k = \{(a_1, ..., a_k) : a_1 \in A_1, ..., a_k \in A_k\}$.

$\sigma(\mathcal{C})$: The smallest σ-field that contains $\mathcal{C}$.

$\sigma(X)$: The smallest σ-field with respect to which X is measurable.

$\nu_1 \times \cdots \times \nu_k$: The product measure of ν_1,...,ν_k on $\sigma(\mathcal{F}_1 \times \cdots \times \mathcal{F}_k)$, where ν_i is a measure on $\mathcal{F}_i$, $i = 1, ..., k$.

$\mathcal{B}$: The Borel σ-field on $\mathcal{R}$.

$\mathcal{B}^k$: The Borel σ-field on $\mathcal{R}^k$.

A^c: The complement of a set A.

$A \cup B$: The union of sets A and B.

$\cup A_i$: The union of sets $A_1, A_2, ...$.

$A \cap B$: The intersection of sets A and B.

$\cap A_i$: The intersection of sets $A_1, A_2, ...$.

I_A: The indicator function of a set A.

$P(A)$: The probability of a set A.

$\int f d\nu$: The integral of a Borel function f with respect to a measure ν.

$\int_A f d\nu$: The integral of f on the set A.

$\int f(x) dF(x)$: The integral of f with respect to the probability measure corresponding to the cumulative distribution function F.

$\lambda \ll \nu$: The measure λ is dominated by the measure ν, i.e., $\nu(A) = 0$ always implies $\lambda(A) = 0$.

$\frac{d\lambda}{d\nu}$: The Radon-Nikodym derivative of λ with respect to ν.

$\mathcal{P}$: A collection of populations (distributions).

a.e.: Almost everywhere.

a.s.: Almost surely.

a.s. $\mathcal{P}$: A statement holds except on the event A with $P(A) = 0$ for all $P \in \mathcal{P}$.

δ_x: The point mass at $x \in \mathcal{R}^k$ or the distribution degenerated at $x \in \mathcal{R}^k$.

$\{a_n\}$: A sequence of elements $a_1, a_2, ...$.

$a_n \to a$ or $\lim_n a_n = a$: $\{a_n\}$ converges to a as n increases to ∞.

$\limsup_n a_n$: The largest limit point of $\{a_n\}$, $\limsup_n a_n = \inf_n \sup_{k \geq n} a_k$.

$\liminf_n a_n$: The smallest limit point of $\{a_n\}$, $\liminf_n a_n = \sup_n \inf_{k \geq n} a_k$.

$\to_p$: Convergence in probability.

$\to_d$: Convergence in distribution.

g': The derivative of a function g on $\mathcal{R}$.

g'': The second-order derivative of a function g on $\mathcal{R}$.

$g^{(k)}$: The kth-order derivative of a function g on $\mathcal{R}$.

$g(x+)$: The right limit of a function g at $x \in \mathcal{R}$.

$g(x-)$: The left limit of a function g at $x \in \mathcal{R}$.

$g_+(x)$: The positive part of a function g, $g_+(x) = \max\{g(x), 0\}$.

$g_-(x)$: The negative part of a function g, $g_-(x) = \max\{-g(x), 0\}$.

$\partial g/\partial x$: The partial derivative of a function g on $\mathcal{R}^k$.

$\partial^2 g/\partial x \partial x^\tau$: The second-order partial derivative of a function g on $\mathcal{R}^k$.

$\exp\{x\}$: The exponential function e^x.

$\log x$ or $\log(x)$: The inverse of e^x, $\log(e^x) = x$.

$\Gamma(t)$: The gamma function defined as $\Gamma(t) = \int_0^\infty x^{t-1}e^{-x}dx$, $t > 0$.

$F^{-1}(p)$: The pth quantile of a cumulative distribution function F on $\mathcal{R}$, $F^{-1}(t) = \inf\{x : F(x) \geq t\}$.

$E(X)$ or EX: The expectation of a random variable (vector or matrix) X.

$\mathrm{Var}(X)$: The variance of a random variable X or the covariance matrix of a random vector X.

$\mathrm{Cov}(X, Y)$: The covariance between random variables X and Y.

$E(X|\mathcal{A})$: The conditional expectation of X given a σ-field $\mathcal{A}$.

$E(X|Y)$: The conditional expectation of X given Y.

$P(A|\mathcal{A})$: The conditional probability of A given a σ-field $\mathcal{A}$.

$P(A|Y)$: The conditional probability of A given Y.

$X_{(i)}$: The ith order statistic of $X_1, ..., X_n$.

$\bar{X}$ or $\bar{X}_\cdot$: The sample mean of $X_1, ..., X_n$, $\bar{X} = n^{-1}\sum_{i=1}^n X_i$.

$\bar{X}_{\cdot j}$: The average of X_{ij}'s over the index i, $\bar{X}_{\cdot j} = n^{-1}\sum_{i=1}^n X_{ij}$.

S^2: The sample variance of $X_1, ..., X_n$, $S^2 = (n-1)^{-1}\sum_{i=1}^n (X_i - \bar{X})^2$.

F_n: The empirical distribution of $X_1, ..., X_n$, $F_n(t) = n^{-1}\sum_{i=1}^n \delta_{X_i}(t)$.

$\ell(\theta)$: The likelihood function.

H_0: The null hypothesis in a testing problem.

H_1: The alternative hypothesis in a testing problem.

$L(P, a)$ or $L(\theta, a)$: The loss function in a decision problem.

$R_T(P)$ or $R_T(\theta)$: The risk function of a decision rule T.

$r_{_T}$: The Bayes risk of a decision rule T.

$N(\mu, \sigma^2)$: The one-dimensional normal distribution with mean μ and variance σ^2.

$N_k(\mu, \Sigma)$: The k-dimensional normal distribution with mean vector μ and covariance matrix Σ.

$\Phi(x)$: The cumulative distribution function of $N(0, 1)$.

z_α: The $(1-\alpha)$th quantile of $N(0, 1)$.

χ_r^2: The chi-square distribution with degrees of freedom r.

$\chi_{r,\alpha}^2$: The $(1-\alpha)$th quantile of the chi-square distribution χ_r^2.

$\chi_r^2(\delta)$: The noncentral chi-square distribution with degrees of freedom r and noncentrality parameter δ.

t_r: The t-distribution with degrees of freedom r.

$t_{r,\alpha}$: The $(1-\alpha)$th quantile of the t-distribution t_r.

$t_r(\delta)$: The noncentral t-distribution with degrees of freedom r and noncentrality parameter δ.

$F_{a,b}$: The F-distribution with degrees of freedom a and b.

$F_{a,b,\alpha}$: The $(1-\alpha)$th quantile of the F-distribution $F_{a,b}$.

$F_{a,b}(\delta)$: The noncentral F-distribution with degrees of freedom a and b and noncentrality parameter δ.

$\blacksquare$: The end of a solution.

Terminology

σ-field: A collection $\mathcal{F}$ of subsets of a set Ω is a σ-field on Ω if (i) the empty set $\emptyset \in \mathcal{F}$; (ii) if $A \in \mathcal{F}$, then the complement $A^c \in \mathcal{F}$; and (iii) if $A_i \in \mathcal{F}$, $i = 1, 2, ...$, then their union $\cup A_i \in \mathcal{F}$.

σ-finite measure: A measure ν on a σ-field $\mathcal{F}$ on Ω is σ-finite if there are $A_1, A_2, ...$ in $\mathcal{F}$ such that $\cup A_i = \Omega$ and $\nu(A_i) < \infty$ for all i.

Action or decision: Let X be a sample from a population P. An action or decision is a conclusion we make about P based on the observed X.

Action space: The set of all possible actions.

Admissibility: A decision rule T is admissible under the loss function $L(P, \cdot)$, where P is the unknown population, if there is no other decision rule T_1 that is better than T in the sense that $E[L(P, T_1)] \leq E[L(P, T)]$ for all P and $E[L(P, T_1)] < E[L(P, T)]$ for some P.

Ancillary statistic: A statistic is ancillary if and only if its distribution does not depend on any unknown quantity.

Asymptotic bias: Let T_n be an estimator of θ for every n satisfying $a_n(T_n - \theta) \to_d Y$ with $E|Y| < \infty$, where $\{a_n\}$ is a sequence of positive numbers satisfying $\lim_n a_n = \infty$ or $\lim_n a_n = a > 0$. An asymptotic bias of T_n is defined to be EY/a_n.

Asymptotic level α test: Let X be a sample of size n from P and $T(X)$ be a test for $H_0 : P \in \mathcal{P}_0$ versus $H_1 : P \in \mathcal{P}_1$. If $\lim_n E[T(X)] \leq \alpha$ for any $P \in \mathcal{P}_0$, then $T(X)$ has asymptotic level α.

Asymptotic mean squared error and variance: Let T_n be an estimator of θ for every n satisfying $a_n(T_n - \theta) \to_d Y$ with $0 < EY^2 < \infty$, where $\{a_n\}$ is a sequence of positive numbers satisfying $\lim_n a_n = \infty$. The asymptotic mean squared error of T_n is defined to be EY^2/a_n^2 and the asymptotic variance of T_n is defined to be $\mathrm{Var}(Y)/a_n^2$.

Asymptotic relative efficiency: Let T_n and T_n' be estimators of θ. The asymptotic relative efficiency of T_n' with respect to T_n is defined to be the asymptotic mean squared error of T_n divided by the asymptotic mean squared error of T_n'.

Asymptotically correct confidence set: Let X be a sample of size n from P and $C(X)$ be a confidence set for θ. If $\lim_n P(\theta \in C(X)) = 1 - \alpha$, then $C(X)$ is $1 - \alpha$ asymptotically correct.

Bayes action: Let X be a sample from a population indexed by $\theta \in \Theta \subset \mathcal{R}^k$. A Bayes action in a decision problem with action space A and loss function $L(\theta, a)$ is the action that minimizes the posterior expected loss $E[L(\theta, a)]$ over $a \in A$, where E is the expectation with respect to the posterior distribution of θ given X.

Bayes risk: Let X be a sample from a population indexed by $\theta \in \Theta \subset \mathcal{R}^k$. The Bayes risk of a decision rule T is the expected risk of T with respect to a prior distribution on Θ.

Bayes rule or Bayes estimator: A Bayes rule has the smallest Bayes risk over all decision rules. A Bayes estimator is a Bayes rule in an estimation problem.

Borel σ-field $\mathcal{B}^k$: The smallest σ-field containing all open subsets of $\mathcal{R}^k$.

Borel function: A function f from Ω to $\mathcal{R}^k$ is Borel with respect to a σ-field $\mathcal{F}$ on Ω if and only if $f^{-1}(B) \in \mathcal{F}$ for any $B \in \mathcal{B}^k$.

Characteristic function: The characteristic function of a distribution F on $\mathcal{R}^k$ is $\int e^{\sqrt{-1}t^\tau x} dF(x)$, $t \in \mathcal{R}^k$.

Complete (or bounded complete) statistic: Let X be a sample from a population P. A statistic $T(X)$ is complete (or bounded complete) for P if and only if, for any Borel (or bounded Borel) f, $E[f(T)] = 0$ for all P implies $f = 0$ except for a set A with $P(X \in A) = 0$ for all P.

Conditional expectation $E(X|\mathcal{A})$: Let X be an integrable random variable on a probability space $(\Omega, \mathcal{F}, P)$ and $\mathcal{A}$ be a σ-field contained in $\mathcal{F}$. The conditional expectation of X given $\mathcal{A}$, denoted by $E(X|\mathcal{A})$, is defined to be the a.s.-unique random variable satisfying (a) $E(X|\mathcal{A})$ is Borel with respect to $\mathcal{A}$ and (b) $\int_A E(X|\mathcal{A}) dP = \int_A X dP$ for any $A \in \mathcal{A}$.

Conditional expectation $E(X|Y)$: The conditional expectation of X given Y, denoted by $E(X|Y)$, is defined as $E(X|Y) = E(X|\sigma(Y))$.

Confidence coefficient and confidence set: Let X be a sample from a population P and $\theta \in \mathcal{R}^k$ be an unknown parameter that is a function of P. A confidence set $C(X)$ for θ is a Borel set on $\mathcal{R}^k$ depending on X. The confidence coefficient of a confidence set $C(X)$ is $\inf_P P(\theta \in C(X))$. A confidence set is said to be a $1 - \alpha$ confidence set for θ if its confidence coefficient is $1 - \alpha$.

Confidence interval: A confidence interval is a confidence set that is an interval.

Consistent estimator: Let X be a sample of size n from P. An estimator $T(X)$ of θ is consistent if and only if $T(X) \to_p \theta$ for any P as $n \to \infty$. $T(X)$ is strongly consistent if and only if $\lim_n T(X) = \theta$ a.s. for any P. $T(X)$ is consistent in mean squared error if and only if $\lim_n E[T(X) - \theta]^2 = 0$ for any P.

Consistent test: Let X be a sample of size n from P. A test $T(X)$ for testing $H_0 : P \in \mathcal{P}_0$ versus $H_1 : P \in \mathcal{P}_1$ is consistent if and only if $\lim_n E[T(X)] = 1$ for any $P \in \mathcal{P}_1$.

Decision rule (nonrandomized): Let X be a sample from a population P. A (nonrandomized) decision rule is a measurable function from the range of X to the action space.

Discrete probability density: A probability density with respect to the counting measure on the set of nonnegative integers.

Distribution and cumulative distribution function: The probability measure corresponding to a random vector is called its distribution (or law). The cumulative distribution function of a distribution or probability measure P on $\mathcal{B}^k$ is $F(x_1, ..., x_k) = P((-\infty, x_1] \times \cdots \times (-\infty, x_k])$, $x_i \in \mathcal{R}$.

Empirical Bayes rule: An empirical Bayes rule is a Bayes rule with parameters in the prior estimated using data.

Empirical distribution: The empirical distribution based on a random sample $(X_1, ..., X_n)$ is the distribution putting mass n^{-1} at each X_i, $i = 1, ..., n$.

Estimability: A parameter θ is estimable if and only if there exists an unbiased estimator of θ.

Estimator: Let X be a sample from a population P and $\theta \in \mathcal{R}^k$ be a function of P. An estimator of θ is a measurable function of X.

Exponential family: A family of probability densities $\{f_\theta : \theta \in \Theta\}$ (with respect to a common σ-finite measure ν), $\Theta \subset \mathcal{R}^k$, is an exponential family if and only if $f_\theta(x) = \exp\big\{[\eta(\theta)]^\tau T(x) - \xi(\theta)\big\}h(x)$, where T is a random p-vector with a fixed positive integer p, η is a function from Θ to $\mathcal{R}^p$, h is a nonnegative Borel function, and $\xi(\theta) = \log\big\{\int \exp\{[\eta(\theta)]^\tau T(x)\}h(x)d\nu\big\}$.

Generalized Bayes rule: A generalized Bayes rule is a Bayes rule when the prior distribution is improper.

Improper or proper prior: A prior is improper if it is a measure but not a probability measure. A prior is proper if it is a probability measure.

Independence: Let $(\Omega, \mathcal{F}, P)$ be a probability space. Events in $\mathcal{C} \subset \mathcal{F}$ are independent if and only if for any positive integer n and distinct events A_1,...,A_n in $\mathcal{C}$, $P(A_1 \cap A_2 \cap \cdots \cap A_n) = P(A_1)P(A_2)\cdots P(A_n)$. Collections $\mathcal{C}_i \subset \mathcal{F}$, $i \in \mathcal{I}$ (an index set that can be uncountable), are independent if and only if events in any collection of the form $\{A_i \in \mathcal{C}_i : i \in \mathcal{I}\}$ are independent. Random elements X_i, $i \in \mathcal{I}$, are independent if and only if $\sigma(X_i)$, $i \in \mathcal{I}$, are independent.

Integration or integral: Let ν be a measure on a σ-field $\mathcal{F}$ on a set Ω. The integral of a nonnegative simple function (i.e., a function of the form $\varphi(\omega) = \sum_{i=1}^k a_i I_{A_i}(\omega)$, where $\omega \in \Omega$, k is a positive integer, $A_1, ..., A_k$ are in $\mathcal{F}$, and $a_1, ..., a_k$ are nonnegative numbers) is defined as $\int \varphi d\nu = \sum_{i=1}^k a_i \nu(A_i)$. The integral of a nonnegative Borel function is defined as $\int f d\nu = \sup_{\varphi \in S_f} \int \varphi d\nu$, where S_f is the collection of all nonnegative simple functions that are bounded by f. For a Borel function f, its integral exists if and only if at least one of $\int \max\{f, 0\} d\nu$ and $\int \max\{-f, 0\} d\nu$ is finite, in which case $\int f d\nu = \int \max\{f, 0\} d\nu - \int \max\{-f, 0\} d\nu$. f is integrable if and only if both $\int \max\{f, 0\} d\nu$ and $\int \max\{-f, 0\} d\nu$ are finite. When ν is a probability measure corresponding to the cumulative distribution function F on $\mathcal{R}^k$, we write $\int f d\nu = \int f(x) dF(x)$. For any event A, $\int_A f d\nu$ is defined as $\int I_A f d\nu$.

Invariant decision rule: Let X be a sample from $P \in \mathcal{P}$ and $\mathcal{G}$ be a group of one-to-one transformations of X ($g_i \in \mathcal{G}$ implies $g_1 \circ g_2 \in \mathcal{G}$ and $g_i^{-1} \in \mathcal{G}$). $\mathcal{P}$ is invariant under $\mathcal{G}$ if and only if $\bar{g}(P_X) = P_{g(X)}$ is a one-to-one transformation from $\mathcal{P}$ onto $\mathcal{P}$ for each $g \in \mathcal{G}$. A decision problem is invariant if and only if $\mathcal{P}$ is invariant under $\mathcal{G}$ and the loss $L(P, a)$ is invariant in the sense that, for every $g \in \mathcal{G}$ and every $a \in A$ (the collection of all possible actions), there exists a unique $\bar{g}(a) \in A$ such that $L(P_X, a) = L\left(P_{g(X)}, \bar{g}(a)\right)$. A decision rule $T(x)$ in an invariant decision problem is invariant if and only if, for every $g \in \mathcal{G}$ and every x in the range of X, $T(g(x)) = \bar{g}(T(x))$.

Invariant estimator: An invariant estimator is an invariant decision rule in an estimation problem.

LR (Likelihood ratio) test: Let $\ell(\theta)$ be the likelihood function based on a sample X whose distribution is P_θ, $\theta \in \Theta \subset \mathcal{R}^p$ for some positive integer p. For testing $H_0 : \theta \in \Theta_0 \subset \Theta$ versus $H_1 : \theta \notin \Theta_0$, an LR test is any test that rejects H_0 if and only if $\lambda(X) < c$, where $c \in [0, 1]$ and $\lambda(X) = \sup_{\theta \in \Theta_0} \ell(\theta) / \sup_{\theta \in \Theta} \ell(\theta)$ is the likelihood ratio.

LSE: The least squares estimator.

Level α test: A test is of level α if its size is at most α.

Level $1 - \alpha$ confidence set or interval: A confidence set or interval is said to be of level $1 - \alpha$ if its confidence coefficient is at least $1 - \alpha$.

Likelihood function and likelihood equation: Let X be a sample from a population P indexed by an unknown parameter vector $\theta \in \mathcal{R}^k$. The joint probability density of X treated as a function of θ is called the likelihood function and denoted by $\ell(\theta)$. The likelihood equation is $\partial \log \ell(\theta) / \partial \theta = 0$.

Location family: A family of Lebesgue densities on $\mathcal{R}$, $\{f_\mu : \mu \in \mathcal{R}\}$, is a location family with location parameter μ if and only if $f_\mu(x) = f(x-\mu)$, where f is a known Lebesgue density.

Location invariant estimator. Let $(X_1, ..., X_n)$ be a random sample from a population in a location family. An estimator $T(X_1, ..., X_n)$ of the location parameter is location invariant if and only if $T(X_1+c, ..., X_n+c) = T(X_1, ..., X_n) + c$ for any X_i's and $c \in \mathcal{R}$.

Location-scale family: A family of Lebesgue densities on $\mathcal{R}$, $\{f_{\mu,\sigma} : \mu \in \mathcal{R}, \sigma > 0\}$, is a location-scale family with location parameter μ and scale parameter σ if and only if $f_{\mu,\sigma}(x) = \frac{1}{\sigma} f\left(\frac{x-\mu}{\sigma}\right)$, where f is a known Lebesgue density.

Location-scale invariant estimator. Let $(X_1, ..., X_n)$ be a random sample from a population in a location-scale family with location parameter μ and scale parameter σ. An estimator $T(X_1, ..., X_n)$ of the location parameter μ is location-scale invariant if and only if $T(rX_1+c, ..., rX_n+c) = rT(X_1, ..., X_n) + c$ for any X_i's, $c \in \mathcal{R}$, and $r > 0$. An estimator $S(X_1, ..., X_n)$ of σ^h with a fixed $h \neq 0$ is location-scale invariant if and only if $S(rX_1+c, ..., rX_n+c) = r^h T(X_1, ..., X_n)$ for any X_i's and $r > 0$.

Loss function: Let X be a sample from a population $P \in \mathcal{P}$ and A be the set of all possible actions we may take after we observe X. A loss function $L(P, a)$ is a nonnegative Borel function on $\mathcal{P} \times A$ such that if a is our action and P is the true population, our loss is $L(P, a)$.

MRIE (minimum risk invariant estimator): The MRIE of an unknown parameter θ is the estimator has the minimum risk within the class of invariant estimators.

MLE (maximum likelihood estimator): Let X be a sample from a population P indexed by an unknown parameter vector $\theta \in \Theta \subset \mathcal{R}^k$ and $\ell(\theta)$ be the likelihood function. A $\hat{\theta} \in \Theta$ satisfying $\ell(\hat{\theta}) = \max_{\theta \in \Theta} \ell(\theta)$ is called an MLE of θ (Θ may be replaced by its closure in the above definition).

Measure: A set function ν defined on a σ-field $\mathcal{F}$ on Ω is a measure if (i) $0 \leq \nu(A) \leq \infty$ for any $A \in \mathcal{F}$; (ii) $\nu(\emptyset) = 0$; and (iii) $\nu\left(\cup_{i=1}^\infty A_i\right) = \sum_{i=1}^\infty \nu(A_i)$ for disjoint $A_i \in \mathcal{F}$, $i = 1, 2, ...$

Measurable function: a function from a set Ω to a set Λ (with a given σ-field $\mathcal{G}$) is measurable with respect to a σ-field $\mathcal{F}$ on Ω if $f^{-1}(B) \in \mathcal{F}$ for any $B \in \mathcal{G}$.

Minimax rule: Let X be a sample from a population P and $R_T(P)$ be the risk of a decision rule T. A minimax rule is the rule minimizes $\sup_P R_T(P)$ over all possible T.

Moment generating function: The moment generating function of a distribution F on $\mathcal{R}^k$ is $\int e^{t^\tau x} dF(x)$, $t \in \mathcal{R}^k$, if it is finite.

Monotone likelihood ratio: The family of densities $\{f_\theta : \theta \in \Theta\}$ with $\Theta \subset \mathcal{R}$ is said to have monotone likelihood ratio in $Y(x)$ if, for any $\theta_1 < \theta_2$, $\theta_i \in \Theta$, $f_{\theta_2}(x)/f_{\theta_1}(x)$ is a nondecreasing function of $Y(x)$ for values x at which at least one of $f_{\theta_1}(x)$ and $f_{\theta_2}(x)$ is positive.

Optimal rule: An optimal rule (within a class of rules) is the rule has the smallest risk over all possible populations.

Pivotal quantity: A known Borel function R of (X, θ) is called a pivotal quantity if and only if the distribution of $R(X, \theta)$ does not depend on any unknown quantity.

Population: The distribution (or probability measure) of an observation from a random experiment is called the population.

Power of a test: The power of a test T is the expected value of T with respect to the true population.

Prior and posterior distribution: Let X be a sample from a population indexed by $\theta \in \Theta \subset \mathcal{R}^k$. A distribution defined on Θ that does not depend on X is called a prior. When the population of X is considered as the conditional distribution of X given θ and the prior is considered as the distribution of θ, the conditional distribution of θ given X is called the posterior distribution of θ.

Probability and probability space: A measure P defined on a σ-field $\mathcal{F}$ on a set Ω is called a probability if and only if $P(\Omega) = 1$. The triple $(\Omega, \mathcal{F}, P)$ is called a probability space.

Probability density: Let $(\Omega, \mathcal{F}, P)$ be a probability space and ν be a σ-finite measure on $\mathcal{F}$. If $P \ll \nu$, then the Radon-Nikodym derivative of P with respect to ν is the probability density with respect to ν (and is called Lebesgue density if ν is the Lebesgue measure on $\mathcal{R}^k$).

Random sample: A sample $X = (X_1, ..., X_n)$, where each X_j is a random d-vector with a fixed positive integer d, is called a random sample of size n from a population or distribution P if $X_1, ..., X_n$ are independent and identically distributed as P.

Randomized decision rule: Let X be a sample with range $\mathcal{X}$, A be the action space, and $\mathcal{F}_A$ be a σ-field on A. A randomized decision rule is a function $\delta(x, C)$ on $\mathcal{X} \times \mathcal{F}_A$ such that, for every $C \in \mathcal{F}_A$, $\delta(X, C)$ is a Borel function and, for every $X \in \mathcal{X}$, $\delta(X, C)$ is a probability measure on $\mathcal{F}_A$. A nonrandomized decision rule T can be viewed as a degenerate randomized decision rule δ, i.e., $\delta(X, \{a\}) = I_{\{a\}}(T(X))$ for any $a \in A$ and $X \in \mathcal{X}$.

Risk: The risk of a decision rule is the expectation (with respect to the true population) of the loss of the decision rule.

Sample: The observation from a population treated as a random element is called a sample.

Scale family: A family of Lebesgue densities on $\mathcal{R}$, $\{f_\sigma : \sigma > 0\}$, is a scale family with scale parameter σ if and only if $f_\sigma(x) = \frac{1}{\sigma} f(x/\sigma)$, where f is a known Lebesgue density.

Scale invariant estimator. Let $(X_1, ..., X_n)$ be a random sample from a population in a scale family with scale parameter σ. An estimator $S(X_1, ..., X_n)$ of σ^h with a fixed $h \neq 0$ is scale invariant if and only if $S(rX_1, ..., rX_n) = r^h T(X_1, ..., X_n)$ for any X_i's and $r > 0$.

Simultaneous confidence intervals: Let $\theta_t \in \mathcal{R}$, $t \in \mathcal{T}$. Confidence intervals $C_t(X)$, $t \in \mathcal{T}$, are $1-\alpha$ simultaneous confidence intervals for θ_t, $t \in \mathcal{T}$, if $P(\theta_t \in C_t(X), t \in \mathcal{T}) = 1 - \alpha$.

Statistic: Let X be a sample from a population P. A known Borel function of X is called a statistic.

Sufficiency and minimal sufficiency: Let X be a sample from a population P. A statistic $T(X)$ is sufficient for P if and only if the conditional distribution of X given T does not depend on P. A sufficient statistic T is minimal sufficient if and only if, for any other statistic S sufficient for P, there is a measurable function ψ such that $T = \psi(S)$ except for a set A with $P(X \in A) = 0$ for all P.

Test and its size: Let X be a sample from a population $P \in \mathcal{P}$ and $\mathcal{P}_i$ $i = 0, 1$, be subsets of $\mathcal{P}$ satisfying $\mathcal{P}_0 \cup \mathcal{P}_1 = \mathcal{P}$ and $\mathcal{P}_0 \cap \mathcal{P}_1 = \emptyset$. A randomized test for hypotheses $H_0 : P \in \mathcal{P}_0$ versus $H_1 : P \in \mathcal{P}_1$ is a Borel function $T(X) \in [0, 1]$ such that after X is observed, we reject H_0 (conclude $P \in \mathcal{P}_1$) with probability $T(X)$. If $T(X) \in \{0, 1\}$, then T is nonrandomized. The size of a test T is $\sup_{P \in \mathcal{P}_0} E[T(X)]$, where E is the expectation with respect to P.

UMA (uniformly most accurate) confidence set: Let $\theta \in \Theta$ be an unknown parameter and Θ' be a subset of Θ that does not contain the true value of θ. A confidence set $C(X)$ for θ with confidence coefficient $1 - \alpha$ is Θ'-UMA if and only if for any other confidence set $C_1(X)$ with significance level $1 - \alpha$, $P\big(\theta' \in C(X)\big) \leq P\big(\theta' \in C_1(X)\big)$ for all $\theta' \in \Theta'$.

UMAU (uniformly most accurate unbiased) confidence set: Let $\theta \in \Theta$ be an unknown parameter and Θ' be a subset of Θ that does not contain the true value of θ. A confidence set $C(X)$ for θ with confidence coefficient $1 - \alpha$ is Θ'-UMAU if and only if $C(X)$ is unbiased and for any other unbiased confidence set $C_1(X)$ with significance level $1-\alpha$, $P\big(\theta' \in C(X)\big) \leq P\big(\theta' \in C_1(X)\big)$ for all $\theta' \in \Theta'$.

UMP (uniformly most powerful) test: A test of size α is UMP for testing $H_0 : P \in \mathcal{P}_0$ versus $H_1 : P \in \mathcal{P}_1$ if and only if, at each $P \in \mathcal{P}_1$, the power of T is no smaller than the power of any other level α test.

UMPU (uniformly most powerful unbiased) test: An unbiased test of size α is UMPU for testing $H_0 : P \in \mathcal{P}_0$ versus $H_1 : P \in \mathcal{P}_1$ if and only if, at each $P \in \mathcal{P}_1$, the power of T is no larger than the power of any other level α unbiased test.

UMVUE (uniformly minimum variance estimator): An estimator is a UMVUE if it has the minimum variance within the class of unbiased estimators.

Unbiased confidence set: A level $1-\alpha$ confidence set $C(X)$ is said to be unbiased if and only if $P(\theta' \in C(X)) \leq 1-\alpha$ for any P and all $\theta' \neq \theta$.

Unbiased estimator: Let X be a sample from a population P and $\theta \in \mathcal{R}^k$ be a function of P. If an estimator $T(X)$ of θ satisfies $E[T(X)] = \theta$ for any P, where E is the expectation with respect to P, then $T(X)$ is an unbiased estimator of θ.

Unbiased test: A test for hypotheses $H_0 : P \in \mathcal{P}_0$ versus $H_1 : P \in \mathcal{P}_1$ is unbiased if its size is no larger than its power at any $P \in \mathcal{P}_1$.

Some Distributions

1. Discrete uniform distribution on the set $\{a_1, ..., a_m\}$: The probability density (with respect to the counting measure) of this distribution is

$$f(x) = \begin{cases} m^{-1} & x = a_i, \ i = 1, ..., m \\ 0 & \text{otherwise,} \end{cases}$$

where $a_i \in \mathcal{R}$, $i = 1, ..., m$, and m is a positive integer. The expectation of this distribution is $\bar{a} = \sum_{j=1}^m a_j/m$ and the variance of this distribution is $\sum_{j=1}^m (a_j - \bar{a})^2/m$. The moment generating function of this distribution is $\sum_{j=1}^m e^{a_j t}/m, \ t \in \mathcal{R}$.

2. The binomial distribution with size n and probability p: The probability density (with respect to the counting measure) of this distribution is

$$f(x) = \begin{cases} \binom{n}{x} p^x (1-p)^{n-x} & x = 0, 1, ..., n \\ 0 & \text{otherwise,} \end{cases}$$

where n is a positive integer and $p \in [0, 1]$. The expectation and variance of this distributions are np and $np(1-p)$, respectively. The moment generating function of this distribution is $(pe^t + 1 - p)^n$, $t \in \mathcal{R}$.

3. The Poisson distribution with mean θ: The probability density (with respect to the counting measure) of this distribution is

$$f(x) \begin{cases} \frac{\theta^x e^{-\theta}}{x!} & x = 0, 1, 2, ... \\ 0 & \text{otherwise,} \end{cases}$$

where $\theta > 0$ is the expectation of this distribution. The variance of this distribution is θ. The moment generating function of this distribution is $e^{\theta(e^t - 1)}, \ t \in \mathcal{R}$.

4. The geometric with mean p^{-1}: The probability density (with respect to the counting measure) of this distribution is

$$f(x) = \begin{cases} (1-p)^{x-1} p & x = 1, 2, ... \\ 0 & \text{otherwise,} \end{cases}$$

where $p \in [0, 1]$. The expectation and variance of this distribution are p^{-1} and $(1-p)/p^2$, respectively. The moment generating function of this distribution is $pe^t/[1-(1-p)e^t]$, $t < -\log(1-p)$.

5. Hypergeometric distribution: The probability density (with respect to the counting measure) of this distribution is

$$f(x) = \begin{cases} \frac{\binom{n}{x}\binom{m}{r-x}}{\binom{N}{r}} & x = 0, 1, ..., \min\{r, n\},\ r - x \leq m \\ 0 & \text{otherwise,} \end{cases}$$

where r, n, and m are positive integers, and $N = n + m$. The expectation and variance of this distribution are equal to rn/N and $rnm(N-r)/[N^2(N-1)]$, respectively.

6. Negative binomial with size r and probability p: The probability density (with respect to the counting measure) of this distribution is

$$f(x) = \begin{cases} \binom{x-1}{r-1} p^r (1-p)^{x-r} & x = r, r+1, ... \\ 0 & \text{otherwise,} \end{cases}$$

where $p \in [0, 1]$ and r is a positive integer. The expectation and variance of this distribution are r/p and $r(1-p)/p^2$, respectively. The moment generating function of this distribution is equal to $p^r e^{rt}/[1-(1-p)e^t]^r$, $t < -\log(1-p)$.

7. Log-distribution with probability p: The probability density (with respect to the counting measure) of this distribution is

$$f(x) = \begin{cases} -(\log p)^{-1} x^{-1} (1-p)^x & x = 1, 2, ... \\ 0 & \text{otherwise,} \end{cases}$$

where $p \in (0, 1)$. The expectation and variance of this distribution are $-(1-p)/(p \log p)$ and $-(1-p)[1+(1-p)/\log p]/(p^2 \log p)$, respectively. The moment generating function of this distribution is equal to $\log[1-(1-p)e^t]/\log p$, $t \in \mathcal{R}$.

8. Uniform distribution on the interval (a, b): The Lebesgue density of this distribution is

$$f(x) = \frac{1}{b-a} I_{(a,b)}(x),$$

where a and b are real numbers with $a < b$. The expectation and variance of this distribution are $(a+b)/2$ and $(b-a)^2/12$, respectively. The moment generating function of this distribution is equal to $(e^{bt} - e^{at})/[(b-a)t]$, $t \in \mathcal{R}$.

9. Normal distribution $N(\mu, \sigma^2)$: The Lebesgue density of this distribution is
$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-(x-\mu)^2/2\sigma^2},$$
where $\mu \in \mathcal{R}$ and $\sigma^2 > 0$. The expectation and variance of $N(\mu, \sigma^2)$ are μ and σ^2, respectively. The moment generating function of this distribution is $e^{\mu t + \sigma^2 t^2/2}$, $t \in \mathcal{R}$.

10. Exponential distribution on the interval (a, ∞) with scale parameter θ: The Lebesgue density of this distribution is
$$f(x) = \frac{1}{\theta} e^{-(x-a)/\theta} I_{(a,\infty)}(x),$$
where $a \in \mathcal{R}$ and $\theta > 0$. The expectation and variance of this distribution are $\theta + a$ and θ^2, respectively. The moment generating function of this distribution is $e^{at}(1 - \theta t)^{-1}$, $t < \theta^{-1}$.

11. Gamma distribution with shape parameter α and scale parameter γ: The Lebesgue density of this distribution is
$$f(x) = \frac{1}{\Gamma(\alpha)\gamma^\alpha} x^{\alpha-1} e^{-x/\gamma} I_{(0,\infty)}(x),$$
where $\alpha > 0$ and $\gamma > 0$. The expectation and variance of this distribution are $\alpha\gamma$ and $\alpha\gamma^2$, respectively. The moment generating function of this distribution is $(1 - \gamma t)^{-\alpha}$, $t < \gamma^{-1}$.

12. Beta distribution with parameter (α, β): The Lebesgue density of this distribution is
$$f(x) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1} I_{(0,1)}(x),$$
where $\alpha > 0$ and $\beta > 0$. The expectation and variance of this distribution are $\alpha/(\alpha+\beta)$ and $\alpha\beta/[(\alpha+\beta+1)(\alpha+\beta)^2]$, respectively.

13. Cauchy distribution with location parameter μ and scale parameter σ: The Lebesgue density of this distribution is
$$f(x) = \frac{\sigma}{\pi[\sigma^2 + (x-\mu)^2]},$$
where $\mu \in \mathcal{R}$ and $\sigma > 0$. The expectation and variance of this distribution do not exist. The characteristic function of this distribution is $e^{\sqrt{-1}\mu t - \sigma|t|}$, $t \in \mathcal{R}$.

14. Log-normal distribution with parameter (μ, σ^2): The Lebesgue density of this distribution is

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma x} e^{-(\log x - \mu)^2/2\sigma^2} I_{(0,\infty)}(x),$$

where $\mu \in \mathcal{R}$ and $\sigma^2 > 0$. The expectation and variance of this distribution are $e^{\mu+\sigma^2/2}$ and $e^{2\mu+\sigma^2}(e^{\sigma^2} - 1)$, respectively.

15. Weibull distribution with shape parameter α and scale parameter θ: The Lebesgue density of this distribution is

$$f(x) = \frac{\alpha}{\theta} x^{\alpha-1} e^{-x^\alpha/\theta} I_{(0,\infty)}(x),$$

where $\alpha > 0$ and $\theta > 0$. The expectation and variance of this distribution are $\theta^{1/\alpha}\Gamma(\alpha^{-1} + 1)$ and $\theta^{2/\alpha}\{\Gamma(2\alpha^{-1} + 1) - [\Gamma(\alpha^{-1} + 1)]^2\}$, respectively.

16. Double exponential distribution with location parameter μ and scale parameter θ: The Lebesgue density of this distribution is

$$f(x) = \frac{1}{2\theta} e^{-|x-\mu|/\theta},$$

where $\mu \in \mathcal{R}$ and $\theta > 0$. The expectation and variance of this distribution are μ and $2\theta^2$, respectively. The moment generating function of this distribution is $e^{\mu t}/(1 - \theta^2 t^2)$, $|t| < \theta^{-1}$.

17. Pareto distribution: The Lebesgue density of this distribution is

$$f(x) = \theta a^\theta x^{-(\theta+1)} I_{(a,\infty)}(x),$$

where $a > 0$ and $\theta > 0$. The expectation this distribution is $\theta a/(\theta - 1)$ when $\theta > 1$ and does not exist when $\theta \leq 1$. The variance of this distribution is $\theta a^2/[(\theta - 1)^2(\theta - 2)]$ when $\theta > 2$ and does not exist when $\theta \leq 2$.

18. Logistic distribution with location parameter μ and scale parameter σ: The Lebesgue density of this distribution is

$$f(x) = \frac{e^{-(x-\mu)/\sigma}}{\sigma[1 + e^{-(x-\mu)/\sigma}]^2},$$

where $\mu \in \mathcal{R}$ and $\sigma > 0$. The expectation and variance of this distribution are μ and $\sigma^2\pi^2/3$, respectively. The moment generating function of this distribution is $e^{\mu t}\Gamma(1 + \sigma t)\Gamma(1 - \sigma t)$, $|t| < \sigma^{-1}$.

19. Chi-square distribution χ_k^2: The Lebesgue density of this distribution is

$$f(x) = \frac{1}{\Gamma(k/2)2^{k/2}} x^{k/2-1} e^{-x/2} I_{(0,\infty)}(x),$$

where k is a positive integer. The expectation and variance of this distribution are k and $2k$, respectively. The moment generating function of this distribution is $(1-2t)^{-k/2}$, $t < 1/2$.

20. Noncentral chi-square distribution $\chi_k^2(\delta)$: This distribution is defined as the distribution of $X_1^2 + \cdots + X_k^2$, where $X_1, ..., X_k$ are independent and identically distributed as $N(\mu_i, 1)$, k is a positive integer, and $\delta = \mu_1^2 + \cdots + \mu_k^2 \geq 0$. δ is called the noncentrality parameter. The Lebesgue density of this distribution is

$$f(x) = e^{-\delta/2} \sum_{j=0}^{\infty} \frac{(\delta/2)^j}{j!} f_{2j+n}(x),$$

where $f_k(x)$ is the Lebesgue density of the chi-square distribution χ_k^2. The expectation and variance of this distribution are $k + \delta$ and $2k + 4\delta$, respectively. The characteristic function of this distribution is $(1 - 2\sqrt{-1}t)^{-k/2} e^{\sqrt{-1}\delta t/(1-2\sqrt{-1}t)}$.

21. t-distribution t_n: The Lebesgue density of this distribution is

$$f(x) = \frac{\Gamma(\frac{n+1}{2})}{\sqrt{n\pi}\Gamma(\frac{n}{2})} \left(1 + \frac{x^2}{n}\right)^{-(n+1)/2},$$

where n is a positive integer. The expectation of t_n is 0 when $n > 1$ and does not exist when $n = 1$. The variance of t_n is $n/(n-2)$ when $n > 2$ and does not exist when $n \leq 2$.

22. Noncentral t-distribution $t_n(\delta)$: This distribution is defined as the distribution of $X/\sqrt{Y/n}$, where X is distributed as $N(\delta, 1)$, Y is distributed as χ_n^2, X and Y are independent, n is a positive integer, and $\delta \in \mathcal{R}$ is called the noncentrality parameter. The Lebesgue density of this distribution is

$$f(x) = \frac{1}{2^{(n+1)/2}\Gamma(\frac{n}{2})\sqrt{\pi n}} \int_0^\infty y^{(n-1)/2} e^{-[(x\sqrt{y/n}-\delta)^2+y]/2} dy.$$

The expectation of $t_n(\delta)$ is $\delta\Gamma(\frac{n-1}{2})\sqrt{n/2}/\Gamma(\frac{n}{2})$ when $n > 1$ and does not exist when $n = 1$. The variance of $t_n(\delta)$ is $[n(1+\delta^2)/(n-2)] - [\Gamma(\frac{n-1}{2})/\Gamma(\frac{n}{2})]^2\delta^2 n/2$ when $n > 2$ and does not exist when $n \leq 2$.

23. F-distribution $F_{n,m}$: The Lebesgue density of this distribution is

$$f(x) = \frac{n^{n/2}m^{m/2}\Gamma(\frac{n+m}{2})x^{n/2-1}}{\Gamma(\frac{n}{2})\Gamma(\frac{m}{2})(m+nx)^{(n+m)/2}} I_{(0,\infty)}(x),$$

where n and m are positive integers. The expectation of $F_{n,m}$ is $m/(m-2)$ when $m > 2$ and does not exist when $m \leq 2$. The variance of $F_{n,m}$ is $2m^2(n+m-2)/[n(m-2)^2(m-4)]$ when $m > 4$ and does not exist when $m \leq 4$.

24. Noncentral F-distribution $F_{n,m}(\delta)$: This distribution is defined as the distribution of $(X/n)/(Y/m)$, where X is distributed as $\chi^2_n(\delta)$, Y is distributed as χ^2_m, X and Y are independent, n and m are positive integers, and $\delta \geq 0$ is called the noncentrality parameter. The Lebesgue density of this distribution is

$$f(x) = e^{-\delta/2} \sum_{j=0}^{\infty} \frac{n_1(\delta/2)^j}{j!(2j+n_1)} f_{2j+n_1,n_2}\left(\frac{n_1 x}{2j+n_1}\right),$$

where $f_{k_1,k_2}(x)$ is the Lebesgue density of F_{k_1,k_2}. The expectation of $F_{n,m}(\delta)$ is $m(n+\delta)/[n(m-2)]$ when $m > 2$ and does not exist when $m \leq 2$. The variance of $F_{n,m}(\delta)$ is $2m^2[(n+\delta)^2 + (m-2)(n+2\delta)]/[n^2(m-2)^2(m-4)]$ when $m > 4$ and does not exist when $m \leq 4$.

25. Multinomial distribution with size n and probability vector $(p_1,...,p_k)$: The probability density (with respect to the counting measure on $\mathcal{R}^k$) is

$$f(x_1,...,x_k) = \frac{n!}{x_1!\cdots x_k!} p_1^{x_1}\cdots p_k^{x_k} I_B(x_1,...,x_k),$$

where $B = \{(x_1,...,x_k) : x_i\text{'s are nonnegative integers}, \sum_{i=1}^k x_i = n\}$, n is a positive integer, $p_i \in [0,1]$, $i = 1,...,k$, and $\sum_{i=1}^k p_i = 1$. The mean-vector (expectation) of this distribution is $(np_1,...,np_k)$. The variance-covariance matrix of this distribution is the $k \times k$ matrix whose ith diagonal element is np_i and (i,j)th off-diagonal element is $-np_ip_j$.

26. Multivariate normal distribution $N_k(\mu,\Sigma)$: The Lebesgue density of this distribution is

$$f(x) = \frac{1}{(2\pi)^{k/2}[\text{Det}(\Sigma)]^{1/2}} e^{-(x-\mu)^\tau\Sigma^{-1}(x-\mu)/2}, \quad x \in \mathcal{R}^k,$$

where $\mu \in \mathcal{R}^k$ and Σ is a positive definite $k \times k$ matrix. The mean-vector (expectation) of this distribution is μ. The variance-covariance matrix of this distribution is Σ. The moment generating function of $N_k(\mu,\Sigma)$ is $e^{t^\tau\mu + t^\tau\Sigma t/2}$, $t \in \mathcal{R}^k$.

Chapter 1

Probability Theory

Exercise 1. Let Ω be a set, $\mathcal{F}$ be σ-field on Ω, and $C \in \mathcal{F}$. Show that $\mathcal{F}_C = \{C \cap A : A \in \mathcal{F}\}$ is a σ-field on C.
Solution. This exercise, similar to many other problems, can be solved by directly verifying the three properties in the definition of a σ-field.
(i) The empty subset of C is $C \cap \emptyset$. Since $\mathcal{F}$ is a σ-field, $\emptyset \in \mathcal{F}$. Then, $C \cap \emptyset \in \mathcal{F}_C$.
(ii) If $B \in \mathcal{F}_C$, then $B = C \cap A$ for some $A \in \mathcal{F}$. Since $\mathcal{F}$ is a σ-field, $A^c \in \mathcal{F}$. Then the complement of B in C is $C \cap A^c \in \mathcal{F}_C$.
(iii) If $B_i \in \mathcal{F}_C$, $i = 1, 2, ...$, then $B_i = C \cup A_i$ for some $A_i \in \mathcal{F}$, $i = 1, 2, ...$. Since $\mathcal{F}$ is a σ-field, $\cup A_i \in \mathcal{F}$. Therefore, $\cup B_i = \cup (C \cap A_i) = C \cap (\cup A_i) \in \mathcal{F}_C$. ∎

Exercise 2 (#1.12)[†]. Let ν and λ be two measures on a σ-field $\mathcal{F}$ on Ω such that $\nu(A) = \lambda(A)$ for any $A \in \mathcal{C}$, where $\mathcal{C} \subset \mathcal{F}$ is a collection having the property that if A and B are in $\mathcal{C}$, then so is $A \cap B$. Assume that there are $A_i \in \mathcal{C}$, $i = 1, 2, ...$, such that $\cup A_i = \Omega$ and $\nu(A_i) < \infty$ for all i. Show that $\nu(A) = \lambda(A)$ for any $A \in \sigma(\mathcal{C})$, where $\sigma(\mathcal{C})$ is the smallest σ-field containing $\mathcal{C}$.
Note. Solving this problem requires knowing properties of measures (Shao, 2003, §1.1.1). The technique used in solving this exercise is called the "good sets principle". All sets in $\mathcal{C}$ have property A and we want to show that all sets in $\sigma(\mathcal{C})$ also have property A. Let $\mathcal{G}$ be the collection of all sets having property A (good sets). Then, all we need to show is that $\mathcal{G}$ is a σ-field.
Solution. Define $\mathcal{G} = \{A \in \mathcal{F} : \nu(A) = \lambda(A)\}$. Since $\mathcal{C} \subset \mathcal{G}$, $\sigma(\mathcal{C}) \subset \mathcal{G}$ if $\mathcal{G}$ is a σ-field. Hence, the result follows if we can show that $\mathcal{G}$ is a σ-field.
(i) Since both ν and λ are measures, $0 = \nu(\emptyset) = \lambda(\emptyset)$ and, thus, the empty set $\emptyset \in \mathcal{G}$.

[†]The number in parentheses is the exercise number in *Mathematical Statistics* (Shao, 2003). The first digit is the chapter number.

(ii) For any $B \in \mathcal{F}$, by the inclusion and exclusion formula,

$$\nu\left(\bigcup_{i=1}^{n} A_i \cap B\right) = \sum_{1\le i\le n} \nu(A_i \cap B) - \sum_{1\le i<j\le n} \nu(A_i \cap A_j \cap B) + \cdots$$

for any positive integer n, where A_i's are the sets given in the description of this exercise. The same result also holds for λ. Since A_j's are in $\mathcal{C}$, $A_i \cap A_j \cap \cdots \cap A_k \in \mathcal{C}$ and, if $B \in \mathcal{G}$,

$$\nu(A_i \cap A_j \cap \cdots \cap A_k \cap B) = \lambda(A_i \cap A_j \cap \cdots \cap A_k \cap B) < \infty.$$

Consequently,

$$\nu(A_i \cap A_j \cap \cdots \cap A_k \cap B^c) = \lambda(A_i \cap A_j \cap \cdots \cap A_k \cap B^c) < \infty.$$

By the inclusion and exclusion formula again, we obtain that

$$\nu\left(\bigcup_{i=1}^{n} A_i \cap B^c\right) = \lambda\left(\bigcup_{i=1}^{n} A_i \cap B^c\right)$$

for any n. From the continuity property of measures (Proposition 1.1(iii) in Shao, 2003), we conclude that $\nu(B^c) = \lambda(B^c)$ by letting $n \to \infty$ in the previous expression. Thus, $B^c \in \mathcal{G}$ whenever $B \in \mathcal{G}$.
(iii) Suppose that $B_i \in \mathcal{G}$, $i = 1, 2, ...$ Note that

$$\nu(B_1 \cup B_2) = \nu(B_1) + \nu(B_1^c \cap B_2) = \lambda(B_1) + \lambda(B_1^c \cap B_2) = \lambda(B_1 \cup B_2),$$

since $B_1^c \cap B_2 \in \mathcal{G}$. Thus, $B_1 \cup B_2 \in \mathcal{G}$. This shows that for any n, $\cup_{i=1}^n B_i \in \mathcal{G}$. By the continuity property of measures,

$$\nu\left(\bigcup_{i=1}^{\infty} B_i\right) = \lim_{n\to\infty} \nu\left(\bigcup_{i=1}^{n} B_i\right) = \lim_{n\to\infty} \lambda\left(\bigcup_{i=1}^{n} B_i\right) = \lambda\left(\bigcup_{i=1}^{\infty} B_i\right).$$

Hence, $\cup B_i \in \mathcal{G}$. ∎

Exercise 3 (#1.14). Show that a real-valued function f on a set Ω is Borel with respect to a σ-field $\mathcal{F}$ on Ω if and only if $f^{-1}(a, \infty) \in \mathcal{F}$ for all $a \in \mathcal{R}$.
Note. Good sets principle is used in this solution.
Solution. The only if part follows directly from the definition of a Borel function. Suppose that $f^{-1}(a, \infty) \in \mathcal{F}$ for all $a \in \mathcal{R}$. Let

$$\mathcal{G} = \{C \subset \mathcal{R} : f^{-1}(C) \in \mathcal{F}\}.$$

Note that (i) $\emptyset \in \mathcal{G}$; (ii) if $C \in \mathcal{G}$, then $f^{-1}(C^c) = (f^{-1}(C))^c \in \mathcal{F}$, i.e., $C^c \in \mathcal{G}$; and (iii) if $C_i \in \mathcal{G}$, $i = 1, 2, ...$, then $f^{-1}(\cup C_i) = \cup f^{-1}(C_i) \in \mathcal{F}$, i.e., $\cup C_i \in \mathcal{G}$. This shows that $\mathcal{G}$ is a σ-field. Thus $\mathcal{B} \subset \mathcal{G}$, i.e., $f^{-1}(B) \in \mathcal{F}$ for any $B \in \mathcal{B}$ and, hence, f is Borel. ∎

Exercise 4 (#1.14). Let f and g be real-valued functions on Ω. Show that if f and g are Borel with respect to a σ-field $\mathcal{F}$ on Ω, then so are fg, f/g (when $g \neq 0$), and $af + bg$, where a and b are real numbers.
Solution. Suppose that f and g are Borel. Consider $af + bg$ with $a > 0$ and $b > 0$. Let $\mathcal{Q}$ be the set of all rational numbers on $\mathcal{R}$. For any $c \in \mathcal{R}$,

$$\{af + bg > c\} = \bigcup_{t \in \mathcal{Q}} \{f > (c - t)/a\} \cap \{g > t/b\}.$$

Since f and g are Borel, $\{af + bg > c\} \in \mathcal{F}$. By Exercise 3, $af + bg$ is Borel. Similar results can be obtained for the case of $a > 0$ and $b < 0$, $a < 0$ and $b > 0$, or $a < 0$ and $b < 0$.

From the above result, $f + g$ and $f - g$ are Borel if f and g are Borel. Note that for any $c > 0$,

$$\{(f + g)^2 > c\} = \{f + g > \sqrt{c}\} \cup \{f + g < -\sqrt{c}\}.$$

Hence, $(f + g)^2$ is Borel. Similarly, $(f - g)^2$ is Borel. Then

$$fg = [(f + g)^2 - (f - g)^2]/4$$

is Borel.

Since any constant function is Borel, this shows that af is Borel if f is Borel and a is a constant. Thus, $af + bg$ is Borel even when one of a and b is 0.

Assume $g \neq 0$. For any c,

$$\{1/g > c\} = \begin{cases} \{0 < g < 1/c\} & c > 0 \\ \{g > 0\} & c = 0 \\ \{g > 0\} \cup \{1/c < g < 0\} & c < 0. \end{cases}$$

Hence $1/g$ is Borel if g is Borel and $g \neq 0$. Then f/g is Borel if both f and g are Borel and $g \neq 0$. ∎

Exercise 5 (#1.14). Let f_i, $i = 1, 2, ...$, be Borel functions on Ω with respect to a σ-field $\mathcal{F}$. Show that $\sup_n f_n$, $\inf_n f_n$, $\limsup_n f_n$, and $\liminf_n f_n$ are Borel with respect to $\mathcal{F}$. Also, show that the set

$$A = \left\{\omega \in \Omega : \lim_n f_n(\omega) \text{ exists}\right\}$$

is in $\mathcal{F}$ and the function

$$h(\omega) = \begin{cases} \lim_n f_n(\omega) & \omega \in A \\ f_1(\omega) & \omega \notin A \end{cases}$$

is Borel with respect to $\mathcal{F}$.

Solution. For any $c \in \mathcal{R}$, $\{\sup_n f_n > c\} = \cup_n \{f_n > c\}$. By Exercise 3, $\sup_n f_n$ is Borel. By Exercise 4, $\inf_n f_n = -\sup_n(-f_n)$ is Borel. Then $\limsup_n f_n = \inf_n \sup_{k\geq n} f_k$ is Borel and $\liminf_n f_n = -\limsup_n(-f_n)$ is Borel. Consequently, $A = \{\limsup_n f_n - \liminf_n f_n = 0\} \in \mathcal{F}$. The function h is equal to $I_A \limsup_n f_n + I_{A^c} f_1$, where I_A is the indicator function of the set A. Since $A \in \mathcal{F}$, I_A is Borel. Thus, h is Borel. ∎

Exercise 6. Let f be a Borel function on $\mathcal{R}^2$. Define a function g from $\mathcal{R}$ to $\mathcal{R}$ as $g(x) = f(x, y_0)$, where y_0 is a fixed point in $\mathcal{R}$. Show that g is Borel. Is it true that f is Borel from $\mathcal{R}^2$ to $\mathcal{R}$ if $f(x, y)$ with any fixed y or fixed x is Borel from $\mathcal{R}$ to $\mathcal{R}$?

Solution. For a fixed y_0, define

$$\mathcal{G} = \{C \subset \mathcal{R}^2 : \{x : (x, y_0) \in C\} \in \mathcal{B}\}.$$

Then, (i) $\emptyset \in \mathcal{G}$; (ii) if $C \in \mathcal{G}$, $\{x : (x, y_0) \in C^c\} = \{x : (x, y_0) \in C\}^c \in \mathcal{B}$, i.e., $C^c \in \mathcal{G}$; (iii) if $C_i \in \mathcal{G}$, $i = 1, 2, ...$, then $\{x : (x, y_0) \in \cup C_i\} = \cup\{x : (x, y_0) \in C_i\} \in \mathcal{B}$, i.e., $\cup C_i \in \mathcal{G}$. Thus, $\mathcal{G}$ is a σ-field. Since any open rectangle $(a, b) \times (c, d) \in \mathcal{G}$, $\mathcal{G}$ is a σ-field containing all open rectangles and, thus, $\mathcal{G}$ contains $\mathcal{B}^2$, the Borel σ-field on $\mathcal{R}^2$. Let $B \in \mathcal{B}$. Since f is Borel, $A = f^{-1}(B) \in \mathcal{B}^2$. Then $A \in \mathcal{G}$ and, thus,

$$g^{-1}(B) = \{x : f(x, y_0) \in B\} = \{x : (x, y_0) \in A\} \in \mathcal{B}.$$

This proves that g is Borel.

If $f(x, y)$ with any fixed y or fixed x is Borel from $\mathcal{R}$ to $\mathcal{R}$, f is not necessarily to be a Borel function from $\mathcal{R}^2$ to $\mathcal{R}$. The following is a counterexample. Let A be a non-Borel subset of $\mathcal{R}$ and

$$f(x, y) = \begin{cases} 1 & x = y \in A \\ 0 & \text{otherwise} \end{cases}$$

Then for any fixed y_0, $f(x, y_0) = 0$ if $y_0 \notin A$ and $f(x, y_0) = I_{\{y_0\}}(x)$ (the indicator function of the set $\{y_0\}$) if $y_0 \in A$. Hence $f(x, y_0)$ is Borel. Similarly, $f(x_0, y)$ is Borel for any fixed x_0. We now show that $f(x, y)$ is not Borel. Suppose that it is Borel. Then $B = \{(x, y) : f(x, y) = 1\} \in \mathcal{B}^2$. Define $\mathcal{G} = \{C \subset \mathcal{R}^2 : \{x : (x, x) \in C\} \in \mathcal{B}\}$. Using the same argument in the proof of the first part, we can show that $\mathcal{G}$ is a σ-field containing $\mathcal{B}^2$. Hence $\{x : (x, x) \in B\} \in \mathcal{B}$. However, by definition $\{x : (x, x) \in B\} = A \notin \mathcal{B}$. This contradiction proves that $f(x, y)$ is not Borel. ∎

Exercise 7 (#1.21). Let $\Omega = \{\omega_i : i = 1, 2, ...\}$ be a countable set, $\mathcal{F}$ be all subsets of Ω, and ν be the counting measure on Ω (i.e., $\nu(A) =$ the number of elements in A for any $A \subset \Omega$). For any Borel function f, the integral of f w.r.t. ν (if it exists) is

$$\int f d\nu = \sum_{i=1}^{\infty} f(\omega_i).$$

Note. The definition of integration and properties of integration can be found in Shao (2003, §1.2). This type of exercise is much easier to solve if we first consider nonnegative functions (or simple nonnegative functions) and then general functions by using f_+ and f_-. See also the next exercise for another example.

Solution. First, consider nonnegative f. Then $f = \sum_{i=1}^{\infty} a_i I_{\{\omega_i\}}$, where $a_i = f(\omega_i) \geq 0$. Since $f_n = \sum_{i=1}^{n} a_i I_{\{\omega_i\}}$ is a nonnegative simple function (a function is simple if it is a linear combination of finitely many indicator functions of sets in $\mathcal{F}$) and $f_n \leq f$, by definition

$$\int f_n d\nu = \sum_{i=1}^{n} a_i \leq \int f d\nu.$$

Letting $n \to \infty$ we obtain that

$$\int f d\nu \geq \sum_{i=1}^{\infty} a_i.$$

Let $s = \sum_{i=1}^{k} b_i I_{\{\omega_i\}}$ be a nonnegative simple function satisfying $s \leq f$. Then $0 \leq b_i \leq a_i$ and

$$\int s d\nu = \sum_{i=1}^{k} b_i \leq \sum_{i=1}^{\infty} a_i.$$

Hence

$$\int f d\nu = \sup\left\{\int s d\nu : s \text{ is simple}, 0 \leq s \leq f\right\} \leq \sum_{i=1}^{\infty} a_i$$

and, thus,

$$\int f d\nu = \sum_{i=1}^{\infty} a_i$$

for nonnegative f.

For general f, let $f_+ = \max\{f, 0\}$ and $f_- = \max\{-f, 0\}$. Then

$$\int f_+ d\nu = \sum_{i=1}^{\infty} f_+(\omega_i) \quad \text{and} \quad \int f_- d\nu = \sum_{i=1}^{\infty} f_-(\omega_i).$$

Then the result follows from

$$\int f d\nu = \int f_+ d\nu - \int f_- d\nu$$

if at least one of $\int f_+ d\nu$ and $\int f_- d\nu$ is finite. ∎

Exercise 8 (#1.22). Let ν be a measure on a σ-field $\mathcal{F}$ on Ω and f and g be Borel functions with respect to $\mathcal{F}$. Show that
(i) if $\int f d\nu$ exists and $a \in \mathcal{R}$, then $\int (af) d\nu$ exists and is equal to $a \int f d\nu$;
(ii) if both $\int f d\nu$ and $\int g d\nu$ exist and $\int f d\nu + \int g d\nu$ is well defined, then $\int (f+g) d\nu$ exists and is equal to $\int f d\nu + \int g d\nu$.
Note. For integrals in calculus, properties such as $\int (af) d\nu = a \int f d\nu$ and $\int (f+g) d\nu \int f d\nu + \int g d\nu$ are obvious. However, the proof of them are complicated for integrals defined on general measure spaces. As shown in this exercise, the proof often has to be broken into several steps: simple functions, nonnegative functions, and then general functions.
Solution. (i) If $a = 0$, then $\int (af) d\nu = \int 0 d\nu = 0 = a \int f d\nu$.

Suppose that $a > 0$ and $f \geq 0$. By definition, there exists a sequence of nonnegative simple functions s_n such that $s_n \leq f$ and $\lim_n \int s_n d\nu = \int f d\nu$. Then $as_n \leq af$ and $\lim_n \int as_n d\nu = a \lim_n \int s_n d\nu = a \int f d\nu$. This shows $\int (af) d\nu \geq a \int f d\nu$. Let $b = a^{-1}$ and consider the function $h = b^{-1} f$. From what we have shown, $\int f d\nu = \int (bh) d\nu \geq b \int h d\nu = a^{-1} \int (af) d\nu$. Hence $\int (af) d\nu = a \int f d\nu$.

For $a > 0$ and general f, the result follows by considering $af = af_+ - af_-$. For $a < 0$, the result follows by considering $af = |a| f_- - |a| f_+$.
(ii) Consider the case where $f \geq 0$ and $g \geq 0$. If both f and g are simple functions, the result is obvious. Let s_n, t_n, and r_n be simple functions such that $0 \leq s_n \leq f$, $\lim_n \int s_n d\nu = \int f d\nu$, $0 \leq t_n \leq g$, $\lim_n \int t_n d\nu = \int g d\nu$, $0 \leq r_n \leq f + g$, and $\lim_n \int r_n d\nu = \int (f+g) d\nu$. Then $s_n + t_n$ is simple, $0 \leq s_n + t_n \leq f + g$, and

$$\begin{aligned} \int f d\nu + \int g d\nu &= \lim_n \int s_n d\nu + \lim_n \int t_n d\nu \\ &= \lim_n \int (s_n + t_n) d\nu, \end{aligned}$$

which implies

$$\int f d\nu + \int g d\nu \leq \int (f+g) d\nu.$$

If any of $\int f d\nu$ and $\int g d\nu$ is infinite, then so is $\int (f+g) d\nu$. Hence, we only need to consider the case where both f and g are integrable. Suppose that g is simple. Then $r_n - g$ is simple and

$$\lim_n \int r_n d\nu - \int g d\nu = \lim_n \int (r_n - g) d\nu \leq \int f d\nu,$$

since $r_n - g \le f$. Hence

$$\int (f+g)d\nu = \lim_n \int r_n d\nu \le \int f d\nu + \int g d\nu$$

and, thus, the result follows if g is simple. For a general g, by the proved result,

$$\lim_n \int r_n d\nu - \int g d\nu = \lim_n \int (r_n - g) d\nu.$$

Hence $\int (f+g)d\nu = \lim_n \int r_n d\nu \le \int f d\nu + \int g d\nu$ and the result follows.

Consider general f and g. Note that

$$(f+g)_+ - (f+g)_- = f + g = f_+ - f_- + g_+ - g_-,$$

which leads to

$$(f+g)_+ + f_- + g_- = (f+g)_- + f_+ + g_+.$$

From the proved result for nonnegative functions,

$$\begin{aligned}\int [(f+g)_+ + f_- + g_-] d\nu &= \int (f+g)_+ d\nu + \int f_- d\nu + \int g_- d\nu \\ &= \int [(f+g)_- + f_+ + g_+] d\nu \\ &= \int (f+g)_- d\nu + \int f_+ d\nu + \int g_+ d\nu.\end{aligned}$$

If both f and g are integrable, then

$$\int (f+g)_+ d\nu - \int (f+g)_- d\nu = \int f_+ d\nu - \int f_- d\nu + \int g_+ d\nu - \int g_- d\nu,$$

i.e.,

$$\int (f+g) d\nu = \int f d\nu + \int g d\nu.$$

Suppose now that $\int f_- d\nu = \infty$. Then $\int f_+ d\nu < \infty$ since $\int f d\nu$ exists. Since $\int f d\nu + \int g d\nu$ is well defined, we must have $\int g_+ d\nu < \infty$. Since $(f+g)_+ \le f_+ + g_+$, $\int (f+g)_+ d\nu < \infty$. Thus, $\int (f+g)_- d\nu = \infty$ and $\int (f+g) d\nu = -\infty$. On the other hand, we also have $\int f d\nu + \int g d\nu = -\infty$. Similarly, we can prove the case where $\int f_+ d\nu = \infty$ and $\int f_- d\nu < \infty$. ∎

Exercise 9 (#1.30). Let F be a cumulative distribution function on the real line $\mathcal{R}$ and $a \in \mathcal{R}$. Show that

$$\int [F(x+a) - F(x)] dx = a.$$

Solution. For $a \geq 0$,

$$\int [F(x+a) - F(x)]dx = \int \int I_{(x,x+a]}(y)dF(y)dx.$$

Since $I_{(x,x+a]}(y) \geq 0$, by Fubini's theorem, the above integral is equal to

$$\int \int I_{(y-a,y]}(x)dxdF(y) = \int adF(y) = a.$$

The proof for the case of $a < 0$ is similar. ∎

Exercise 10 (#1.31). Let F and G be two cumulative distribution functions on the real line. Show that if F and G have no common points of discontinuity in the interval $[a, b]$, then

$$\int_{(a,b]} G(x)dF(x) = F(b)G(b) - F(a)G(a) - \int_{(a,b]} F(x)dG(x).$$

Solution. Let P_F and P_G be the probability measures corresponding to F and G, respectively, and let $P = P_F \times P_G$ be the product measure (see Shao, 2003, §1.1.1). Consider the following three Borel sets in $\mathcal{R}^2$: $A = \{(x, y) : x \leq y, a < y \leq b\}$, $B = \{(x, y) : y \leq x, a < x \leq b\}$, and $C = \{(x, y) : a < x \leq b, x = y\}$. Since F and G have no common points of discontinuity, $P(C) = 0$. Then,

$$\begin{aligned}
F(b)G(b) - F(a)G(a) &= P\big((-\infty, b]\times(-\infty, b]\big) - P\big((-\infty, a]\times(-\infty, a]\big) \\
&= P(A) + P(B) - P(C) \\
&= P(A) + P(B) \\
&= \int_A dP + \int_B dP \\
&= \int_{(a,b]} \int_{(-\infty,y]} dP_F dP_G + \int_{(a,b]} \int_{(-\infty,x]} dP_G dP_F \\
&= \int_{(a,b]} F(y)dP_G + \int_{(a,b]} G(x)dP_F \\
&= \int_{(a,b]} F(y)dG(y) + \int_{(a,b]} G(x)dF(x) \\
&= \int_{(a,b]} F(x)dG(x) + \int_{(a,b]} G(x)dF(x),
\end{aligned}$$

where the fifth equality follows from Fubini's theorem. ∎

Exercise 11. Let Y be a random variable and m be a median of Y, i.e., $P(Y \leq m) \geq 1/2$ and $P(Y \geq m) \geq 1/2$. Show that, for any real numbers

a and b such that $m \le a \le b$ or $m \ge a \ge b$, $E|Y-a| \le E|Y-b|$.
Solution. We can assume $E|Y| < \infty$, otherwise $\infty = E|Y-a| \le E|Y-b| = \infty$. Assume $m \le a \le b$. Then

$$\begin{aligned} E|Y-b| - E|Y-a| &= E[(b-Y)I_{\{Y\le b\}}] + E[(Y-b)I_{\{Y>b\}}] \\ &\quad - E[(a-Y)I_{\{Y\le a\}}] - E[(Y-a)I_{\{Y>a\}}] \\ &= 2E[(b-Y)I_{\{a<Y\le b\}}] \\ &\quad + (a-b)[E(I_{\{Y>a\}}) - E(I_{\{Y\le a\}})] \\ &\ge (a-b)[1-2P(Y\le a)] \\ &\ge 0, \end{aligned}$$

since $P(Y \le a) \ge P(Y \le m) \ge 1/2$. If $m \ge a \ge b$, then $-m \le -a \le -b$ and $-m$ is a median of $-Y$. From the proved result, $E|(-Y)-(-b)| \ge E|(-Y)-(-a)|$, i.e., $E|Y-a| \le E|Y-b|$. ∎

Exercise 12. Let X and Y be independent random variables satisfying $E|X+Y|^a < \infty$ for some $a > 0$. Show that $E|X|^a < \infty$.
Solution. Let $c \in \mathcal{R}$ such that $P(Y > c) > 0$ and $P(Y \le c) > 0$. Note that

$$\begin{aligned} E|X+Y|^a &\ge E(|X+Y|^a I_{\{Y>c, X+c>0\}}) + E(|X+Y|^a I_{\{Y\le c, X+c\le 0\}}) \\ &\ge E(|X+c|^a I_{\{Y>c, X+c>0\}}) + E(|X+c|^a I_{\{Y\le c, X+c\le 0\}}) \\ &= P(Y>c)E(|X+c|^a I_{\{X+c>0\}}) \\ &\quad + P(Y\le c)E(|X+c|^a I_{\{X+c\le 0\}}), \end{aligned}$$

where the last inequality follows from the independence of X and Y. Since $E|X+Y|^a < \infty$, both $E(|X+c|^a I_{\{X+c>0\}})$ and $E(|X+c|^a I_{\{X+c\le 0\}})$ are finite and

$$E|X+c|^a = E(|X+c|^a I_{\{X+c>0\}}) + E(|X+c|^a I_{\{X+c\le 0\}}) < \infty.$$

Then,

$$E|X|^a \le 2^a(E|X+c|^a + |c|^a) < \infty. \quad ∎$$

Exercise 13 (#1.34). Let ν be a σ-finite measure on a σ-field $\mathcal{F}$ on Ω, λ be another measure with $\lambda \ll \nu$, and f be a nonnegative Borel function on Ω. Show that

$$\int f d\lambda = \int f \frac{d\lambda}{d\nu} d\nu,$$

where $\frac{d\lambda}{d\nu}$ is the Radon-Nikodym derivative.
Note. Two measures λ and ν satisfying $\lambda \ll \nu$ if $\nu(A) = 0$ always implies

$\lambda(A) = 0$, which ensures the existence of the Radon-Nikodym derivative $\frac{d\lambda}{d\nu}$ when ν is σ-finite (see Shao, 2003, §1.1.2).

Solution. By the definition of the Radon-Nikodym derivative and the linearity of integration, the result follows if f is a simple function. For a general nonnegative f, there is a sequence $\{s_n\}$ of nonnegative simple functions such that $s_n \leq s_{n+1}$, $n = 1, 2, \ldots$, and $\lim_n s_n = f$. Then $0 \leq s_n \frac{d\lambda}{d\nu} \leq s_{n+1} \frac{d\lambda}{d\nu}$ and $\lim_n s_n \frac{d\lambda}{d\nu} = f \frac{d\lambda}{d\nu}$. By the monotone convergence theorem (e.g., Theorem 1.1 in Shao, 2003),

$$\int f d\lambda = \lim_n \int s_n d\lambda = \lim_n \int s_n \frac{d\lambda}{d\nu} d\nu = \int f \frac{d\lambda}{d\nu} d\nu. \quad \blacksquare$$

Exercise 14 (#1.34). Let $\mathcal{F}_i$ be a σ-field on Ω_i, ν_i be a σ-finite measure on $\mathcal{F}_i$, and λ_i be a measure on $\mathcal{F}_i$ with $\lambda_i \ll \nu_i$, $i = 1, 2$. Show that $\lambda_1 \times \lambda_2 \ll \nu_1 \times \nu_2$ and

$$\frac{d(\lambda_1 \times \lambda_2)}{d(\nu_1 \times \nu_2)} = \frac{d\lambda_1}{d\nu_1} \frac{d\lambda_2}{d\nu_2} \quad \text{a.e. } \nu_1 \times \nu_2,$$

where $\nu_1 \times \nu_2$ (or $\lambda_1 \times \lambda_2$) denotes the product measure of ν_1 and ν_2 (or λ_1 and λ_2).

Solution. Suppose that $A \in \sigma(\mathcal{F}_1 \times \mathcal{F}_2)$ and $\nu_1 \times \nu_2(A) = 0$. By Fubini's theorem,

$$0 = \nu_1 \times \nu_2(A) = \int I_A d(\nu_1 \times \nu_2) = \int \left(\int I_A d\nu_1 \right) d\nu_2.$$

Since $I_A \geq 0$, this implies that there is a $B \in \mathcal{F}_2$ such that $\nu_2(B^c) = 0$ and on the set B, $\int I_A d\nu_1 = 0$. Since $\lambda_1 \ll \nu_1$, on the set B

$$\int I_A d\lambda_1 = \int I_A \frac{d\lambda_1}{d\nu_1} d\nu_1 = 0.$$

Since $\lambda_2 \ll \nu_2$, $\lambda_2(B^c) = 0$. Then

$$\lambda_1 \times \lambda_2(A) = \int I_A d(\lambda_1 \times \lambda_2) = \int_B \left(\int_A d\lambda_1 \right) d\lambda_2 = 0.$$

Hence $\lambda_1 \times \lambda_2 \ll \nu_1 \times \nu_2$.

For the second assertion, it suffices to show that for any $A \in \sigma(\mathcal{F}_1 \times \mathcal{F}_2)$, $\lambda(A) = \nu(A)$, where

$$\lambda(A) = \int_A \frac{d(\lambda_1 \times \lambda_2)}{d(\nu_1 \times \nu_2)} d(\nu_1 \times \nu_2)$$

and

$$\nu(A) = \int_A \frac{d\lambda_1}{d\nu_1} \frac{d\lambda_2}{d\nu_2} d(\nu_1 \times \nu_2).$$

Let $\mathcal{C} = \mathcal{F}_1 \times \mathcal{F}_2$. Then $\mathcal{C}$ satisfies the conditions specified in Exercise 2. For $A_1 \times A_2 \in \mathcal{F}_1 \times \mathcal{F}_2$,

$$\begin{aligned}\lambda(A) &= \int_{A_1 \times A_2} \frac{d(\lambda_1 \times \lambda_2)}{d(\nu_1 \times \nu_2)} d(\nu_1 \times \nu_2) \\ &= \int_{A_1 \times A_2} d(\lambda_1 \times \lambda_2) \\ &= \lambda_1(A_1)\lambda_2(A_2)\end{aligned}$$

and, by Fubini's theorem,

$$\begin{aligned}\nu(A) &= \int_{A_1 \times A_2} \frac{d\lambda_1}{d\nu_1}\frac{d\lambda_2}{d\nu_2} d(\nu_1 \times \nu_2) \\ &= \int_{A_1} \frac{d\lambda_1}{d\nu_1} d\nu_1 \int_{A_2} \frac{d\lambda_2}{d\nu_2} d\nu_2 \\ &= \lambda_1(A_1)\lambda_2(A_2).\end{aligned}$$

Hence $\lambda(A) = \nu(A)$ for any $A \in \mathcal{C}$ and the second assertion of this exercise follows from the result in Exercise 2. ∎

Exercise 15. Let P and Q be two probability measures on a σ-field $\mathcal{F}$. Assume that $f = \frac{dP}{d\nu}$ and $g = \frac{dQ}{d\nu}$ exists for a measure ν on $\mathcal{F}$. Show that

$$\int |f - g| d\nu = 2 \sup\{|P(C) - Q(C)| : C \in \mathcal{F}\}.$$

Solution. Let $A = \{f \geq g\}$ and $B = \{f < g\}$. Then $A \in \mathcal{F}$, $B \in \mathcal{F}$, and

$$\begin{aligned}\int |f - g| d\nu &= \int_A (f - g) d\nu + \int_B (g - f) d\nu \\ &= P(A) - Q(A) + Q(B) - P(B) \\ &\leq |P(A) - Q(A)| + |P(B) - Q(B)| \\ &\leq 2 \sup\{|P(C) - Q(C)| : C \in \mathcal{F}\}.\end{aligned}$$

For any $C \in \mathcal{F}$,

$$\begin{aligned}P(C) - Q(C) &= \int_C (f - g) d\nu \\ &= \int_{C \cap A} (f - g) d\nu + \int_{C \cap B} (f - g) d\nu \\ &\leq \int_A (f - g) d\nu.\end{aligned}$$

Since

$$\int_C (f - g) d\nu + \int_{C^c} (f - g) d\nu = \int (f - g) d\nu = 1 - 1 = 0,$$

we have

$$\begin{aligned} P(C) - Q(C) &= \int_{C^c} (g-f)d\nu \\ &= \int_{C^c\cap A} (g-f)d\nu + \int_{C^c\cap B} (g-f)d\nu \\ &\le \int_B (g-f)d\nu. \end{aligned}$$

Hence

$$2[P(C) - Q(C)] \le \int_A (f-g)d\nu + \int_B (g-f)d\nu = \int |f-g|d\nu.$$

Similarly, $2[Q(C)-P(C)] \le \int |f-g|d\nu$. Thus, $2|P(C)-Q(C)| \le \int |f-g|d\nu$ and, consequently, $\int |f-g|d\nu \ge 2\sup\{|P(C)-Q(C)| : C \in \mathcal{F}\}$. ∎

Exercise 16 (#1.36). Let F_i be a cumulative distribution function on the real line having a Lebesgue density f_i, $i = 1, 2$. Assume that there is a real number c such that $F_1(c) < F_2(c)$. Define

$$F(x) = \begin{cases} F_1(x) & -\infty < x < c \\ F_2(x) & c \le x < \infty. \end{cases}$$

Show that the probability measure P corresponding to F satisfies $P \ll m + \delta_c$, where m is the Lebesgue measure and δ_c is the point mass at c, and find the probability density of F with respect to $m + \delta_c$.

Solution. For any $A \in \mathcal{B}$,

$$P(A) = \int_{(-\infty,c)\cap A} f_1(x)dm + a\int_{\{c\}\cap A} d\delta_c + \int_{(c,\infty)\cap A} f_2(x)dm,$$

where $a = F_2(c) - F_1(c)$. Note that $\int_{(-\infty,c)\cap A} d\delta_c = 0$, $\int_{(c,\infty)\cap A} d\delta_c = 0$, and $\int_{\{c\}\cap A} dm = 0$. Hence,

$$\begin{aligned} P(A) &= \int_{(-\infty,c)\cap A} f_1(x)d(m+\delta_c) + a\int_{\{c\}\cap A} d(m+\delta_c) \\ &\quad + \int_{(c,\infty)\cap A} f_2(x)d(m+\delta_c) \\ &= \int_A [I_{(-\infty,c)}(x)f_1(x) + aI_{\{c\}}(x) + I_{(c,\infty)}f_2(x)]d(m+\delta_c). \end{aligned}$$

This shows that $P \ll m + \delta_c$ and

$$\frac{dP}{d(m+\delta_c)} = I_{(-\infty,c)}(x)f_1(x) + aI_{\{c\}}(x) + I_{(c,\infty)}f_2(x).$$ ∎

Exercise 17 (#1.46). Let X_1 and X_2 be independent random variables having the standard normal distribution. Obtain the joint Lebesgue density of (Y_1, Y_2), where $Y_1 = \sqrt{X_1^2 + X_2^2}$ and $Y_2 = X_1/X_2$. Are Y_1 and Y_2 independent?

Note. For this type of problem, we may apply the following result. Let X be a random k-vector with a Lebesgue density f_X and let $Y = g(X)$, where g is a Borel function from $(\mathcal{R}^k, \mathcal{B}^k)$ to $(\mathcal{R}^k, \mathcal{B}^k)$. Let $A_1, ..., A_m$ be disjoint sets in $\mathcal{B}^k$ such that $\mathcal{R}^k - (A_1 \cup \cdots \cup A_m)$ has Lebesgue measure 0 and g on A_j is one-to-one with a nonvanishing Jacobian, i.e., the determinant $\text{Det}(\partial g(x)/\partial x) \neq 0$ on A_j, $j = 1, ..., m$. Then Y has the following Lebesgue density:

$$f_Y(x) = \sum_{j=1}^{m} \left| \text{Det}\left(\partial h_j(x)/\partial x\right) \right| f_X\left(h_j(x)\right),$$

where h_j is the inverse function of g on A_j, $j = 1, ..., m$.

Solution. Let $A_1 = \{(x_1, x_2)\colon x_1 > 0, x_2 > 0\}$, $A_2 = \{(x_1, x_2)\colon x_1 > 0, x_2 < 0\}$, $A_3 = \{(x_1, x_2)\colon x_1 < 0, x_2 > 0\}$, and $A_4 = \{(x_1, x_2)\colon x_1 < 0, x_2 < 0\}$. Then the Lebesgue measure of $\mathcal{R}^2 - (A_1 \cup A_2 \cup A_3 \cup A_4)$ is 0. On each A_i, the function $(y_1, y_2) = (\sqrt{x_1^2 + x_2^2}, x_1/x_2)$ is one-to-one with

$$\text{Det}\left(\frac{\partial(x_1, x_2)}{\partial(y_1, y_2)}\right) = \begin{vmatrix} \frac{y_2}{\sqrt{1+y_2^2}} & \frac{y_1}{\sqrt{1+y_2^2}} - \frac{y_1 y_2^2}{(1+y_2^2)^{3/2}} \\ \frac{1}{\sqrt{1+y_2^2}} & -\frac{y_1 y_2}{(1+y_2^2)^{3/2}} \end{vmatrix} = \frac{y_1}{1+y_2^2}.$$

Since the joint Lebesgue density of (X_1, X_2) is

$$\frac{1}{2\pi} e^{-(x_1^2 + x_2^2)/2}$$

and $x_1^2 + x_2^2 = y_1^2$, the joint Lebesgue density of (Y_1, Y_2) is

$$\sum_{i=1}^{4} \frac{1}{2\pi} e^{-(x_1^2+x_2^2)/2} \left| \text{Det}\left(\frac{\partial(x_1, x_2)}{\partial(y_1, y_2)}\right) \right| = \frac{2}{\pi} e^{-y_1^2} \frac{y_1}{1+y_2^2}.$$

Since the joint Lebesgue density of (Y_1, Y_2) is a product of two functions that are functions of one variable, Y_1 and Y_2 are independent. ∎

Exercise 18 (#1.45). Let X_i, $i = 1, 2, 3$, be independent random variables having the same Lebesgue density $f(x) = e^{-x} I_{(0,\infty)}(x)$. Obtain the joint Lebesgue density of (Y_1, Y_2, Y_3), where $Y_1 = X_1 + X_2 + X_3$, $Y_2 = X_1/(X_1 + X_2)$, and $Y_3 = (X_1 + X_2)/(X_1 + X_2 + X_3)$. Are Y_i's independent?

Solution: Let $x_1 = y_1 y_2 y_3$, $x_2 = y_1 y_3 - y_1 y_2 y_3$, and $x_3 = y_1 - y_1 y_3$. Then,

$$\text{Det}\left(\frac{\partial(x_1, x_2, x_3)}{\partial(y_1, y_2, y_3)}\right) = y_1^2 y_3.$$

Using the same argument as that in the previous exercise, we obtain the joint Lebesgue density of (Y_1, Y_2, Y_3) as

$$e^{-y_1} y_1^2 I_{(0,\infty)}(y_1) I_{(0,1)}(y_2) y_3 I_{(0,1)}(y_3).$$

Because this function is a product of three functions, $e^{-y_1} y_1^2 I_{(0,\infty)}(y_1)$, $I_{(0,1)}(y_2)$, and $y_3 I_{(0,1)}(y_3)$, Y_1, Y_2, and Y_3 are independent. ∎

Exercise 19 (#1.47). Let X and Y be independent random variables with cumulative distribution functions F_X and F_Y, respectively. Show that
(i) the cumulative distribution function of $X + Y$ is

$$F_{X+Y}(t) = \int F_Y(t-x) dF_X(x);$$

(ii) F_{X+Y} is continuous if one of F_X and F_Y is continuous;
(iii) $X+Y$ has a Lebesgue density if one of X and Y has a Lebesgue density.
Solution. (i) Note that

$$\begin{aligned} F_{X+Y}(t) &= \int_{x+y\le t} dF_X(x) dF_Y(y) \\ &= \int \left(\int_{y \le t-x} dF_Y(y) \right) dF_X(x) \\ &= \int F_Y(t-x) dF_X(x), \end{aligned}$$

where the second equality follows from Fubini's theorem.
(ii) Without loss of generality, we assume that F_Y is continuous. Since F_Y is bounded, by the dominated convergence theorem (e.g., Theorem 1.1 in Shao, 2003),

$$\begin{aligned} \lim_{\Delta t \to 0} F_{X+Y}(t+\Delta t) &= \lim_{\Delta t \to 0} \int F_Y(t+\Delta t - x) dF_X(x) \\ &= \int \lim_{\Delta t \to 0} F_Y(t+\Delta t - x) dF_X(x) \\ &= \int F_Y(t-x) dF_X(x) \\ &= F_{X+Y}(t). \end{aligned}$$

(iii) Without loss of generality, we assume that Y has a Lebesgue density f_Y. Then

$$\begin{aligned} F_{X+Y}(t) &= \int F_Y(t-x) dF_X(x) \\ &= \int \left(\int_{-\infty}^{t-x} f_Y(s) ds \right) dF_X(x) \end{aligned}$$

$$= \int \left(\int_{-\infty}^{t} f_Y(y-x)dy \right) dF_X(x)$$
$$= \int_{-\infty}^{t} \left(\int f_Y(y-x)dF_X(x) \right) dy,$$

where the last equality follows from Fubini's theorem. Hence, $X+Y$ has the Lebesgue density $f_{X+Y}(t) = \int f_Y(t-x)dF_X(x)$. ∎

Exercise 20 (#1.94). Show that a random variable X is independent of itself if and only if X is constant a.s. Can X and $f(X)$ be independent, where f is a Borel function?

Solution. Suppose that $X = c$ a.s. for a constant $c \in \mathcal{R}$. For any $A \in \mathcal{B}$ and $B \in \mathcal{B}$,

$$P(X \in A, X \in B) = I_A(c)I_B(c) = P(X \in A)P(X \in B).$$

Hence X and X are independent. Suppose now that X is independent of itself. Then, for any $t \in \mathcal{R}$,

$$P(X \le t) = P(X \le t, X \le t) = [P(X \le t)]^2.$$

This means that $P(X \le t)$ can only be 0 or 1. Since $\lim_{t\to\infty} P(X \le t) = 1$ and $\lim_{t\to-\infty} P(X \le t) = 0$, there must be a $c \in \mathcal{R}$ such that $P(X \le c) = 1$ and $P(X < c) = 0$. This shows that $X = c$ a.s.

If X and $f(X)$ are independent, then so are $f(X)$ and $f(X)$. From the previous result, this occurs if and only if $f(X)$ is constant a.s. ∎

Exercise 21 (#1.38). Let (X, Y, Z) be a random 3-vector with the following Lebesgue density:

$$f(x,y,z) = \begin{cases} \frac{1-\sin x \sin y \sin z}{8\pi^3} & 0 \le x, y, z, \le 2\pi \\ 0 & \text{otherwise} \end{cases}$$

Show that X, Y, Z are pairwise independent, but not independent.

Solution. The Lebesgue density for (X, Y) is

$$\int_0^{2\pi} f(x,y,z)dz = \int_0^{2\pi} \frac{1-\sin x \sin y \sin z}{8\pi^3} dz = \frac{1}{4\pi^2},$$

$0 \le x, y, \le 2\pi$. The Lebesgue density for X or Y is

$$\int_0^{2\pi} \int_0^{2\pi} f(x,y,z)dydz = \int_0^{2\pi} \frac{1}{4\pi^2} dy = \frac{1}{2\pi},$$

$0 \le x \le 2\pi$. Hence X and Y are independent. Similarly, X and Z are independent and Y and Z are independent. Note that

$$P(X \le \pi) = P(Y \le \pi) = P(Z \le \pi) = \int_0^{\pi} \frac{1}{2\pi} dx = \frac{1}{2}.$$

Hence $P(X \le \pi)P(Y \le \pi)P(Z \le \pi) = 1/8$. On the other hand,

$$\begin{aligned} P(X \le \pi, Y \le \pi, Z \le \pi) &= \int_0^\pi \int_0^\pi \int_0^\pi \frac{1 - \sin x \sin y \sin z}{8\pi^3} dxdydz \\ &= \frac{1}{8} - \frac{1}{8\pi^3} \left(\int_0^\pi \sin x dx \right)^3 \\ &= \frac{1}{8} - \frac{1}{\pi^3}. \end{aligned}$$

Hence X, Y, and Z are not independent. ∎

Exercise 22 (#1.51, #1.53). Let X be a random n-vector having the multivariate normal distribution $N_n(\mu, I_n)$.
(i) Apply Cochran's theorem to show that if $A^2 = A$, then $X^\tau AX$ has the noncentral chi-square distribution $\chi^2_r(\delta)$, where A is an $n \times n$ symmetric matrix, $r =$ rank of A, and $\delta = \mu^\tau A\mu$.
(ii) Let A_i be an $n \times n$ symmetric matrix satisfying $A_i^2 = A_i$, $i = 1, 2$. Show that a necessary and sufficient condition that $X^\tau A_1 X$ and $X^\tau A_2 X$ are independent is $A_1 A_2 = 0$.
Note. If $X_1, ..., X_k$ are independent and X_i has the normal distribution $N(\mu_i, \sigma^2)$, $i = 1, ..., k$, then the distribution of $(X_1^2 + \cdots + X_k^2)/\sigma^2$ is called the noncentral chi-square distribution $\chi^2_k(\delta)$, where $\delta = (\mu_1^2 + \cdots + \mu_k^2)/\sigma^2$. When $\delta = 0$, χ^2_k is called the central chi-square distribution.
Solution. (i) Since $A^2 = A$, i.e., A is a projection matrix,

$$(I_n - A)^2 = I_n - A - A + A^2 = I_n - A.$$

Hence, $I_n - A$ is a projection matrix with rank $\text{tr}(I_n - A) = \text{tr}(I_n) - \text{tr}(A) = n - r$. The result then follows by applying Cochran's theorem (e.g., Theorem 1.5 in Shao, 2003) to

$$X^\tau X = X^\tau AX + X^\tau (I_n - A)X.$$

(ii) Suppose that $A_1 A_2 = 0$. Then

$$\begin{aligned} (I_n - A_1 - A_2)^2 &= I_n - A_1 - A_2 - A_1 + A_1^2 + A_2 A_1 - A_2 + A_1 A_2 + A_2^2 \\ &= I_n - A_1 - A_2, \end{aligned}$$

i.e., $I_n - A_1 - A_2$ is a projection matrix with rank $= \text{tr}(I_n - A_1 - A_2) = n - r_1 - r_2$, where $r_i = \text{tr}(A_i)$ is the rank of A_i, $i = 1, 2$. By Cochran's theorem and

$$X^\tau X = X^\tau A_1 X + X^\tau A_2 X + X^\tau (I_n - A_1 - A_2)X,$$

$X^\tau A_1 X$ and $X^\tau A_2 X$ are independent.

Assume that $X^\tau A_1 X$ and $X^\tau A_2 X$ are independent. Since $X^\tau A_i X$ has the noncentral chi-square distribution $\chi^2_{r_i}(\delta_i)$, where r_i is the rank of A_i and $\delta_i = \mu^\tau A_i \mu$, $X^\tau(A_1 + A_2)X$ has the noncentral chi-square distribution $\chi^2_{r_1+r_2}(\delta_1 + \delta_2)$. Consequently, $A_1 + A_2$ is a projection matrix, i.e.,

$$(A_1 + A_2)^2 = A_1 + A_2,$$

which implies

$$A_1 A_2 + A_2 A_1 = 0.$$

Since $A_1^2 = A_1$, we obtain that

$$0 = A_1(A_1 A_2 + A_2 A_1) = A_1 A_2 + A_1 A_2 A_1$$

and

$$0 = A_1(A_1 A_2 + A_2 A_1)A_1 = 2A_1 A_2 A_1,$$

which imply $A_1 A_2 = 0$. ∎

Exercise 23 (#1.55). Let X be a random variable having a cumulative distribution function F. Show that if EX exists, then

$$EX = \int_0^\infty [1 - F(x)]dx - \int_{-\infty}^0 F(x)dx.$$

Solution. By Fubini's theorem,

$$\begin{aligned}\int_0^\infty [1 - F(x)]dx &= \int_0^\infty \int_{(x,\infty)} dF(y)dx \\ &= \int_0^\infty \int_{(0,y)} dx dF(y) \\ &= \int_0^\infty y dF(y).\end{aligned}$$

Similarly,

$$\int_{-\infty}^0 F(x)dx = \int_{-\infty}^0 \int_{(-\infty,x]} dF(y)dx = -\int_{-\infty}^0 y dF(y).$$

If EX exists, then at least one of $\int_0^\infty y dF(y)$ and $\int_{-\infty}^0 y dF(y)$ is finite and

$$EX = \int_{-\infty}^\infty y dF(y) = \int_0^\infty [1 - F(x)]dx - \int_{-\infty}^0 F(x)dx. \ ∎$$

Exercise 24 (#1.58(c)). Let X and Y be random variables having the bivariate normal distribution with $EX = EY = 0$, $\text{Var}(X) = \text{Var}(Y) = 1$,

and $\text{Cov}(X,Y) = \rho$. Show that $E(\max\{X,Y\}) = \sqrt{(1-\rho)/\pi}$.
Solution. Note that

$$|X - Y| = \max\{X,Y\} - \min\{X,Y\} = \max\{X,Y\} + \max\{-X,-Y\}.$$

Since the joint distribution of (X,Y) is symmetric about 0, the distribution of $\max\{X,Y\}$ and $\max\{-X,-Y\}$ are the same. Hence, $E|X-Y| = 2E(\max\{X,Y\})$. From the property of the normal distribution, $X - Y$ is normally distributed with mean 0 and variance $\text{Var}(X-Y) = \text{Var}(X) + \text{Var}(Y) - 2\text{Cov}(X,Y) = 2 - 2\rho$. Then,

$$E(\max\{X,Y\}) = 2^{-1}E|X-Y| = 2^{-1}\sqrt{2/\pi}\sqrt{2-2\rho} = \sqrt{(1-\rho)/\pi}. \quad \blacksquare$$

Exercise 25 (#1.60). Let X be a random variable with $EX^2 < \infty$ and let $Y = |X|$. Suppose that X has a Lebesgue density symmetric about 0. Show that X and Y are uncorrelated, but they are not independent.
Solution. Let f be the Lebesgue density of X. Then $f(x) = f(-x)$. Since X and $XY = X|X|$ are odd functions of X, $EX = 0$ and $E(X|X|) = 0$. Hence,

$$\text{Cov}(X,Y) = E(XY) - EXEY = E(X|X|) - EXE|X| = 0.$$

Let t be a positive constant such that $p = P(0 < X < t) > 0$. Then

$$\begin{aligned} P(0 < X < t, Y < t) &= P(0 < X < t, -t < X < t) \\ &= P(0 < X < t) \\ &= p \end{aligned}$$

and

$$\begin{aligned} P(0 < X < t)P(Y < t) &= P(0 < X < t)P(-t < X < t) \\ &= 2P(0 < X < t)P(0 < X < t) \\ &= 2p^2, \end{aligned}$$

i.e., $P(0 < X < t, Y < t) \neq P(0 < X < t)P(Y < t)$. Hence X and Y are not independent. $\blacksquare$

Exercise 26 (#1.61). Let (X,Y) be a random 2-vector with the following Lebesgue density:

$$f(x,y) = \begin{cases} \pi^{-1} & x^2 + y^2 \leq 1 \\ 0 & x^2 + y^2 > 1. \end{cases}$$

Show that X and Y are uncorrelated, but they are not independent.
Solution. Since X and Y are uniformly distributed on the Borel set

$\{(x, y) : x^2 + y^2 \leq 1\}$, $EX = EY = 0$ and $E(XY) = 0$. Hence $\text{Cov}(X, Y) = 0$. A direct calculation shows that

$$P(0 < X < 1/\sqrt{2}, 0 < Y < 1/\sqrt{2}) = \frac{1}{2\pi}$$

and

$$P(0 < X < 1/\sqrt{2}) = P(0 < Y < 1/\sqrt{2}) = \frac{1}{4} + \frac{1}{2\pi}.$$

Hence,

$$P(0 < X < 1/\sqrt{2}, 0 < Y < 1/\sqrt{2}) \neq P(0 < X < 1/\sqrt{2})P(0 < Y < 1/\sqrt{2})$$

and X and Y are not independent. ∎

Exercise 27 (#1.48, #1.70). Let Y be a random variable having the noncentral chi-square distribution $\chi_k^2(\delta)$, where k is a positive integer. Show that
(i) the Lebesgue density of Y is

$$g_{\delta,k}(t) = e^{-\delta/2} \sum_{j=0}^{\infty} \frac{(\delta/2)^j}{j!} f_{2j+k}(t),$$

where $f_j(t) = [\Gamma(j/2)2^{j/2}]^{-1} t^{j/2-1} e^{-t/2} I_{(0,\infty)}(t)$ is the Lebesgue density of the central chi-square distribution χ_j^2, $j = 1, 2, ...$;
(ii) the characteristic function of Y is $(1 - 2\sqrt{-1}t)^{-k/2} e^{\sqrt{-1}t/(1-2\sqrt{-1}t)}$;
(iii) $E(Y) = k + \delta$ and $\text{Var}(Y) = 2k + 4\delta$.

Solution A. (i) Consider first $k = 1$. By the definition of the noncentral chi-square distribution (e.g., Shao, 2003, p. 26), the distribution of Y is the same as that of X^2, where X has the normal distribution with mean $\sqrt{\delta}$ and variance 1. Since

$$P(Y \leq t) = P(X \leq \sqrt{t}) - P(X \leq -\sqrt{t})$$

for $t > 0$, the Lebesgue density of Y is

$$f_Y(t) = \frac{1}{2\sqrt{t}} [f_X(\sqrt{t}) + f_X(-\sqrt{t})] I_{(0,\infty)}(t),$$

where f_X is the Lebesgue density of X. Using the fact that X has a normal distribution, we obtain that, for $t > 0$,

$$\begin{aligned} f_Y(t) &= \frac{1}{2\sqrt{2\pi t}} \left(e^{-(\sqrt{t}-\sqrt{\delta})^2/2} + e^{-(-\sqrt{t}-\sqrt{\delta})^2/2} \right) \\ &= \frac{e^{-\delta/2} e^{-t/2}}{2\sqrt{2\pi t}} \left(e^{\sqrt{\delta t}} + e^{\sqrt{-\delta t}} \right) \end{aligned}$$

$$= \frac{e^{-\delta/2}e^{-t/2}}{2\sqrt{2\pi t}}\left(\sum_{j=0}^{\infty}\frac{(\sqrt{\delta t})^j}{j!} + \sum_{j=0}^{\infty}\frac{(-\sqrt{\delta t})^j}{j!}\right)$$

$$= \frac{e^{-\delta/2}e^{-t/2}}{\sqrt{2\pi t}}\sum_{j=0}^{\infty}\frac{(\delta t)^j}{(2j)!}.$$

On the other hand, for $k = 1$ and $t > 0$,

$$g_{\delta,1}(t) = e^{-\delta/2}\sum_{j=0}^{\infty}\frac{(\delta/2)^j}{j!}\left(\frac{1}{\Gamma(j+1/2)2^{j+1/2}}t^{j-1/2}e^{-t/2}\right)$$

$$= \frac{e^{-\delta/2}e^{-t/2}}{\sqrt{2t}}\sum_{j=0}^{\infty}\frac{(\delta t)^j}{j!\Gamma(j+1/2)2^{2j}}.$$

Since $j!2^{2j}\Gamma(j+1/2) = \sqrt{\pi}(2j)!$, $f_Y(t) = g_{\delta,1}(t)$ holds.

We then use induction. By definition, $Y = X_1 + X_2$, where X_1 has the noncentral chi-square distribution $\chi^2_{k-1}(\delta)$, X_2 has the central chi-square distribution χ^2_1, and X_1 and X_2 are independent. By the induction assumption, the Lebesgue density of X_1 is $g_{\delta,k-1}$. Note that the Lebesgue density of X_2 is f_1. Using the convolution formula (e.g., Example 1.15 in Shao, 2003), the Lebesgue density of Y is

$$f_Y(t) = \int g_{\delta,k-1}(u)f_1(t-u)du$$

$$= e^{-\delta/2}\sum_{j=0}^{\infty}\frac{(\delta/2)^j}{j!}\int f_{2j+k-1}(u)f_1(t-u)du$$

for $t > 0$. By the convolution formula again, $\int f_{2j+k-1}(u)f_1(t-u)du$ is the Lebesgue density of $Z+X_2$, where Z has density f_{2j+k-1} and is independent of X_2. By definition, $Z + X_2$ has the central chi-square distribution χ^2_{2j+k}, i.e.,

$$\int f_{2j+k-1}(u)f_1(t-u)du = f_{2j+k}(t).$$

Hence, $f_Y = g_{\delta,k}$.

(ii) Note that the moment generating function of the central chi-square distribution χ^2_k is, for $t < 1/2$,

$$\int e^{tu}f_k(u)du = \frac{1}{\Gamma(k/2)2^{k/2}}\int_0^{\infty} u^{k/2-1}e^{-(1-2t)u/2}du$$

$$= \frac{1}{\Gamma(k/2)2^{k/2}(1-2t)^{k/2}}\int_0^{\infty} s^{k/2-1}e^{-s/2}ds$$

$$= \frac{1}{(1-2t)^{k/2}},$$

where the second equality follows from the following change of variable in the integration: $s = (1-2t)u$. By the result in (i), the moment generating function for Y is

$$\begin{aligned}
\int e^{tx} g_{\delta,k}(x)dx &= e^{-\delta/2} \sum_{j=0}^{\infty} \frac{(\delta/2)^j}{j!} \int e^{tx} f_{2j+k}(x)dx \\
&= e^{-\delta/2} \sum_{j=0}^{\infty} \frac{(\delta/2)^j}{j!(1-2t)^{(j+k/2)}} \\
&= \frac{e^{-\delta/2}}{(1-2t)^{k/2}} \sum_{j=0}^{\infty} \frac{\{\delta/[2(1-2t)]\}^j}{j!} \\
&= \frac{e^{-\delta/2+\delta/[2(1-2t)]}}{(1-2t)^{k/2}} \\
&= \frac{e^{\delta t/(1-2t)}}{(1-2t)^{k/2}}.
\end{aligned}$$

Substituting t by $\sqrt{-1}t$ in the moment generating function of Y, we obtain the characteristic function of Y as $(1-2\sqrt{-1}t)^{-k/2} e^{\sqrt{-1}\delta t/(1-2\sqrt{-1}t)}$.
(iii) Let $\psi_Y(t)$ be the moment generating function of Y. By the result in (ii),

$$\psi'(t) = \psi(t)\left(\frac{\delta}{1-2t} + \frac{2\delta t}{(1-2t)^2} + \frac{k}{1-2t}\right)$$

and

$$\begin{aligned}
\psi''(t) = \psi'(t)&\left(\frac{\delta}{1-2t} + \frac{2\delta t}{(1-2t)^2} + \frac{k}{1-2t}\right) \\
&+ \psi(t)\left(\frac{4\delta}{(1-2t)^2} + \frac{2\delta t}{(1-2t)^3} + \frac{2k}{(1-2t)^2}\right).
\end{aligned}$$

Hence, $EY = \psi'(0) = \delta + k$, $EY^2 = \psi''(0) = (\delta+k)^2 + 4\delta + 2k$, and $\mathrm{Var}(Y) = EY^2 - (EY)^2 = 4\delta + 2k$.
Solution B. (i) We first derive result (ii). Let X be a random variable having the standard normal distribution and μ be a real number. The moment generating function of $(X+\mu)^2$ is

$$\begin{aligned}
\psi_\mu(t) &= \frac{1}{\sqrt{2\pi}} \int e^{-x^2/2} e^{t(x+\mu)^2} dx \\
&= \frac{e^{\mu^2 t/(1-2t)}}{\sqrt{2\pi}} \int e^{-(1-2t)[x-2\mu t/(1-2t)]^2/2} dx \\
&= \frac{e^{\mu^2 t/(1-2t)}}{\sqrt{1-2t}}.
\end{aligned}$$

By definition, Y has the same distribution as $X_1^2+\cdots+X_{k-1}^2+(X_k+\sqrt{\delta})^2$, where X_i's are independent and have the standard normal distribution. From the obtained result, the moment generating function of Y is

$$Ee^{t[X_1^2+\cdots+X_{k-1}^2+(X_k+\sqrt{\delta})^2]} = [\psi_0(t)]^{k-1}\psi_{\sqrt{\delta}}(t) = \frac{e^{\mu^2 t/(1-2t)}}{(1-2t)^{k/2}}.$$

(ii) We now use the result in (ii) to prove the result in (i). From part (ii) of Solution A, the moment generating function of $g_{\delta,k}$ is $e^{\mu^2 t/(1-2t)}(1-2t)^{-k/2}$, which is the same as the moment generating function of Y derived in part (i) of this solution. By the uniqueness theorem (e.g., Theorem 1.6 in Shao, 2003), we conclude that $g_{\delta,k}$ is the Lebesgue density of Y.
(iii) Let X_i's be as defined in (i). Then,

$$\begin{aligned} EY &= EX_1^2 + \cdots + EX_{k-1}^2 + E(X_k + \sqrt{\delta})^2 \\ &= k - 1 + EX_k^2 + \delta + E(2\sqrt{\delta}X_k) \\ &= k + \delta \end{aligned}$$

and

$$\begin{aligned} \mathrm{Var}(Y) &= \mathrm{Var}(X_1^2) + \cdots + \mathrm{Var}(X_{k-1}^2) + \mathrm{Var}\big((X_k + \sqrt{\delta})^2\big) \\ &= 2(k-1) + \mathrm{Var}(X_k^2 + 2\sqrt{\delta}X_k) \\ &= 2(k-1) + \mathrm{Var}(X_k^2) + \mathrm{Var}(2\sqrt{\delta}X_k) + 2\mathrm{Cov}(X_k^2, 2\sqrt{\delta}X_k) \\ &= 2k + 4\delta, \end{aligned}$$

since $\mathrm{Var}(X_i^2) = 2$ and $\mathrm{Cov}(X_k^2, X_k) = EX_k^3 - EX_k^2 EX_k = 0$. ∎

Exercise 28 (#1.57). Let U_1 and U_2 be independent random variables having the $\chi^2_{n_1}(\delta)$ and $\chi^2_{n_2}$ distributions, respectively, and let $F = (U_1/n_1)/(U_2/n_2)$. Show that
(i) $E(F) = \frac{n_2(n_1+\delta)}{n_1(n_2-2)}$ when $n_2 > 2$;
(ii) $\mathrm{Var}(F) = \frac{2n_2^2[(n_1+\delta)^2+(n_2-2)(n_1+2\delta)]}{n_1^2(n_2-2)^2(n_2-4)}$ when $n_2 > 4$.
Note. The distribution of F is called the noncentral F-distribution and denoted by $F_{n_1,n_2}(\delta)$.
Solution. From the previous exercise, $EU_1 = n_1+\delta$ and $EU_1^2 = \mathrm{Var}(U_1)+(EU_1)^2 = 2n_1 + 4\delta + (n_1 + \delta)^2$. Also,

$$\begin{aligned} EU_2^{-1} &= \frac{1}{\Gamma(n_2/2)2^{n_2/2}} \int_0^\infty x^{n_2/2-2} e^{-x/2} dx \\ &= \frac{\Gamma(n_2/2-1)2^{n_2/2-1}}{\Gamma(n_2/2)2^{n_2/2}} \\ &= \frac{1}{n_2-2} \end{aligned}$$

for $n_2 > 2$ and

$$\begin{aligned} EU_2^{-2} &= \frac{1}{\Gamma(n_2/2)2^{n_2/2}} \int_0^\infty x^{n_2/2-3} e^{-x/2} dx \\ &= \frac{\Gamma(n_2/2-2)2^{n_2/2-2}}{\Gamma(n_2/2)2^{n_2/2}} \\ &= \frac{1}{(n_2-2)(n_2-4)} \end{aligned}$$

for $n_2 > 4$. Then,

$$E(F) = E\frac{U_1/n_1}{U_2/n_2} = \frac{n_2}{n_1} EU_1 EU_2^{-1} = \frac{n_2(n_1+\delta)}{n_1(n_2-2)}$$

when $n_2 > 2$ and

$$\begin{aligned} \text{Var}(F) &= E\frac{U_1^2/n_1^2}{U_2^2/n_2^2} - [E(F)]^2 \\ &= \frac{n_2^2}{n_1^2} EU_1^2 EU_2^{-2} - \left(\frac{n_2(n_1+\delta)}{n_1(n_2-2)}\right)^2 \\ &= \frac{n_2^2}{n_1^2}\left(\frac{2n_1+4\delta+(N-1+\delta)^2}{(n_2-2)(n_2-4)} - \frac{(n_1+\delta)^2}{(n_2-2)^2}\right) \\ &= \frac{2n_2^2[(n_1+\delta)^2+(n_2-2)(n_1+2\delta)]}{n_1^2(n_2-2)^2(n_2-4)} \end{aligned}$$

when $n_2 > 4$. ▮

Exercise 29 (#1.74). Let ϕ_n be the characteristic function of a probability measure P_n, $n = 1, 2, ...$ Let $\{a_n\}$ be a sequence of nonnegative numbers with $\sum_{n=1}^\infty a_n = 1$. Show that $\sum_{n=1}^\infty a_n \phi_n$ is a characteristic function and find its corresponding probability measure.
Solution A. For any event A, define

$$P(A) = \sum_{n=1}^\infty a_n P_n(A).$$

Then P is a probability measure and $P_n \ll P$ for any n. Denote the Radon-Nikodym derivative of P_n with respect to P as f_n, $n = 1, 2, ...$ By Fubini's theorem, for any event A,

$$\begin{aligned} \int_A \sum_{n=1}^\infty a_n f_n dP &= \sum_{n=1}^\infty a_n \int_A f_n dP \\ &= \sum_{n=1}^\infty a_n P_n(A) \\ &= P(A). \end{aligned}$$

Hence, $\sum a_n f_n = 1$ a.s. P. Then,

$$\begin{aligned}\sum_{n=1}^{\infty} a_n \phi_n(t) &= \sum_{n=1}^{\infty} a_n \int e^{\sqrt{-1}tx} dP_n(x) \\ &= \sum_{n=1}^{\infty} a_n \int e^{\sqrt{-1}tx} f_n(x) dP \\ &= \int e^{\sqrt{-1}tx} \sum_{n=1}^{\infty} a_n f_n(x) dP \\ &= \int e^{\sqrt{-1}tx} dP.\end{aligned}$$

Hence, $\sum_{n=1}^{\infty} a_n \phi_n$ is the characteristic function of P.
Solution B. Let X be a discrete random variable satisfying $P(X = n) = a_n$ and Y be a random variable such that given $X = n$, the conditional distribution of Y is P_n, $n = 1, 2, \ldots$. The characteristic function of Y is

$$\begin{aligned}E(e^{\sqrt{-1}tY}) &= E[E(e^{\sqrt{-1}tY}|X)] \\ &= \sum_{n=1}^{\infty} a_n E(e^{\sqrt{-1}tY}|X = n) \\ &= \sum_{n=1}^{\infty} a_n \int e^{\sqrt{-1}ty} dP_n(y) \\ &= \sum_{n=1}^{\infty} a_n \phi_n(t).\end{aligned}$$

This shows that $\sum_{n=1}^{\infty} a_n \phi_n$ is the characteristic function of the marginal distribution of Y. ∎

Exercise 30 (#1.79). Find an example of two random variables X and Y such that X and Y are not independent but their characteristic functions ϕ_X and ϕ_Y satisfy $\phi_X(t)\phi_Y(t) = \phi_{X+Y}(t)$ for all $t \in \mathcal{R}$.
Solution. Let $X = Y$ be a random variable having the Cauchy distribution with $\phi_X(t) = \phi_Y(t) = e^{-|t|}$. Then X and Y are not independent (see Exercise 20). The characteristic function of $X + Y = 2X$ is

$$\phi_{X+Y}(t) = E(e^{\sqrt{-1}t(2X)}) = \phi_X(2t) = e^{-|2t|} = e^{-|t|}e^{-|t|} = \phi_X(t)\phi_Y(t). \quad ∎$$

Exercise 31 (#1.75). Let X be a random variable whose characteristic function ϕ satisfies $\int |\phi(t)|dt < \infty$. Show that $(2\pi)^{-1} \int e^{-\sqrt{-1}xt}\phi(t)dt$ is the Lebesgue density of X.

Solution. Define $g(t,x) = (e^{-\sqrt{-1}ta} - e^{-\sqrt{-1}tx})/(\sqrt{-1}t)$ for a fixed real number a. For any x, $|g(t,x)| \leq |x-a|$. Under the condition $\int |\phi(t)|dt < \infty$,

$$F(x) - F(a) = \frac{1}{2\pi}\int_{-\infty}^{\infty} \phi(t)g(t,x)dt$$

(e.g., Theorem 1.6 in Shao, 2003), where F is the cumulative distribution function of X. Since

$$\left|\frac{\partial g(t,x)}{\partial x}\right| = |e^{-\sqrt{-1}tx}| = 1,$$

by the dominated convergence theorem (Theorem 1.1 and Example 1.8 in Shao, 2003),

$$\begin{aligned} F'(x) &= \frac{d}{dx}\left(\frac{1}{2\pi}\int_{-\infty}^{\infty} \phi(t)g(t,x)dt\right) \\ &= \frac{1}{2\pi}\int_{-\infty}^{\infty} \phi(t)\frac{\partial g(t,x)}{\partial x}dt \\ &= \frac{1}{2\pi}\int_{-\infty}^{\infty} \phi(t)e^{-\sqrt{-1}tx}dt. \quad \blacksquare \end{aligned}$$

Exercise 32 (#1.73(g)). Let ϕ be a characteristic function and G be a cumulative distribution function on the real line. Show that $\int \phi(ut)dG(u)$ is a characteristic function on the real line.

Solution. Let F be the cumulative distribution function corresponding to ϕ and let X and U be independent random variables having distributions F and G, respectively. The characteristic function of UX is

$$\begin{aligned} Ee^{\sqrt{-1}tUX} &= \int\int e^{\sqrt{-1}tux}dF(x)dG(u) \\ &= \int \phi(ut)dG(u). \quad \blacksquare \end{aligned}$$

Exercise 33. Let X and Y be independent random variables. Show that if X and $X-Y$ are independent, then X must be degenerate.

Solution. We denote the characteristic function of any random variable Z by ϕ_Z. Since X and Y are independent, so are $-X$ and Y. Hence,

$$\phi_{Y-X}(t) = \phi_Y(t)\phi_{-X}(t) = \phi_Y(t)\phi_X(-t), \quad t \in \mathcal{R}.$$

If X and $X-Y$ are independent, then X and $Y-X$ are independent. Then

$$\phi_Y(t) = \phi_{X+(Y-X)}(t) = \phi_X(t)\phi_{Y-X}(t) = \phi_X(t)\phi_X(-t)\phi_Y(t), \quad t \in \mathcal{R}.$$

Since $\phi_Y(0) = 1$ and ϕ_Y is continuous, $\phi_Y(t) \neq 0$ for a neighborhood of 0. Hence $\phi_X(t)\phi_X(-t) = |\phi_X(t)|^2 = 1$ on this neighborhood of 0. Thus, X is degenerate. ∎

Exercise 34 (#1.98). Let P_Y be a discrete distribution on $\{0, 1, 2, ...\}$ and given $Y = y$, the conditional distribution of X be the binomial distribution with size y and probability p. Show that
(i) if Y has the Poisson distribution with mean θ, then the marginal distribution of X is the Poisson distribution with mean $p\theta$;
(ii) if $Y + r$ has the negative binomial distribution with size r and probability π, then the marginal distribution of $X + r$ is the negative binomial distribution with size r and probability $\pi/[1 - (1 - p)(1 - \pi)]$.
Solution. (i) The moment generating function of X is

$$E(e^{tX}) = E[E(e^{tX}|Y)] = E[(pe^t + 1 - p)^Y] = e^{\theta p(e^t - 1)},$$

which is the moment generating function of the Poisson distribution with mean $p\theta$.
(ii) The moment generating function of $X + r$ is

$$\begin{aligned} E(e^{t(X+r)}) &= e^{tr}E[E(e^{tX}|Y)] \\ &= e^{tr}E[(pe^t + 1 - p)^Y] \\ &= \frac{e^{tr}}{(pe^t + 1 - p)^r}E[(pe^t + 1 - p)^{Y+r}] \\ &= \frac{e^{tr}}{(pe^t + 1 - p)^r}\frac{\pi^r(pe^t + 1 - p)^r}{[1 - (1 - \pi)(pe^t + 1 - p)]^r} \\ &= \frac{\pi^r e^{rt}}{[1 - (1 - \pi)(pe^t + 1 - p)]^r}. \end{aligned}$$

Then the result follows from the fact that

$$\frac{1 - (1 - \pi)(pe^t + 1 - p)}{1 - (1 - p)(1 - \pi)} = 1 - \left[1 - \frac{\pi}{1 - (1 - p)(1 - \pi)}\right]e^t. \quad ∎$$

Exercise 35 (#1.85). Let X and Y be integrable random variables on the probability space $(\Omega, \mathcal{F}, P)$ and $\mathcal{A}$ be a sub-σ-field of $\mathcal{F}$. Show that
(i) if $X \leq Y$ a.s., then $E(X|\mathcal{A}) \leq E(Y|\mathcal{A})$ a.s.;
(ii) if a and b are constants, then $E(aX + bY|\mathcal{A}) = aE(X|\mathcal{A}) + bE(X|\mathcal{A})$ a.s.
Solution. (i) Suppose that $X \leq Y$ a.s. By the definition of the conditional expectation and the property of integration,

$$\int_A E(X|\mathcal{A})dP = \int_A XdP \leq \int_A YdP = \int_A E(Y|\mathcal{A})dP,$$

where

$$A = \{E(X|\mathcal{A}) > E(Y|\mathcal{A})\} \in \mathcal{A}.$$

Hence $P(A) = 0$, i.e., $E(X|\mathcal{A}) \leq E(Y|\mathcal{A})$ a.s.
(ii) Note that $aE(X|\mathcal{A}) + bE(Y|\mathcal{A})$ is measurable from $(\Omega, \mathcal{A})$ to $(\mathcal{R}, \mathcal{B})$. For any $A \in \mathcal{A}$, by the linearity of integration,

$$\begin{aligned} \int_A (aX + bY)dP &= a \int_A XdP + b \int_A YdP \\ &= a \int_A E(X|\mathcal{A})dP + b \int_A E(Y|\mathcal{A})dP \\ &= \int_A [aE(X|\mathcal{A}) + bE(Y|\mathcal{A})]dP. \end{aligned}$$

By the a.s.-uniqueness of the conditional expectation, $E(aX + bY|\mathcal{A}) = aE(X|\mathcal{A}) + bE(X|\mathcal{A})$ a.s. ∎

Exercise 36 (#1.85). Let X be an integrable random variable on the probability space $(\Omega, \mathcal{F}, P)$ and $\mathcal{A}$ and $\mathcal{A}_0$ be σ-fields satisfying $\mathcal{A}_0 \subset \mathcal{A} \subset \mathcal{F}$. Show that $E[E(X|\mathcal{A})|\mathcal{A}_0] = E(X|\mathcal{A}_0) = E[E(X|\mathcal{A}_0)|\mathcal{A}]$ a.s.
Solution. Note that $E(X|\mathcal{A}_0)$ is measurable from $(\Omega, \mathcal{A}_0)$ to $(\mathcal{R}, \mathcal{B})$ and $\mathcal{A}_0 \subset \mathcal{A}$. Hence $E(X|\mathcal{A}_0)$ is measurable from $(\Omega, \mathcal{A})$ to $(\mathcal{R}, \mathcal{B})$ and, thus, $E(X|\mathcal{A}_0) = E[E(X|\mathcal{A}_0)|\mathcal{A}]$ a.s. Since $E[E(X|\mathcal{A})|\mathcal{A}_0]$ is measurable from $(\Omega, \mathcal{A}_0)$ to $(\mathcal{R}, \mathcal{B})$ and for any $A \in \mathcal{A}_0 \subset \mathcal{A}$,

$$\int_A E[E(X|\mathcal{A})|\mathcal{A}_0]dP = \int_A E(X|\mathcal{A})dP = \int_A XdP,$$

we conclude that $E[E(X|\mathcal{A})|\mathcal{A}_0] = E(X|\mathcal{A}_0)$ a.s. ∎

Exercise 37 (#1.85). Let X be an integrable random variable on the probability space $(\Omega, \mathcal{F}, P)$, $\mathcal{A}$ be a sub-σ-field of $\mathcal{F}$, and Y be another random variable satisfying $\sigma(Y) \subset \mathcal{A}$ and $E|XY| < \infty$. Show that

$$E(XY|\mathcal{A}) = YE(X|\mathcal{A}) \quad \text{a.s.}$$

Solution. Since $\sigma(Y) \subset \mathcal{A}$, $YE(X|\mathcal{A})$ is measurable from $(\Omega, \mathcal{A})$ to $(\mathcal{R}, \mathcal{B})$. The result follows if we can show that for any $A \in \mathcal{A}$,

$$\int_A YE(X|\mathcal{A})dP = \int_A XYdP.$$

(1) If $Y = aI_B$, where $a \in \mathcal{R}$ and $B \in \mathcal{A}$, then $A \cap B \in \mathcal{A}$ and

$$\int_A XYdP = a \int_{A\cap B} XdP = a \int_{A\cap B} E(X|\mathcal{A})dP = \int_A YE(X|\mathcal{A})dP.$$

(2) If $Y = \sum_{i=1}^k a_i I_{B_i}$, where $B_i \in \mathcal{A}$, then

$$\int_A XYdP = \sum_{i=1}^k a_i \int_A XI_{B_i}dP = \sum_{i=1}^k a_i \int_A I_{B_i}E(X|\mathcal{A})dP = \int_A YE(X|\mathcal{A})dP.$$

(3) Suppose that $X \geq 0$ and $Y \geq 0$. There exists a sequence of increasing simple functions Y_n such that $\sigma(Y_n) \subset \mathcal{A}$, $Y_n \leq Y$ and $\lim_n Y_n = Y$. Then $\lim_n XY_n = XY$ and $\lim_n Y_n E(X|\mathcal{A}) = YE(X|\mathcal{A})$. By the monotone convergence theorem and the result in (2),

$$\int_A XYdP = \lim_n \int_A XY_n dP = \lim_n \int_A Y_n E(X|\mathcal{A})dP = \int_A YE(X|\mathcal{A})dP.$$

(4) For general X and Y, consider X_+, X_-, Y_+, and Y_-. Since $\sigma(Y) \subset \mathcal{A}$, so are $\sigma(Y_+)$ and $\sigma(Y_-)$. Then, by the result in (3),

$$\begin{aligned}
\int_A XYdP &= \int_A X_+Y_+dP - \int_A X_+Y_-dP \\
&\quad - \int_A X_-Y_+dP + \int_A X_-Y_-dP \\
&= \int_A Y_+E(X_+|\mathcal{A})dP - \int_A Y_-E(X_+|\mathcal{A})dP \\
&\quad - \int_A Y_+E(X_-|\mathcal{A})dP + \int_A Y_-E(X_-|\mathcal{A})dP \\
&= \int_A YE(X_+|\mathcal{A})dP - \int_A YE(X_-|\mathcal{A})dP \\
&= \int_A YE(X|\mathcal{A})dP,
\end{aligned}$$

where the last equality follows from the result in Exercise 35. ▮

Exercise 38 (#1.85). Let $X_1, X_2, ...$ and X be integrable random variables on the probability space $(\Omega, \mathcal{F}, P)$. Assume that $0 \leq X_1 \leq X_2 \leq \cdots \leq X$ and $\lim_n X_n = X$ a.s. Show that for any σ-field $\mathcal{A} \subset \mathcal{F}$,

$$E(X|\mathcal{A}) = \lim_n E(X_n|\mathcal{A}) \quad \text{a.s.}$$

Solution. Since each $E(X_n|\mathcal{A})$ is measurable from $(\Omega, \mathcal{A})$ to $(\mathcal{R}, \mathcal{B})$, so is the limit $\lim_n E(X_n|\mathcal{A})$. We need to show that

$$\int_A \lim_n E(X_n|\mathcal{A})dP = \int_A XdP$$

for any $A \in \mathcal{A}$. By Exercise 35, $0 \leq E(X_1|\mathcal{A}) \leq E(X_2|\mathcal{A}) \leq \cdots \leq E(X|\mathcal{A})$ a.s. By the monotone convergence theorem (e.g., Theorem 1.1 in Shao, 2003), for any $A \in \mathcal{A}$,

$$\begin{aligned}\int_A \lim_n E(X_n|\mathcal{A})dP &= \lim_n \int_A E(X_n|\mathcal{A})dP \\ &= \lim_n \int_A X_n dP \\ &= \int_A \lim_n X_n dP \\ &= \int_A X dP. \quad \blacksquare\end{aligned}$$

Exercise 39 (#1.85). Let $X_1, X_2, ...$ be integrable random variables on the probability space $(\Omega, \mathcal{F}, P)$. Show that for any σ-field $\mathcal{A} \subset \mathcal{F}$,
(i) $E(\liminf_n X_n|\mathcal{A}) \leq \liminf_n E(X_n|\mathcal{A})$ a.s. if $X_n \geq 0$ for any n;
(ii) $\lim_n E(X_n|\mathcal{A}) = E(X|\mathcal{A})$ a.s. if $\lim_n X_n = X$ a.s. and $|X_n| \leq Y$ for any n and an integrable random variable Y.
Solution. (i) For any $m \geq n$, by Exercise 35, $E(\inf_{m\geq n} X_m|\mathcal{A}) \leq E(X_m|\mathcal{A})$ a.s. Hence, $E(\inf_{m\geq n} X_m|\mathcal{A}) \leq \inf_{m\geq n} E(X_m|\mathcal{A})$ a.s. Let $Y_n = \inf_{m\geq n} X_m$. Then $0 \leq Y_1 \leq Y_2 \leq \cdots \leq \lim_n Y_n$ and Y_n's are integrable. Hence,

$$\begin{aligned}E(\liminf_n X_n|\mathcal{A}) &= E(\lim_n Y_n|\mathcal{A}) \\ &= \lim_n E(Y_n|\mathcal{A}) \\ &= \lim_n E(\inf\nolimits_{m\geq n} X_m|\mathcal{A}) \\ &\leq \lim_n \inf_{m\geq n} E(X_m|\mathcal{A}) \\ &= \liminf_n E(X_n|\mathcal{A})\end{aligned}$$

a.s., where the second equality follows from the result in the previous exercise and the first and the last equalities follow from the fact that $\liminf_n f_n = \lim_n \inf_{m\geq n} f_m$ for any sequence of functions $\{f_n\}$.
(ii) Note that $Y + X_n \geq 0$ for any n. Applying the result in (i) to $Y + X_n$,

$$\liminf_n E(Y + X_n|\mathcal{A}) \leq E(\liminf_n (Y + X_n)|\mathcal{A}) = E(Y + X|\mathcal{A}) \quad \text{a.s.}$$

Since Y is integrable, so is X and, consequently, $E(Y + X|\mathcal{A}) = E(Y|\mathcal{A}) + E(X|\mathcal{A})$ a.s. and $\liminf_n E(Y + X_n|\mathcal{A}) = E(Y|\mathcal{A}) + \liminf_n E(X_n|\mathcal{A})$ a.s. Hence,

$$\liminf_n E(X_n|\mathcal{A}) \leq E(X|\mathcal{A}) \quad \text{a.s.}$$

Applying the same argument to $Y - X_n$, we obtain that

$$\liminf_n E(-X_n|\mathcal{A}) \leq E(-X|\mathcal{A}) \quad \text{a.s.}$$

Since $\liminf_n E(-X_n|\mathcal{A}) = -\limsup_n E(X_n|\mathcal{A})$, we obtain that

$$\limsup_n E(X_n|\mathcal{A}) \geq E(X|\mathcal{A}) \quad \text{a.s.}$$

Combining the results, we obtain that

$$\limsup_n E(X_n|\mathcal{A}) = \liminf_n E(X_n|\mathcal{A}) = \lim_n E(X_n|\mathcal{A}) = E(X|\mathcal{A}) \quad \text{a.s.} \quad \blacksquare$$

Exercise 40 (#1.86). Let X and Y be integrable random variables on the probability space $(\Omega, \mathcal{F}, P)$ and $\mathcal{A} \subset \mathcal{F}$ be a σ-field. Show that $E[YE(X|\mathcal{A})] = E[XE(Y|\mathcal{A})]$, assuming that both integrals exist.

Solution. (1) The problem is much easier if we assume that Y is bounded. When Y is bounded, both $YE(X|\mathcal{A})$ and $XE(Y|\mathcal{A})$ are integrable. Using the result in Exercise 37 and the fact that $E[E(X|\mathcal{A})] = EX$, we obtain that

$$\begin{aligned} E[YE(X|\mathcal{A})] &= E\{E[YE(X|\mathcal{A})|\mathcal{A}]\} \\ &= E[E(X|\mathcal{A})E(Y|\mathcal{A})] \\ &= E\{E[XE(Y|\mathcal{A})|\mathcal{A}]\} \\ &= E[XE(Y|\mathcal{A})]. \end{aligned}$$

(2) Assume that $Y \geq 0$. Let Z be another nonnegative integrable random variable. We now show that if $\sigma(Z) \subset \mathcal{A}$, then $E(YZ) = E[ZE(Y|\mathcal{A})]$. (Note that this is a special case of the result in Exercise 37 if $E(YZ) < \infty$.)

Let $Y_n = \max\{Y, n\}$, $n = 1, 2, \ldots.$ Then $0 \leq Y_1 \leq Y_2 \leq \cdots \leq Y$ and $\lim_n Y_n = Y$. By the results in Exercises 35 and 39, $0 \leq E(Y_1|\mathcal{A}) \leq E(Y_2|\mathcal{A}) \leq \cdots$ a.s. and $\lim_n E(Y_n|\mathcal{A}) = E(Y|\mathcal{A})$ a.s. Since Y_n is bounded, $Y_n Z$ is integrable. By the result in Exercise 37,

$$E[ZE(Y_n|\mathcal{A})] = E(Y_n Z), \quad n = 1, 2, \ldots.$$

By the monotone convergence theorem,

$$E(YZ) = \lim_n E(Y_n Z) = \lim_n E[ZE(Y_n|\mathcal{A})] = E[ZE(Y|\mathcal{A})].$$

Consequently, if $X \geq 0$, then the result follows by taking $Z = E(X|\mathcal{A})$.

(3) We now consider general X and Y. Let f_+ and f_- denote the positive and negative parts of a function f. Note that

$$E\{[XE(Y|\mathcal{A})]_+\} = E\{X_+[E(Y|\mathcal{A})]_+\} + E\{X_-[E(Y|\mathcal{A})]_-\}$$

and

$$E\{[XE(Y|\mathcal{A})]_-\} = E\{X_+[E(Y|\mathcal{A})]_-\} + E\{X_-[E(Y|\mathcal{A})]_+\}.$$

Since $E[XE(Y|\mathcal{A})]$ exists, without loss of generality we assume that

$$E\{[XE(Y|\mathcal{A})]_+\} = E\{X_+[E(Y|\mathcal{A})]_+\} + E\{X_-[E(Y|\mathcal{A})]_-\} < \infty.$$

Then, both

$$E[X_+E(Y|\mathcal{A})] = E\{X_+[E(Y|\mathcal{A})]_+\} - E\{X_+[E(Y|\mathcal{A})]_-\}$$

and

$$E[X_-E(Y|\mathcal{A})] = E\{X_-[E(Y|\mathcal{A})]_+\} - E\{X_-[E(Y|\mathcal{A})]_-\}$$

are well defined and their difference is also well defined. Applying the result established in (2), we obtain that

$$\begin{aligned} E[X_+E(Y|\mathcal{A})] &= E\{E(X_+|\mathcal{A})[E(Y|\mathcal{A})]_+\} - E\{E(X_+|\mathcal{A})[E(Y|\mathcal{A})]_-\} \\ &= E[E(X_+|\mathcal{A})E(Y|\mathcal{A})], \end{aligned}$$

where the last equality follows from the result in Exercise 8. Similarly,

$$\begin{aligned} E[X_-E(Y|\mathcal{A})] &= E\{E(X_-|\mathcal{A})[E(Y|\mathcal{A})]_+\} - E\{E(X_-|\mathcal{A})[E(Y|\mathcal{A})]_-\} \\ &= E[E(X_-|\mathcal{A})E(Y|\mathcal{A})]. \end{aligned}$$

By Exercise 8 again,

$$\begin{aligned} E[XE(Y|\mathcal{A})] &= E[X_+E(Y|\mathcal{A})] - E[X_-E(Y|\mathcal{A})] \\ &= E[E(X_+|\mathcal{A})E(Y|\mathcal{A})] - E[E(X_-|\mathcal{A})E(Y|\mathcal{A})] \\ &= E[E(X|\mathcal{A})E(Y|\mathcal{A})]\}. \end{aligned}$$

Switching X and Y, we also conclude that

$$E[YE(X|\mathcal{A})] = E[E(X|\mathcal{A})E(Y|\mathcal{A})].$$

Hence, $E[XE(Y|\mathcal{A})] = E[YE(X|\mathcal{A})]$. ∎

Exercise 41 (#1.87). Let $X, X_1, X_2, ...$ be a sequence of integrable random variables on the probability space $(\Omega, \mathcal{F}, P)$ and $\mathcal{A} \subset \mathcal{F}$ be a σ-field. Suppose that $\lim_n E(X_nY) = E(XY)$ for every integrable (or bounded) random variable Y. Show that $\lim_n E[E(X_n|\mathcal{A})Y] = E[E(X|\mathcal{A})Y]$ for every integrable (or bounded) random variable Y.

Solution. Assume that Y is integrable. Then $E(Y|\mathcal{A})$ is integrable. By the condition, $E[X_nE(Y|\mathcal{A})] \to E[XE(Y|\mathcal{A})]$. By the result of the previous exercise, $E[X_nE(Y|\mathcal{A})] = E[E(X_n|\mathcal{A})Y]$ and $E[XE(Y|\mathcal{A})] = E[E(X|\mathcal{A})Y]$. Hence, $E[E(X_n|\mathcal{A})Y] \to E[E(X|\mathcal{A})Y]$ for every integrable Y. The same result holds if "integrable" is changed to "bounded". ∎

Exercise 42 (#1.88). Let X be a nonnegative integrable random variable on the probability space $(\Omega, \mathcal{F}, P)$ and $\mathcal{A} \subset \mathcal{F}$ be a σ-field. Show that

$$E(X|\mathcal{A}) = \int_0^\infty P\big(X > t|\mathcal{A}\big)dt.$$

Note. For any $B \in \mathcal{F}$, $P(B|\mathcal{A})$ is defined to be $E(I_B|\mathcal{A})$.
Solution. From the theory of conditional distribution (e.g., Theorem 1.7 in Shao, 2003), there exists $\tilde{P}(B, \omega)$ defined on $\mathcal{F} \times \Omega$ such that (i) for any $\omega \in \Omega$, $\tilde{P}(\cdot, \omega)$ is a probability measure on $(\Omega, \mathcal{F})$ and (ii) for any $B \in \mathcal{F}$, $\tilde{P}(B, \omega) = P(B|\mathcal{A})$ a.s. From Exercise 23,

$$\begin{aligned} \int X d\tilde{P}(\cdot, \omega) &= \int_0^\infty \tilde{P}(\{X > t\}, \omega)dt \\ &= \int_0^\infty P(X > t|\mathcal{A})dt \quad \text{a.s.} \end{aligned}$$

Hence, the result follows if

$$E(X|\mathcal{A})(\omega) = \int X d\tilde{P}(, \omega) \quad \text{a.s.}$$

This is certainly true if $X = I_B$ for a $B \in \mathcal{F}$. By the linearity of the integration and conditional expectation, this equality also holds when X is a nonnegative simple function. For general nonnegative X, there exists a sequence of simple functions $X_1, X_2, ...,$ such that $0 \leq X_1 \leq X_2 \leq \cdots \leq X$ and $\lim_n X_n = X$ a.s. From Exercise 38,

$$\begin{aligned} E(X|\mathcal{A}) &= \lim_n E(X_n|\mathcal{A}) \\ &= \lim_n \int X_n d\tilde{P}(\cdot, \omega) \\ &= \int X d\tilde{P}(\cdot, \omega) \quad \text{a.s.} \ ▮ \end{aligned}$$

Exercise 43 (#1.97). Let X and Y be independent integrable random variables on a probability space and f be a nonnegative convex function. Show that $E[f(X + Y)] \geq E[f(X + EY)]$.
Note. We need to apply the following Jensen's inequality for conditional expectations. Let f be a convex function and X be an integrable random variable satisfying $E|f(X)| < \infty$. Then $f(E(X|\mathcal{A})) \leq E(f(X)|\mathcal{A})$ a.s. (e.g., Theorem 9.1.4 in Chung, 1974).
Solution. If $E[f(X + Y)] = \infty$, then the inequality holds. Hence, we may assume that $f(X + Y)$ is integrable. Using Jensen's inequality and some

properties of conditional expectations, we obtain that

$$\begin{aligned} E[f(X+Y)] &= E\{E[f(X+Y)|X]\} \\ &\geq E\{f(E(X+Y|X))\} \\ &= E\{f(X+E(Y|X))\} \\ &= E[f(X+EY)], \end{aligned}$$

where the last equality follows from $E(Y|X) = EY$ since X and Y are independent. ∎

Exercise 44 (#1.83). Let X be an integrable random variable with a Lebesgue density f and let $Y = g(X)$, where g is a function with positive derivative on $(0,\infty)$ and $g(x) = g(-x)$. Find an expression for $E(X|Y)$ and verify that it is indeed the conditional expectation.

Solution. Let h be the inverse function of g on $(0,\infty)$ and

$$\psi(y) = h(y)\frac{f(h(y)) - f(-h(y))}{f(h(y)) + f(-h(y))}.$$

We now show that $E(X|Y) = \psi(Y)$ a.s. It is clear that $\psi(y)$ is a Borel function. Also, the σ-field generated by Y is generated by the sets of the form $A_a = \{y : g(0) \leq y \leq a\}$, $a > g(0)$. Hence, it suffices to show that for any $a > g(0)$,

$$\int_{A_a} XdP = \int_{A_a} \psi(Y)dP.$$

Note that

$$\begin{aligned} \int_{A_a} XdP &= \int_{g(0)\leq g(x)\leq a} xf(x)dx \\ &= \int_{-h(a)}^{h(a)} xf(x)dx \\ &= \int_{-h(a)}^{0} xf(x)dx + \int_{0}^{h(a)} xf(x)dx \\ &= \int_{h(a)}^{0} xf(-x)dx + \int_{0}^{h(a)} xf(x)dx \\ &= \int_{0}^{h(a)} x[f(x) - f(-x)]dx \\ &= \int_{g(0)}^{a} h(y)[f(h(y)) - f(-h(y))]h'(y)dy. \end{aligned}$$

On the other hand, $h'(y)[f(h(y)) + f(-h(y))]I_{(g(0),\infty)}(y)$ is the Lebesgue density of Y (see the note in Exercise 17). Hence,

$$\begin{aligned}\int_{A_a} \psi(Y)dP &= \int_{g(0)}^{a} \psi(y)h'(y)[f(h(y)) + f(-h(y))]dy \\ &= \int_{g(0)}^{a} h(y)[f(h(y)) - f(-h(y))]h'(y)dy\end{aligned}$$

by the definition of $\psi(y)$. ∎

Exercise 45 (#1.91). Let X, Y, and Z be random variables on a probability space. Suppose that $E|X| < \infty$ and $Y = h(Z)$ with a Borel h. Show that
(i) $E(XZ|Y) = E(X)E(Z|Y)$ a.s. if X and Z are independent and $E|Z| < \infty$;
(ii) if $E[f(X)|Z] = f(Y)$ for all bounded continuous functions f on $\mathcal{R}$, then $X = Y$ a.s.;
(iii) if $E[f(X)|Z] \geq f(Y)$ for all bounded, continuous, nondecreasing functions f on $\mathcal{R}$, then $X \geq Y$ a.s.
Solution. (i) It suffices to show

$$\int_{Y^{-1}(B)} XZdP = E(X) \int_{Y^{-1}(B)} ZdP$$

for any Borel set B. Since $Y = h(Z)$, $Y^{-1}(B) = Z^{-1}(h^{-1}(B))$. Then

$$\int_{Y^{-1}(B)} XZdP = \int XZI_{h^{-1}(B)}(Z)dP = E(X) \int ZI_{h^{-1}(B)}(Z)dP,$$

since X and Z are independent. On the other hand,

$$\int_{Y^{-1}(B)} ZdP = \int_{h^{-1}(B)} ZdP = \int ZI_{h^{-1}(B)}(Z)dP.$$

(ii) Let $f(t) = e^t/(1+e^t)$. Then both f and f^2 are bounded and continuous. Note that

$$\begin{aligned}E[f(X) - f(Y)]^2 &= EE\{[f(X) - f(Y)]^2|Z\} \\ &= E\{E[f^2(X)|Z] + E[f^2(Y)|Z] - 2E[f(X)f(Y)|Z]\} \\ &= E\{E[f^2(X)|Z] + f^2(Y) - 2f(Y)E[f(X)|Z]\} \\ &= E\{f^2(Y) + f^2(Y) - 2f(Y)f(Y)\} \\ &= 0,\end{aligned}$$

where the third equality follows from the result in Exercise 37 and the fourth equality follows from the condition. Hence $f(X) = f(Y)$ a.s. Since

f is strictly increasing, $X = Y$ a.s.
(iii) For any real number c, there exists a sequence of bounded, continuous and nondecreasing functions $\{f_n\}$ such that $\lim_n f_n(t) = I_{(c,\infty)}(t)$ for any real number t. Then,

$$\begin{aligned} P(X > c, Y > c) &= E\{E(I_{\{X>c\}} I_{\{Y>c\}}|Z)\} \\ &= E\{I_{\{Y>c\}} E(I_{\{X>c\}}|Z)\} \\ &= E\{I_{\{Y>c\}} E[\lim_n f_n(X)|Z]\} \\ &= E\{I_{\{Y>c\}} \lim_n E[f_n(X)|Z]\} \\ &\geq E\{I_{\{Y>c\}} \lim_n f_n(Y)\} \\ &= E\{I_{\{Y>c\}} I_{\{Y>c\}}\} \\ &= P(Y > c), \end{aligned}$$

where the fourth and fifth equalities follow from Exercise 39 (since f_n is bounded) and the inequality follows from the condition. This implies that $P(X \leq c, Y > c) = P(Y > c) - P(X > c, Y > c) = 0$. For any integer k and positive integer n, let $a_{k,i} = k + i/n$, $i = 1, ..., n$. Then

$$P(X < Y) = \lim_n \sum_{k=-\infty}^{\infty} \sum_{i=0}^{n-1} P(X \leq a_{k,i}, a_{k,i} < Y \leq a_{k,i+1}) = 0.$$

Hence, $X \geq Y$ a.s. ∎

Exercise 46 (#1.115). Let $X_1, X_2, ...$ be a sequence of identically distributed random variables with $E|X_1| < \infty$ and let $Y_n = n^{-1} \max_{1\leq i\leq n} |X_i|$. Show that $\lim_n E(Y_n) = 0$ and $\lim_n Y_n = 0$ a.s.
Solution. (i) Let $g_n(t) = n^{-1} P(\max_{1\leq i\leq n} |X_i| > t)$. Then $\lim_n g_n(t) = 0$ for any t and

$$0 \leq g_n \leq \frac{1}{n} \sum_{i=1}^{n} P(|X_i| > t) = P(|X_1| > t).$$

Since $E|X_1| < \infty$, $\int_0^\infty P(|X_1| > t)dt < \infty$ (Exercise 23). By the dominated convergence theorem,

$$\lim_n E(Y_n) = \lim_n \int_0^\infty g_n(t)dt = \int_0^\infty \lim_n g_n(t)dt = 0.$$

(ii) Since $E|X_1| < \infty$,

$$\sum_{n=1}^{\infty} P(|X_n|/n > \epsilon) = \sum_{n=1}^{\infty} P(|X_1| > \epsilon n) < \infty,$$

which implies that $\lim_n |X_n|/n = 0$ a.s. (see, e.g., Theorem 1.8(v) in Shao, 2003). Let $\Omega^0 = \{\omega : \lim_n |X_n(\omega)|/n = 0\}$. Then $P(\Omega^0) = 1$. Let $\omega \in \Omega^0$. For any $\epsilon > 0$, there exists an $N_{\epsilon,\omega}$ such that $|X_n(\omega)| < n\epsilon$ whenever $n > N_{\epsilon,\omega}$. Also, there exists an $M_{\epsilon,\omega} > N_{\epsilon,\omega}$ such that $\max_{1\leq i\leq N_{\epsilon\omega}} |X_i(\omega)| \leq n\epsilon$ whenever $n > M_{\epsilon,\omega}$. Then, whenever $n > M_{\epsilon,\omega}$,

$$\begin{aligned} Y_n(\omega) &= \frac{\max_{1\leq i\leq n} |X_i(\omega)|}{n} \\ &\leq \frac{\max_{1\leq i\leq N_{\epsilon\omega}} |X_i(\omega)|}{n} + \frac{\max_{N_{\epsilon\omega}<i\leq n} |X_i(\omega)|}{n} \\ &\leq \epsilon + \max_{N_{\epsilon\omega}<i\leq n} \frac{|X_i(\omega)|}{i} \\ &\leq 2\epsilon, \end{aligned}$$

i.e., $\lim_n Y_n(\omega) = 0$. Hence, $\lim_n Y_n = 0$ a.s., since $P(\Omega^0) = 1$. ∎

Exercise 47 (#1.116). Let $X, X_1, X_2, ...$ be random variables. Find an example for each of the following cases:
(i) $X_n \to_p X$, but $\{X_n\}$ does not converge to X a.s.;
(ii) $X_n \to_p X$, but $E|X_n - X|^p$ does not converge for any $p > 0$;
(iii) $X_n \to_d X$, but $\{X_n\}$ does not converge to X in probability;
(iv) $X_n \to_p X$, but $g(X_n)$ does not converge to $g(X)$ in probability for some function g;
(v) $\lim_n E|X_n| = 0$, but $|X_n|$ cannot be bounded by any integrable function.
Solution: Consider the probability space $([0,1], \mathcal{B}_{[0,1]}, P)$, where $\mathcal{B}_{[0,1]}$ is the Borel σ-field and P is the Lebesgue measure on $[0,1]$.
(i) Let $X = 0$. For any positive integer n, there exist integers m and k such that $n = 2^m - 2 + k$ and $0 \leq k < 2^{m+1}$. Define

$$X_n(\omega) = \begin{cases} 1 & k/2^m \leq \omega \leq k + 1/2^m \\ 0 & \text{otherwise} \end{cases}$$

for any $\omega \in [0,1]$. Note that

$$P(|X_n - X| > \epsilon) \leq P(\{\omega : k/2^m \leq \omega \leq (k+1)/2^m\}) = \frac{1}{2^m} \to 0$$

as $n \to \infty$ for any $\epsilon > 0$. Thus $X_n \to_p X$. However, for any fixed $\omega \in [0,1]$ and m, there exists k with $1 \leq k \leq 2^m$ such that $(k-1)/2^m \leq \omega \leq k/2^m$. Let $n_m = 2^m - 2 + k$. Then $X_{n_m}(\omega) = 1$. Since m is arbitrarily selected, we can find an infinite sequence $\{n_m\}$ such that $X_{n_m}(\omega) = 1$. This implies $X_n(\omega)$ does not converge to $X(\omega) = 0$. Since ω is arbitrary, X_n does not converge to X a.s.
(ii) Let $X = 0$ and

$$X_n(\omega) = \begin{cases} 0 & 1/n < \omega \leq 1 \\ e^n & 0 \leq \omega \leq 1/n. \end{cases}$$

For any $\epsilon \in (0,1)$

$$P(|X_n - X| > \epsilon) = P(|X_n| \neq 0) = \frac{1}{n} \to 0$$

as $n \to \infty$, i.e., $X_n \to_p X$. On the other hand, for any $p > 0$,

$$E|X_n - X|^p = E|X_n|^p = e^{np}/n \to \infty.$$

(iii) Define

$$X(\omega) = \begin{cases} 1 & 0 \leq \omega \leq 1/2 \\ 0 & 1/2 < \omega \leq 1 \end{cases}$$

and

$$X_n(\omega) = \begin{cases} 0 & 0 \leq \omega \leq 1/2 \\ 1 & 1/2 < \omega \leq 1. \end{cases}$$

For any t,

$$P(X \leq t) = P(X_n \leq t) = \begin{cases} 1 & t \geq 1 \\ 1/2 & 0 \leq t < 1 \\ 0 & t < 0, \end{cases}$$

Therefore, $X_n \to_d X$. However, $|X_n - X| = 1$ and thus $P(|X_n - X| > \epsilon) = 1$ for any $\epsilon \in (0,1)$.
(iv) let $g(t) = 1 - I_{\{0\}}(t)$, $X = 0$, and $X_n = 1/n$. Then, $X_n \to_p X$, but $g(X_n) = 1$ and $g(X) = 0$.
(v) Define

$$X_{n,m}(\omega) = \begin{cases} \sqrt{n} & \frac{m-1}{n} < \omega \leq \frac{m}{n} \\ 0 & \text{otherwise,} \end{cases} \qquad m = 1, ..., n, \ n = 1, 2, ...$$

Then,

$$E|X_{n,m}| = \int_{(m-1)/n}^{m/n} \sqrt{n} dx = \frac{1}{\sqrt{n}} \to 0$$

as $n \to \infty$. Hence, the sequence $\{X_{n,m} : m = 1, ..., n, n = 1, 2, ...\}$ satisfies the requirement. If there is a function f such that $|X_{n,m}| \leq f$, then $f(\omega) = \infty$ for any $\omega \in [0,1]$. Hence, f cannot be integrable. ∎

Exercise 48. Let X_n be a random variable and m_n be a median of X_n, $n = 1, 2, ...$ Show that if $X_n \to_d X$ for a random variable X, then any limit point of m_n is a median of X.
Solution. Without loss of generality, assume that $\lim_n m_n = m$. For $\epsilon > 0$ such that $m + \epsilon$ and $m - \epsilon$ are continuity points of the distribution of X, $m - \epsilon < m_n < m + \epsilon$ for sufficiently large n and

$$\frac{1}{2} \leq P(X_n \leq m_n) \leq P(X_n \leq m + \epsilon)$$

and

$$\frac{1}{2} \leq P(X_n \geq m_n) \leq P(X_n \geq m - \epsilon).$$

Letting $n \to \infty$, we obtain that $\frac{1}{2} \leq P(X \leq m+\epsilon)$ and $\frac{1}{2} \leq P(X \geq m-\epsilon)$. Letting $\epsilon \to 0$, we obtain that $\frac{1}{2} \leq P(X \leq m)$ and $\frac{1}{2} \leq P(X \geq m)$. Hence m is a median of X. ∎

Exercise 49 (#1.126). Show that if $X_n \to_d X$ and $X = c$ a.s. for a real number c, then $X_n \to_p X$.
Solution. Note that the cumulative distribution function of X has only one discontinuity point c. For any $\epsilon > 0$,

$$\begin{aligned} P(|X_n - X| > \epsilon) &= P(|X_n - c| > \epsilon) \\ &\leq P(X_n > c + \epsilon) + P(X_n \geq c - \epsilon) \\ &\to P(X > c + \epsilon) + P(X \leq c - \epsilon) \\ &= 0 \end{aligned}$$

as $n \to \infty$. Thus, $X_n \to_p X$. ∎

Exercise 50 (#1.117(b), #1.118). Let $X_1, X_2, ...$ be random variables. Show that $\{|X_n|\}$ is uniformly integrable if one of the following condition holds:
(i) $\sup_n E|X_n|^{1+\delta} < \infty$ for a $\delta > 0$;
(ii) $P(|X_n| \geq c) \leq P(|X| \geq c)$ for all n and $c > 0$, where X is an integrable random variable.
Note. A sequence of random variables $\{X_n\}$ is uniformly integrable if $\lim_{t\to\infty} \sup_n E(|X_n| I_{\{|X_n|>t\}}) = 0$.
Solution. (i) Denote $p = 1 + \delta$ and $q = 1 + \delta^{-1}$. Then

$$\begin{aligned} E(|X_n| I_{\{|X_n|>t\}}) &\leq (E|X_n|^p)^{1/p} [E(I_{\{|X_n|>t\}})^q]^{1/q} \\ &= (E|X_n|^p)^{1/p} [P(|X_n| > t)]^{1/q} \\ &\leq (E|X_n|^p)^{1/p} (E|X_n|^p)^{1/q} t^{-p/q} \\ &= E|X_n|^{1+\delta} t^{-\delta}, \end{aligned}$$

where the first inequality follows from Hölder's inequality (e.g., Shao, 2003, p. 29) and the second inequality follows from

$$P(|X_n| > t) \leq t^{-p} E|X_n|^p.$$

Hence

$$\lim_{t\to\infty} \sup_n E(|X_n| I_{\{|X_n|>t\}}) \leq \sup_n E|X_n|^{1+\delta} \lim_{t\to\infty} t^{-\delta} = 0.$$

(ii) By Exercise 23,

$$\begin{aligned}
\sup_n E(|X_n| I_{\{|X_n|>t\}}) &= \sup_n \int_0^\infty P(|X_n| I_{\{|X_n|>t\}} > s) ds \\
&= \sup_n \int_0^\infty P(|X_n| > s, |X_n| > t) ds \\
&= \sup_n \left(tP(|X_n| > t) + \int_t^\infty P(|X_n| > s) ds \right) \\
&\le tP(|X| > t) + \int_t^\infty P(|X| > s) ds \\
&\to 0
\end{aligned}$$

as $t \to \infty$ when $E|X| < \infty$. ▮

Exercise 51. Let $\{X_n\}$ and $\{Y_n\}$ be sequences of random variables such that X_n diverges to ∞ in probability and Y_n is bounded in probability. Show that $X_n + Y_n$ diverges to ∞ in probability.

Solution. By the definition of bounded in probability, for any $\epsilon > 0$, there is $C_\epsilon > 0$ such that $\sup_n P(|Y_n| > C_\epsilon) < \epsilon/2$. By the definition of divergence to ∞ in probability, for any $M > 0$ and $\epsilon > 0$, there is $n_\epsilon > 0$ such that $P(|X_n| \le M + C_\epsilon) < \epsilon/2$ whenever $n > n_\epsilon$. Then, for $n > n_\epsilon$,

$$\begin{aligned}
P(|X_n + Y_n| \le M) &\le P(|X_n| \le M + |Y_n|) \\
&= P(|X_n| \le M + |Y_n|, |Y_n| \le C_\epsilon) \\
&\quad + P(|X_n| \le M + |Y_n|, |Y_n| > C_\epsilon) \\
&\le P(|X_n| \le M + C_\epsilon) + P(|Y_n| > C_\epsilon) \\
&\le \epsilon/2 + \epsilon/2 \\
&= \epsilon.
\end{aligned}$$

This means that $X_n + Y_n$ diverges to ∞ in probability. ▮

Exercise 52. Let $X, X_1, X_2, ...$ be random variables. Show that if $\lim_n X_n = X$ a.s., then $\sup_{m\ge n} |X_m|$ is bounded in probability.

Solution. Since $\sup_{m\ge n} |X_m| \le \sup_{m\ge 1} |X_m|$ for any n, it suffices to show that for any $\epsilon > 0$, there is a $C > 0$ such that $P(\sup_{n\ge 1} |X_n| > C) \le \epsilon$. Note that $\lim_n X_n = X$ implies that, for any $\epsilon > 0$ and any fixed $c_1 > 0$, there exists a sufficiently large N such that $P\left(\cup_{n=N}^\infty \{|X_n - X| > c_1\}\right) < \epsilon/3$ (e.g., Lemma 1.4 in Shao, 2003). For this fixed N, there exist constants $c_2 > 0$ and $c_3 > 0$ such that

$$\sum_{n=1}^N P(|X_n| > c_2) < \frac{\epsilon}{3} \quad \text{and} \quad P(|X| > c_3) < \frac{\epsilon}{3}.$$

Let $C = \max\{c_1, c_2\} + c_3$. Then the result follows from

$$\begin{aligned}
P\left(\sup_{n\geq 1}|X_n| > C\right) &= P\left(\bigcup_{n=1}^{\infty}\{|X_n| > C\}\right) \\
&\leq \sum_{n=1}^{N} P(|X_n| > C) + P\left(\bigcup_{n=N}^{\infty}\{|X_n| > C\}\right) \\
&\leq \frac{\epsilon}{3} + P(|X| > c_3) + P\left(\bigcup_{n=N}^{\infty}\{|X_n| > C, |X| \leq c_3\}\right) \\
&\leq \frac{\epsilon}{3} + \frac{\epsilon}{3} + P\left(\bigcup_{n=N}^{\infty}\{|X_n - X| > c_1\}\right) \\
&\leq \frac{\epsilon}{3} + \frac{\epsilon}{3} + \frac{\epsilon}{3} = \epsilon. \ \blacksquare
\end{aligned}$$

Exercise 53 (#1.128). Let $\{X_n\}$ and $\{Y_n\}$ be two sequences of random variables such that X_n is bounded in probability and, for any real number t and $\epsilon > 0$, $\lim_n[P(X_n \leq t, Y_n \geq t+\epsilon) + P(X_n \geq t+\epsilon, Y_n \leq t)] = 0$. Show that $X_n - Y_n \to_p 0$.

Solution. For any $\epsilon > 0$, there exists an $M > 0$ such that $P(|X_n| \geq M) \leq \epsilon$ for any n, since X_n is bounded in probability. For this fixed M, there exists an N such that $2M/N < \epsilon/2$. Let $t_i = -M + 2Mi/N$, $i = 0, 1, ..., N$. Then,

$$\begin{aligned}
P(|X_n - Y_n| \geq \epsilon) &\leq P(|X_n| \geq M) + P(|X_n| < M, |X_n - Y_n| \geq \epsilon) \\
&\leq \epsilon + \sum_{i=1}^{N} P(t_{i-1} \leq X_n \leq t_i, |X_n - Y_n| \geq \epsilon) \\
&\leq \epsilon + \sum_{i=1}^{N} P(Y_n \leq t_{i-1} - \epsilon/2, t_{i-1} \leq X_n) \\
&\quad + \sum_{i=1}^{N} P(Y_n \geq t_i + \epsilon/2, X_n \leq t_i).
\end{aligned}$$

This, together with the given condition, implies that

$$\limsup_n P(|X_n - Y_n| \geq \epsilon) \leq \epsilon.$$

Since ϵ is arbitrary, we conclude that $X_n - Y_n \to_p 0$. ∎

Exercise 54 (#1.133). Let F_n, $n = 0, 1, 2, ...$, be cumulative distribution functions such that $F_n \to F_0$ for every continuity point of F_0. Let U be a random variable having the uniform distribution on the interval $[0, 1]$ and let

$G_n(U) = \sup\{x : F_n(x) \leq U\}$, $n = 0, 1, 2, ...$ Show that $G_n(U) \to_p G_0(U)$.
Solution. For any n and real number t, $G_n(U) \leq t$ if and only if $F_n(t) \geq U$ a.s. Similarly, $G_n(U) \geq t$ if and only if $F_n(t) \leq U$ a.s. Hence, for any n, t and $\epsilon > 0$,

$$\begin{aligned} P\left(G_n(U) \leq t, G_0(U) \geq t + \epsilon\right) &= P\left(F_n(t) \geq U, F_0(t+\epsilon) \leq U\right) \\ &= \max\{0, F_n(t) - F_0(t+\epsilon)\} \end{aligned}$$

and

$$\begin{aligned} P\left(G_n(U) \geq t + \epsilon, G_0(U) \leq t\right) &= P\left(F_n(t+\epsilon) \leq U, F_0(t) \leq U\right) \\ &= \max\{0, F_0(t) - F_n(t+\epsilon)\}. \end{aligned}$$

If both t and $t+\epsilon$ are continuity points of F_0, then $\lim_n[F_n(t) - F_0(t+\epsilon)] = F_0(t) - F_0(t+\epsilon) \leq 0$ and $\lim_n[F_0(t) - F_n(t+\epsilon)] = F_0(t) - F_0(t+\epsilon) \leq 0$. Hence,

$$\lim_n \left[P\left(G_n(U) \leq t, G_0(U) \geq t+\epsilon\right) + P\left(G_n(U) \geq t+\epsilon, G_0(U) \leq t\right)\right] = 0$$

when both t and $t+\epsilon$ are continuity points of F_0. Since the set of discontinuity points of F_0 is countable, $G_n(U) - G_0(U) \to_p 0$ follows from the result in the previous exercise, since $G_0(U)$ is obviously bounded in probability. ∎

Exercise 55. Let $\{X_n\}$ be a sequence of independent and identically distributed random variables. Show that there does not exist a sequence of real numbers $\{c_n\}$ such that $\lim_n \sum_{i=1}^n (X_i - c_i)$ exists a.s., unless the distribution of X_1 is degenerate.
Solution. Suppose that $\lim_n \sum_{i=1}^n (X_i - c_i)$ exists a.s. Let ϕ and g be the characteristic functions of X_1 and $\lim_n \sum_{i=1}^n (X_i - c_i)$, respectively. For any n, the characteristic function of $\sum_{i=1}^n (X_i - c_i)$ is

$$\prod_{i=1}^n \phi(t) e^{-\sqrt{-1}tc_i} = [\phi(t)]^n e^{-\sqrt{-1}t(c_1+\cdots+c_n)},$$

which converges to $g(t)$ for any t. Then

$$\lim_n \left| [\phi(t)]^n e^{-\sqrt{-1}t(c_1+\cdots+c_n)} \right| = \lim_n |\phi(t)|^n = |g(t)|.$$

Since $|g(t)|$ is continuous and $g(0) = 1$, $|g(t)| \neq 0$ on a neighborhood of 0. Hence, $|\phi(t)| = 1$ on this neighborhood and, thus, X_1 is degenerate. ∎

Exercise 56. Let $P, P_1, P_2, ...$ be probability measures such that $\lim_n P_n(O) = P(O)$ for any open set O with $P(\partial O) = 0$, where ∂A is the boundary of the set A. Show that $\lim_n P_n(A) = P(A)$ for any Borel A with $P(\partial A) = 0$.

Solution. Let A be a Borel set with $P(\partial A) = 0$. Let A_0 be the interior of A and A_1 be the closure of A. Then $A_0 \subset A \subset A_1 = A_0 \cup \partial A$. Since $\partial A_0 \subset \partial A$, $P(\partial A_0) = 0$ and, by assumption, $\lim_n P_n(A_0) = P(A_0)$. Since $\partial A_1^c \subset \partial A$ and A_1^c is an open set, $\lim_n P_n(A_1^c) = P(A_1^c)$, which implies $\lim_n P_n(A_1) = P(A_1)$. Then,

$$P(A) \leq P(A_0) + P(\partial A) = P(A_0) = \lim_n P_n(A_0) \leq \liminf_n P_n(A)$$

and

$$P(A) \geq P(A_0) = P(A_0) + P(\partial A) = P(A_1) = \lim_n P_n(A_1) \geq \limsup_n P_n(A).$$

Hence $\liminf_n P_n(A) = \limsup_n P_n(A) = \lim_n P_n(A) = P(A)$. ∎

Exercise 57. Let $X, X_1, X_2, ...$ be random variables such that, for any continuous cumulative distribution function F, $\lim_n E[F(X_n)] = E[F(X)]$. Show that $X_n \to_d X$.

Solution. Let y be a continuity point of the cumulative distribution function of X. Define

$$F_m(x) = \begin{cases} 0 & x \leq y - m^{-1} \\ mx + 1 - my & y - m^{-1} < x \leq y \\ 1 & x > y \end{cases}$$

and

$$H_m(x) = \begin{cases} 0 & x \leq y \\ mx - my & y < x \leq y + m^{-1} \\ 1 & x > y + m^{-1}. \end{cases}$$

Then, F_m, H_m, $m = 1, 2, ...$, are continuous cumulative distribution functions and $\lim_m F_m(x) = I_{(-\infty,y]}(x)$ and $\lim_m H_m(x) = I_{(-\infty,y]}(x)$, since y is a continuity point of the cumulative distribution function of X. By the dominated convergence theorem,

$$\lim_m E[F_m(X)] = \lim_m E[H_m(X)] = E[I_{(-\infty,y]}(X)] = P(X \leq y).$$

Since $F_m(x)$ decreases as m increases,

$$E[F_m(X_n)] \geq E[I_{(-\infty,y]}(X_n)] = P(X_n \leq y).$$

By the assumption, $\lim_n E[F_m(X_n)] = E[F_m(X)]$ for any m. Hence,

$$E[F_m(X)] \geq \limsup_n P(X_n \leq y)$$

for any m. Letting $m \to \infty$, we obtain that

$$P(X \leq y) \geq \limsup_n P(X_n \leq y).$$

Similarly,

$$E[H_m(X_n)] \le E[I_{(-\infty,y]}(X_n)] = P(X_n \le y)$$

and

$$\liminf_n P(X_n \le y) \le \lim_m \lim_n E[H_m(X_n)] = \lim_m E[H_m(X)] = P(X \le y).$$

Hence $\lim_n P(X_n \le y) = P(X \le y)$. ∎

Exercise 58 (#1.137). Let $\{X_n\}$ and $\{Y_n\}$ be two sequences of random variables. Suppose that $X_n \to_d X$ and that, for almost every given sequence $\{X_n\}$, the conditional distribution of Y_n given X_n converges to the distribution of Y at every continuity point of the distribution of Y, where X and Y are independent random variables. Show that $X_n + Y_n \to_d X + Y$.
Solution. From the assumed conditions and the continuity theorem (e.g., Theorem 1.9 in Shao, 2003), for any real number t, $\lim_n E(e^{\sqrt{-1}tX_n}) = E(e^{\sqrt{-1}tX})$ and $\lim_n E(e^{\sqrt{-1}tY_n}|X_n) = E(e^{\sqrt{-1}tY})$ a.s. By the dominated convergence theorem,

$$\lim_n E\{e^{\sqrt{-1}tX_n}[E(e^{\sqrt{-1}tY_n}|X_n) - E(e^{\sqrt{-1}tY})]\} = 0.$$

Then

$$\begin{aligned}
\lim_n E[e^{\sqrt{-1}t(X_n+Y_n)}] &= \lim_n E[E(e^{\sqrt{-1}t(X_n+Y_n)}|X_n)] \\
&= \lim_n E[e^{\sqrt{-1}tX_n} E(e^{\sqrt{-1}tY_n}|X_n)] \\
&= \lim_n E\{e^{\sqrt{-1}tX_n}[E(e^{\sqrt{-1}tY_n}|X_n) - E(e^{\sqrt{-1}tY})]\} \\
&\quad + \lim_n E(e^{\sqrt{-1}tY})E(e^{\sqrt{-1}tX_n}) \\
&= E(e^{\sqrt{-1}tY})E(e^{\sqrt{-1}tX}) \\
&= E[e^{\sqrt{-1}t(X+Y)}].
\end{aligned}$$

By the continuity theorem again, $X_n + Y_n \to_d X + Y$. ∎

Exercise 59 (#1.140). Let X_n be a random variable distributed as $N(\mu_n, \sigma_n^2)$, $n = 1, 2, ...$, and X be a random variable distributed as $N(\mu, \sigma^2)$. Show that $X_n \to_d X$ if and only if $\lim_n \mu_n = \mu$ and $\lim_n \sigma_n^2 = \sigma^2$.
Solution. The characteristic function of X is $\phi_X(t) = e^{\sqrt{-1}\mu t - \sigma^2 t^2/2}$ and the characteristic function of X_n is $\phi_{X_n}(t) = e^{\sqrt{-1}\mu_n t - \sigma_n^2 t^2/2}$. If $\lim_n \mu_n = \mu$ and $\lim_n \sigma_n^2 = \sigma^2$, then $\lim_n \phi_{X_n}(t) = \phi_X(t)$ for any t and, by the continuity theorem, $X_n \to_d X$.

Assume now $X_n \to_d X$. By the continuity theorem, $\lim_n \phi_{X_n}(t) = \phi_X(t)$ for any t. Then $\lim_n |\phi_{X_n}(t)| = |\phi_X(t)|$ for any t. Since $|\phi_{X_n}(t)| =$

$e^{-\sigma_n^2 t^2/2}$ and $|\phi_X(t)| = e^{-\sigma^2 t^2/2}$, $\lim_n \sigma_n^2 = \sigma^2$. Note that

$$\lim_n e^{\sqrt{-1}\mu_n t} = \lim_n \frac{\phi_{X_n}(t)}{|\phi_{X_n}(t)|} = \frac{\phi_X(t)}{|\phi_X(t)|} = e^{\sqrt{-1}\mu t}.$$

Hence $\lim_n \mu_n = \mu$. ∎

Exercise 60 (#1.146). Let $U_1, U_2, ...$ be independent random variables having the uniform distribution on $[0, 1]$ and $Y_n = \left(\prod_{i=1}^n U_i\right)^{-1/n}$. Show that $\sqrt{n}(Y_n - e) \to_d N(0, e^2)$.

Solution. Let $X_i = -\log U_i$. Then $X_1, X_2, ...$ are independent and identically distributed random variables with $EX_1 = 1$ and $\text{Var}(X_1) = 1$. By the central limit theorem, $\sqrt{n}(\bar{X}_n - 1) \to_d N(0, 1)$, where $\bar{X}_n = n^{-1}\sum_{i=1}^n X_i$. Note that $Y_n = e^{\bar{X}_n}$. Applying the δ-method with $g(t) = e^t$ to $\bar{X}_n$ (e.g., Theorem 1.12 in Shao, 2003), we obtain that $\sqrt{n}(Y_n - e) \to_d N(0, e^2)$, since $g'(0) = 1$. ∎

Exercise 61 (#1.161). Suppose that X_n is a random variable having the binomial distribution with size n and probability $\theta \in (0, 1)$, $n = 1, 2, ...$. Define $Y_n = \log(X_n/n)$ when $X_n \geq 1$ and $Y_n = 1$ when $X_n = 0$. Show that $\lim_n Y_n = \log\theta$ a.s. and $\sqrt{n}(Y_n - \log\theta) \to_d N\left(0, \frac{1-\theta}{\theta}\right)$.

Solution. (i) Let $Z_1, Z_2, ...$ be independent and identically distributed random variables with $P(Z_1 = 1) = \theta$ and $P(Z_1 = 0) = 1 - \theta$. Then the distribution of X_n is the same as that of $\sum_{j=1}^n Z_j$. For any $\epsilon > 0$,

$$\begin{aligned} P\left(\left|\frac{X_n}{n} - \theta\right| \geq \epsilon\right) &\leq \frac{1}{\epsilon^4} E\left|\frac{X_n}{n} - \theta\right|^4 \\ &= \frac{\theta^4(1-\theta) + (1-\theta)^4\theta}{\epsilon^4 n^3} + \frac{\theta^2(1-\theta)^2(n-1)}{\epsilon^4 n^3}. \end{aligned}$$

Hence,

$$\sum_{n=1}^{\infty} P\left(\left|\frac{X_n}{n} - \theta\right| \geq \epsilon\right) < \infty$$

and, by Theorem 1.8(v) in Shao (2003), $\lim_n X_n/n = \theta$ a.s.

Define $W_n = I_{\{X_n \neq 0\}} X_n/n$. Then $Y_n = \log(W_n + eI_{\{X_n = 0\}})$. Note that

$$\sum_{n=1}^{\infty} P(\sqrt{n} I_{\{X_n = 0\}} > \epsilon) = \sum_{n=1}^{\infty} P(X_n = 0) = \sum_{n=1}^{\infty} (1-\theta)^n < \infty.$$

Hence, $\lim_n \sqrt{n} I_{\{X_n = 0\}} = 0$ a.s., which implies that $\lim_n I_{\{X_n = 0\}} = 0$ a.s. and $\lim_n I_{\{X_n \neq 0\}} = 1$ a.s. By the continuity of the log function on $(0, \infty)$, $\lim_n Y_n = \log\theta$ a.s.

Since X_n has the same distribution as $\sum_{j=1}^n Z_j$, by the central limit theorem, $\sqrt{n}(X_n/n - \theta) \to_d N(0, \theta(1-\theta))$. Since we have shown that

$\lim_n \sqrt{n} I_{\{X_n=0\}} = 0$ a.s. and $\lim_n X_n/n = \theta$ a.s., $\lim_n I_{\{X_n=0\}} X_n/\sqrt{n} = 0$ a.s. By Slutsky's theorem (e.g., Theorem 1.11 in Shao, 2003),

$$\begin{aligned}\sqrt{n}(W_n - \theta) &= \sqrt{n}\left(\frac{X_n}{n} - \theta\right) - I_{\{X_n=0\}}\frac{X_n}{\sqrt{n}} \\ &\to_d N(0, \theta(1-\theta)).\end{aligned}$$

Then, by the δ-method with $g(t) = \log t$ and $g'(t) = t^{-1}$ (e.g., Theorem 1.12 in Shao, 2003), $\sqrt{n}(\log W_n - \log\theta) \to_d N\left(0, \frac{1-\theta}{\theta}\right)$. Since $\sqrt{n}(Y_n - \log\theta) = \sqrt{n}(\log W_n - \log\theta) + \sqrt{n} I_{\{X_n=0\}}$, by Slutsky's theorem again, we obtain that $\sqrt{n}(Y_n - \log\theta) \to_d N\left(0, \frac{1-\theta}{\theta}\right)$. ∎

Exercise 62 (#1.149). Let $X_1, ..., X_n$ be independent and identically distributed random variables such that for $x = 3, 4, ...$, $P(X_1 = \pm x) = (2cx^2 \log x)^{-1}$, where $c = \sum_{x=3}^{\infty} x^{-2}/\log x$. Show that $E|X_1| = \infty$ but $n^{-1}\sum_{i=1}^n X_i \to_p 0$.
Solution. Note that

$$E|X_1| = c^{-1}\sum_{x=3}^{\infty}\frac{1}{x\log x} \geq c^{-1}\int_3^{\infty}\frac{1}{x\log x}dx = \infty.$$

For any positive integer n, $E[X_1 I_{(-n,n)}(X_1)] = 0$. For sufficiently large x,

$$\begin{aligned}x[1 - F(x) + F(-x)] &< c^{-1}x\sum_{k=x}^{\infty}\frac{1}{k^2\log k} \\ &\leq c^{-1}x\int_{x-1}^{\infty}\frac{1}{t^2\log t}dt \\ &\leq \frac{c^{-1}x}{\log(x-1)}\int_{x-1}^{\infty}\frac{1}{t^2}dt \\ &= \frac{c^{-1}x}{\log(x-1)}\cdot\frac{1}{x-1} \\ &\to 0\end{aligned}$$

as $x \to \infty$. By the weak law of large numbers (e.g., Theorem 1.13(i) in Shao, 2003), $n^{-1}\sum_{i=1}^n X_i \to_p 0$. ∎

Exercise 63 (#1.151). Let $X_1, X_2, ...$ be independent random variables. Assume that $\lim_n \sum_{i=1}^n P(|X_i| > n) = 0$ and $\lim_n n^{-2}\sum_{i=1}^n E(X_i^2 I_{\{|X_i| \leq n\}}) = 0$. Show that $(\sum_{i=1}^n X_i - b_n)/n \to_p 0$, where $b_n = \sum_{i=1}^n E(X_i I_{\{|X_i|\leq n\}})$.
Solution. For any n, let $Y_{ni} = X_i I_{\{|X_i|\leq n\}}$, $i = 1, ..., n$. Define $T_n = \sum_{i=1}^n X_i$ and $Z_n = \sum_{i=1}^n Y_{ni}$. Then

$$P(T_n \neq Z_n) \leq \sum_{i=1}^n P(Y_{ni} \neq X_i) = \sum_{i=1}^n P(|X_i| > n) \to 0$$

as $n \to \infty$. For any $\epsilon > 0$,

$$\begin{aligned} P\left(\frac{|Z_n - EZ_n|}{n} \geq \epsilon\right) &\leq \frac{\mathrm{Var}(Z_n)}{\epsilon^2 n^2} \\ &= \frac{1}{\epsilon^2 n^2}\sum_{i=1}^n \mathrm{Var}(Y_{ni}) \\ &\leq \frac{1}{\epsilon^2 n^2}\sum_{i=1}^n EY_{ni}^2 \\ &= \frac{1}{\epsilon^2 n^2}\sum_{i=1}^n E(X_i^2 I_{\{|X_i|\leq n\}}) \\ &\to 0 \end{aligned}$$

as $n \to \infty$, where the first equality follows from the fact that $Y_{n1}, ..., Y_{nn}$ are independent since $X_1, ..., X_n$ are independent. Thus,

$$\begin{aligned} P\left(\frac{|T_n - EZ_n|}{n} \geq \epsilon\right) &\leq P\left(\frac{|T_n - EZ_n|}{n} \geq \epsilon, T_n = Z_n\right) + P(T_n \neq Z_n) \\ &\leq P\left(\frac{|Z_n - EZ_n|}{n} \geq \epsilon\right) + P(T_n \neq Z_n) \\ &\to 0 \end{aligned}$$

under the established results. Hence the result follows from the fact that $b_n = EZ_n$. ∎

Exercise 64 (#1.154). Let $X_1, ..., X_n$ be independent and identically distributed random variables with $\mathrm{Var}(X_1) < \infty$. Show that

$$\frac{1}{n(n+1)}\sum_{j=1}^n jX_j \to_p EX_1.$$

Note. A simple way to solve a problem of showing $Y_n \to_p a$ is to establish $\lim_n EY_n = a$ and $\lim_n \mathrm{Var}(Y_n) = 0$.

Solution. Note that

$$E\left(\frac{2}{n(n+1)}\sum_{j=1}^n jX_j\right) = \frac{2}{n(n+1)}\sum_{j=1}^n jEX_j = EX_1.$$

Let $\sigma^2 = \mathrm{Var}(X_1)$. Then,

$$\mathrm{Var}\left(\frac{2}{n(n+1)}\sum_{j=1}^n jX_j\right) = \frac{4\sigma^2}{n^2(n+1)^2}\sum_{j=1}^n j^2 = \frac{2\sigma^2(2n+1)}{3n(n+1)} \to 0$$

as $n \to \infty$. ∎

Exercise 65 (#1.165). Let $X_1, X_2, ...$ be independent random variables. Suppose that $\sum_{j=1}^n (X_j - EX_j)/\sigma_n \to_d N(0,1)$, where $\sigma_n^2 = \text{Var}(\sum_{j=1}^n X_j)$. Show that $n^{-1}\sum_{j=1}^n (X_j - EX_j) \to_p 0$ if and only if $\lim_n \sigma_n/n = 0$.

Solution. If $\lim_n \sigma_n/n = 0$, then by Slutsky's theorem (e.g., Theorem 1.11 in Shao, 2003),

$$\frac{1}{n}\sum_{j=1}^n (X_j - EX_j) = \frac{\sigma_n}{n}\frac{1}{\sigma_n}\sum_{j=1}^n (X_j - EX_j) \to_d 0.$$

Assume now σ_n/n does not converge to 0 but $n^{-1}\sum_{j=1}^n (X_j - EX_j) \to_p 0$. Without loss of generality, assume that $\lim_n \sigma_n/n = c \in (0, \infty]$. By Slutsky's theorem,

$$\frac{1}{\sigma_n}\sum_{j=1}^n (X_j - EX_j) = \frac{n}{\sigma_n}\frac{1}{n}\sum_{j=1}^n (X_j - EX_j) \to_p 0.$$

This contradicts the fact that $\sum_{j=1}^n (X_j - EX_j)/\sigma_n \to_d N(0,1)$. Hence, $n^{-1}\sum_{j=1}^n (X_j - EX_j)$ does not converge to 0 in probability. ∎

Exercise 66 (#1.152, #1.166). Let $T_n = \sum_{i=1}^n X_i$, where X_n's are independent random variables satisfying $P(X_n = \pm n^\theta) = 0.5$ and $\theta > 0$ is a constant. Show that

(i) $T_n/\sqrt{\text{Var}(T_n)} \to_d N(0,1)$;

(ii) when $\theta < 0.5$, $\lim_n T_n/n = 0$ a.s.;

(iii) when $\theta \geq 0.5$, T_n/n does not converge to 0 in probability.

Solution. (i) Note that $ET_n = 0$ and $\text{Var}(X_n) = n^{2\theta}$ for any n. Hence, $\sigma_n^2 = \text{Var}(T_n) = \sum_{j=1}^n j^{2\theta}$. Since

$$\sum_{j=1}^{n-1} j^{2\theta} \leq \sum_{j=1}^{n-1}\int_j^{j+1} x^{2\theta}dx \leq \sum_{j=2}^n j^{2\theta}$$

and

$$\sum_{j=1}^{n-1}\int_j^{j+1} x^{2\theta}dx = \int_1^n x^{2\theta}dx = \frac{n^{2\theta+1}-1}{2\theta+1},$$

we conclude that

$$\lim_n \frac{\sigma_n^2}{n^{2\theta+1}} = \lim_n \frac{1}{n^{2\theta+1}}\sum_{j=1}^n j^{2\theta} = \frac{1}{2\theta+1}.$$

Then $\lim_n n^\theta/\sigma_n = 0$ and, for any $\epsilon > 0$, $n^\theta < \epsilon\sigma_n$ for sufficiently large n. Since $|X_n| \leq n^\theta$, when n is sufficiently large, $I_{\{|X_j|>\epsilon\sigma_n\}} = 0$, $j = 1, ..., n$,

and, hence, $\sum_{j=1}^n E(X_j^2 I_{\{|X_j|>\epsilon\sigma_n\}}) = 0$. Thus, Lindeberg's condition holds and, by the Lindeberg central limit theorem (e.g., Theorem 1.15 in Shao, 2003), $T_n/\sqrt{\text{Var}(T_n)} \to_d N(0,1)$.
(ii) When $\theta < 0.5$,

$$\sum_{n=1}^{\infty} \frac{EX_n^2}{n^2} = \sum_{n=1}^{\infty} \frac{n^{2\theta}}{n^2} < \infty.$$

By the Kolmogorov strong law of large numbers (e.g., Theorem 1.14 in Shao, 2003), $\lim_n T_n/n = 0$ a.s.
(iii) From the result in (i) and the result in the previous exercise, $T_n/n \to_p 0$ if and only if $\lim_n \sigma_n/n = 0$. In part (i), we have shown that $\lim_n \sigma_n^2/n^{2\theta+1}$ equals a positive constant. Hence, the result follows since $\lim_n \sigma_n/n \neq 0$ when $\theta \geq 0.5$. ∎

Exercise 67 (#1.162). Let $X_1, X_2, \ldots$ be independent random variables such that X_j has the uniform distribution on $[-j, j]$, $j = 1, 2, \ldots$. Show that Lindeberg's condition is satisfied.
Solution. Note that $EX_j = 0$ and $\text{Var}(X_j) = \int_{-j}^{j} x^2 dx = 2j^3/3$ for all j. Hence

$$\sigma_n^2 = \text{Var}\left(\sum_{j=1}^n X_j\right) = \frac{2}{3}\sum_{j=1}^n j^3 = \frac{n^2(n+1)^2}{6}.$$

For any $\epsilon > 0$, $n < \epsilon\sigma_n$ for sufficiently large n, since $\lim_n n/\sigma_n = 0$. Since $|X_j| \leq j \leq n$, when n is sufficiently large,

$$\sum_{j=1}^n E(X_j^2 I_{\{|X_j|>\epsilon\sigma_n\}}) = 0.$$

Thus, Lindeberg's condition holds. ∎

Exercise 68 (#1.163). Let $X_1, X_2, \ldots$ be independent random variables such that for $j = 1, 2, \ldots$, $P(X_j = \pm j^a) = 6^{-1} j^{-2(a-1)}$ and $P(X_j = 0) = 1 - 3^{-1} j^{-2(a-1)}$, where $a > 1$ is a constant. Show that Lindeberg's condition is satisfied if and only if $a < 1.5$.
Solution. Note that $EX_j = 0$ and

$$\sigma_n^2 = \sum_{j=1}^n \text{Var}(X_j) = \sum_{j=1}^n \frac{2j^{2a}}{6j^{2(a-1)}} = \frac{1}{3}\sum_{j=1}^n j^2 = \frac{n(n+1)(2n+1)}{18}.$$

Assume first $a < 1.5$. For any $\epsilon > 0$, $n^a < \epsilon\sigma_n$ for sufficiently large n, since $\lim_n n^a/\sigma_n = 0$. Note that $|X_n| \leq n^a$ for all n. Therefore, when n is sufficiently large, $I_{\{|X_j|>\epsilon\sigma_n\}} = 0$, $j = 1, \ldots, n$, and, hence,

$$\sum_{j=1}^n E(X_j^2 I_{\{|X_j|>\epsilon\sigma_n\}}) = 0.$$

Thus, Lindeberg's condition holds.

Assume now $a \geq 1.5$. For $\epsilon \in (0, \frac{1}{3})$, let k_n be the integer part of $(\epsilon\sigma_n)^{1/a}$. Then

$$\frac{1}{\sigma_n^2}\sum_{j=1}^{n} E(X_j^2 I_{\{|X_j|>\epsilon\sigma_n\}}) = \frac{1}{\sigma_n^2}\left(\sum_{j=1}^{n}\frac{j^2}{3} - \sum_{j=1}^{k_n}\frac{j^2}{3}\right)$$
$$= 1 - \frac{k_n(k_n+1)(2k_n+1)}{18\sigma_n^2},$$

which converges, as $n \to \infty$, to 1 if $a > 1.5$ (since $\lim_n k_n/n = 0$) and to $1 - \epsilon^2/9$ if $a = 1.5$ (since $\lim_n k_n^3/\sigma_n^2 = \epsilon^2$). Hence, Lindeberg's condition does not hold. ∎

Exercise 69 (#1.155). Let $\{X_n\}$ be a sequence of random variables and let $\bar{X}_n = \sum_{i=1}^n X_i/n$. Show that
(i) if $\lim_n X_n = 0$ a.s., then $\lim_n \bar{X}_n = 0$ a.s.;
(ii) if $\sup_n E|X_n|^r < \infty$ and $\lim_n E|X_n|^r = 0$, then $\lim_n E|\bar{X}_n|^r = 0$, where $r \geq 1$ is a constant;
(iii) the result in part (ii) may not be true for $r \in (0, 1)$;
(iv) $X_n \to_p 0$ may not imply $\bar{X}_n \to_p 0$.

Solution. (i) The result in this part is actually a well known result in mathematical analysis. It suffices to show that if $\{x_n\}$ is a sequence of real numbers satisfying $\lim_n x_n = 0$, then $n^{-1}\sum_{i=1}^n x_i = 0$. Assume that $\lim_n x_n = 0$. Then $M = \sup_n |x_n| < \infty$ and, for any $\epsilon > 0$, there is an N such that $|x_n| \leq \epsilon$ for all $n > N$. Then, for $n > \max\{N, NM/\epsilon\}$,

$$\left|\frac{1}{n}\sum_{i=1}^{n} x_i\right| \leq \frac{1}{n}\left(\sum_{i=1}^{N}|x_i| + \sum_{i=N+1}^{n}|x_i|\right)$$
$$\leq \frac{1}{n}\left(\sum_{i=1}^{N} M + \sum_{i=N+1}^{n}\epsilon\right)$$
$$= \frac{NM}{n} + \frac{\epsilon(n-N)}{n}$$
$$< \epsilon + \epsilon = 2\epsilon.$$

(ii) When $r \geq 1$, $|x|^r$ is a convex function. By Jensen's inequality, $E|\bar{X}_n|^r \leq n^{-1}\sum_{i=1}^n E|X_i|^r$. When $\lim_n E|X_n|^r = 0$, $\lim_n n^{-1}\sum_{i=1}^n E|X_i|^r = 0$ (the result in part (i)). Hence, $\lim_n E|\bar{X}_n|^r = 0$.
(iii)-(iv) Consider the X_i's in the previous exercise, i.e., $X_1, X_2, \ldots$ are independent, $P(X_j = \pm j^a) = 6^{-1}j^{-2(a-1)}$ and $P(X_j = 0) = 1 - 3^{-1}j^{-2(a-1)}$, where a is a constant satisfying $1 < a < 1.5$. Let r be a constant such that $0 < r < 2(a-1)/a$. Then $0 < r < 1$. Note that

$$\lim_n E|X_n|^r = \lim_n 3^{-1}n^{ar-2(a-1)} = 0$$

and, hence, $X_n \to_p 0$. From the result in the previous exercise, $\sum_{i=1}^n X_i/\sigma_n \to_d N(0,1)$ with $\sigma_n^2 = n(n+1)(2n+1)/18$. Since $\lim_n \sigma_n/n = \infty$, $\bar{X}_n$ does not converge to 0 in probability. This shows that $X_n \to_p 0$ may not imply $\bar{X}_n \to_p 0$. Furthermore, $E|\bar{X}_n|^r$ does not converge to 0, because if it does, then $\bar{X}_n \to_p 0$. This shows that the result in part (ii) may not be true for $r \in (0,1)$. ∎

Exercise 70 (#1.164). Let $X_1, X_2, ...$ be independent random variables satisfying $P(X_j = \pm j^a) = P(X_j = 0) = 1/3$, where $a > 0$, $j = 1, 2, ...$ Show that Liapounov's condition holds, i.e.,

$$\lim_n \frac{1}{\sigma_n^{2+\delta}} \sum_{j=1}^n E|X_j - EX_j|^{2+\delta} = 0$$

for some $\delta > 0$, where $\sigma_n^2 = \mathrm{Var}(\sum_{j=1}^n X_j)$.
Solution. Note that $EX_j = 0$ and

$$\sigma_n^2 = \sum_{j=1}^n \mathrm{Var}(X_j) = \frac{2}{3}\sum_{j=1}^n j^{2a}.$$

For any $\delta > 0$,

$$\sum_{j=1}^n E|X_j - E(X_j)|^{2+\delta} = \frac{2}{3}\sum_{j=1}^n j^{(2+\delta)a}.$$

From the proof of Exercise 66,

$$\lim_n \frac{1}{n^{t+1}} \sum_{j=1}^n j^t = \frac{1}{t+1}$$

for any $t > 0$. Thus,

$$\begin{aligned}\lim_n \frac{1}{\sigma_n^{2+\delta}} \sum_{j=1}^n E|X_j - EX_j|^{2+\delta} &= \lim_n \frac{\frac{2}{3}\sum_{j=1}^n j^{(2+\delta)a}}{\left(\frac{2}{3}\sum_{j=1}^n j^{2a}\right)^{1+\delta/2}} \\ &= \lim_n \left(\frac{3}{2}\right)^{\delta/2} \frac{(2a+1)^{1+\delta/2}}{(2+\delta)a+1} \frac{n^{(2+\delta)a+1}}{n^{(2a+1)(1+\delta)}} \\ &= \left(\frac{3}{2}\right)^{\delta/2} \frac{(2a+1)^{1+\delta/2}}{(2+\delta)a+1} \lim_n \frac{1}{n^{\delta/2}} \\ &= 0. \quad ∎\end{aligned}$$

Chapter 2

Fundamentals of Statistics

Exercise 1 (#2.9). Consider the family of double exponential distributions: $\mathcal{P} = \left\{2^{-1}e^{-|t-\mu|} : \ -\infty < \mu < \infty\right\}$. Show that $\mathcal{P}$ is not an exponential family.

Solution. Assume that $\mathcal{P}$ is an exponential family. Then there exist p-dimensional Borel functions $T(X)$ and $\eta(\mu)$ ($p \geq 1$) and one-dimensional Borel functions $h(X)$ and $\xi(\mu)$ such that

$$2^{-1}\exp\{-|t-\mu|\} = \exp\{[\eta(\mu)]^\tau T(t) - \xi(\mu)\}h(t)$$

for any t and μ. Let $X = (X_1, ..., X_n)$ be a random sample from $P \in \mathcal{P}$ (i.e., $X_1, ..., X_n$ are independent and identically distributed with $P \in \mathcal{P}$), where $n > p$, $T_n(X) = \sum_{i=1}^n T(X_i)$, and $h_n(X) = \Pi_{i=1}^n h(X_i)$. Then the joint Lebesgue density of X is

$$2^{-n}\exp\left\{-\sum_{i=1}^n |x_i - \mu|\right\} = \exp\left\{[\eta(\mu)]^\tau T_n(x) - n\xi(\mu)\right\} h_n(x)$$

for any $x = (x_1, ..., x_n)$ and μ, which implies that

$$\sum_{i=1}^n |x_i| - \sum_{i=1}^n |x_i - \mu| = [\tilde{\eta}(\mu)]^\tau T_n(x) - n\tilde{\xi}(\mu)$$

for any x and μ, where $\tilde{\eta}(\mu) = \eta(\mu) - \eta(0)$ and $\tilde{\xi}(\mu) = \xi(\mu) - \xi(0)$. Define $\psi_\mu(x) = \sum_{i=1}^n |x_i| - \sum_{i=1}^n |x_i - \mu|$. We conclude that if $x = (x_1, ..., x_n)$ and $y = (y_1, ..., y_n)$ such that $T_n(x) = T_n(y)$, then $\psi_\mu(x) = \psi_\mu(y)$ for all μ, which implies that vector of the ordered x_i's is the same as the vector of the ordered y_i's.

On the other hand, we may choose real numbers $\mu_1, ..., \mu_p$ such that $\tilde{\eta}(\mu_i)$, $i = 1, ..., p$, are linearly independent vectors. Since

$$\psi_{\mu_i}(x) = [\tilde{\eta}(\mu_i)]^\tau T_n(x) - n\tilde{\xi}(\mu_i), \quad i = 1, ..., p,$$

for any x, $T_n(x)$ is then a function of the p functions $\psi_{\mu_i}(x)$, $i = 1, ..., p$. Since $n > p$, it can be shown that there exist x and y in $\mathcal{R}^n$ such that $\psi_{\mu_i}(x) = \psi_{\mu_i}(y)$, $i = 1, ..., p$, (which implies $T_n(x) = T_n(y)$), but the vector of ordered x_i's is not the same as the vector of ordered y_i's. This contradicts the previous conclusion. Hence, $\mathcal{P}$ is not an exponential family. ∎

Exercise 2 (#2.13). A discrete random variable X with

$$P(X = x) = \gamma(x)\theta^x/c(\theta), \quad x = 0, 1, 2, ...,$$

where $\gamma(x) \geq 0$, $\theta > 0$, and $c(\theta) = \sum_{x=0}^{\infty} \gamma(x)\theta^x$, is called a random variable with a power series distribution. Show that
(i) $\{\gamma(x)\theta^x/c(\theta) : \theta > 0\}$ is an exponential family;
(ii) if $X_1, ..., X_n$ are independent and identically distributed with a power series distribution $\gamma(x)\theta^x/c(\theta)$, then $\sum_{i=1}^n X_i$ has the power series distribution $\gamma_n(x)\theta^x/[c(\theta)]^n$, where $\gamma_n(x)$ is the coefficient of θ^x in the power series expansion of $[c(\theta)]^n$.
Solution. (i) Note that

$$\gamma(x)\theta^x/c(\theta) = \exp\{x \log\theta - \log(c(\theta))\}\gamma(x).$$

Thus, $\{\gamma(x)\theta^x/c(\theta) : \theta > 0\}$ is an exponential family.
(ii) From part (i), we know that the natural parameter $\eta = \log\theta$, and also $\zeta(\eta) = \log(c(e^\eta))$. From the properties of exponential families (e.g., Theorem 2.1 in Shao, 2003), the moment generating function of X is $\psi_X(t) = e^{\zeta(\eta+t)}/e^{\zeta(\eta)} = c(\theta e^t)/c(\theta)$. The moment generating function of $\sum_{i=1}^n X_i$ is $[c(\theta e^t)]^n/[c(\theta)]^n$, which is the moment generating function of the power series distribution $\gamma_n(x)\theta^x/[c(\theta)]^n$. ∎

Exercise 3 (#2.17). Let X be a random variable having the gamma distribution with shape parameter α and scale parameter γ, where α is know and γ is unknown. let $Y = \sigma \log X$. Show that
(i) if $\sigma > 0$ is unknown, then the distribution of Y is in a location-scale family;
(ii) if $\sigma > 0$ is known, then the distribution of Y is in an exponential family.
Solution. (i) The Lebesgue density of X is

$$\frac{1}{\Gamma(\alpha)\gamma^\alpha} x^{\alpha-1} e^{-x/\gamma} I_{(0,\infty)}(x).$$

Applying the result in the note of Exercise 17 in Chapter 1, the Lebesgue density for $Y = \sigma \log X$ is

$$\frac{1}{\Gamma(\alpha)\sigma} e^{\alpha(y-\sigma\log\gamma)/\sigma} \exp\left\{-e^{(y-\sigma\log\gamma)/\sigma}\right\}.$$

It belongs to a location-scale family with location parameter $\eta = \sigma\log\gamma$ and scale parameter σ.

(ii) When σ is known, we rewrite the density of Y as

$$\frac{1}{\sigma\Gamma(\alpha)}\exp\{\alpha y/\sigma\}\exp\left\{-\frac{e^{y/\sigma}}{\gamma}-\alpha\log\gamma\right\}.$$

Therefore, the distribution of Y is from an exponential family. ∎

Exercise 4. Let $(X_1,...,X_n)$ be a random sample from $N(0,1)$. Show that $X_i^2/\sum_{j=1}^n X_j^2$ and $\sum_{j=1}^n X_j^2$ are independent, $i=1,...,n$.
Solution. Note that $X_1^2,...,X_n^2$ are independent and have the chi-square distribution χ_1^2. Hence their joint Lebesgue density is

$$\frac{ce^{-(y_1+\cdots+y_n)/2}}{\sqrt{y_1\cdots y_n}},\quad y_j>0,$$

where c is a constant. Let $U=\sum_{j=1}^n X_j^2$ and $V_i=X_i^2/U$, $i=1,...,n$. Then $X_i^2=UV_i$ and $\sum_{j=1}^n V_j=1$. The Lebesgue density for $(U,V_1,...,V_{n-1})$ is

$$\frac{ce^{-u/2}v_nu^{n-1}}{\sqrt{u^nv_1\cdots v_n}}=cu^{n/2-1}e^{-u/2}\sqrt{\frac{1-v_1\cdots v_{n-1}}{v_1\cdots v_{n-1}}},\quad u>0,v_j>0.$$

Hence U and $(V_1/U,...,V_{n-1}/U)$ are independent. Since $V_n=1-(V_1+\cdots+V_{n-1})$, we conclude that U and V_n/U are independent.

An alternative solution can be obtained by using Basu's theorem (e.g., Theorem 2.4 in Shao, 2003). ∎

Exercise 5. Let $X=(X_1,...,X_n)$ be a random n-vector having the multivariate normal distribution $N_n(\mu J,D)$, where J is the n-vector of 1's,

$$D=\sigma^2\begin{pmatrix}1&\rho&\cdots&\rho\\ \rho&1&\cdots&\rho\\ \cdots&\cdots&\cdots&\cdots\\ \rho&\rho&\cdots&1\end{pmatrix},$$

and $|\rho|<1$. Show that $\bar{X}=n^{-1}\sum_{i=1}^n X_i$ and $W=\sum_{i=1}^n(X_i-\bar{X})^2$ are independent, $\bar{X}$ has the normal distribution $N\left(\mu,\frac{1+(n-1)\rho}{n}\sigma^2\right)$, and $W/[(1-\rho)\sigma^2]$ has the chi-square distribution χ_{n-1}^2.
Solution. Define

$$A=\begin{pmatrix}\frac{1}{\sqrt{n}}&\frac{1}{\sqrt{n}}&\frac{1}{\sqrt{n}}&\frac{1}{\sqrt{n}}&\cdots&\frac{1}{\sqrt{n}}\\ \frac{1}{\sqrt{2\cdot1}}&\frac{-1}{\sqrt{2\cdot1}}&0&0&\cdots&0\\ \frac{1}{\sqrt{3\cdot2}}&\frac{1}{\sqrt{3\cdot2}}&\frac{-2}{\sqrt{3\cdot2}}&0&\cdots&0\\ \cdots&\cdots&\cdots&\cdots&\cdots&\cdots\\ \frac{1}{\sqrt{n(n-1)}}&\frac{1}{\sqrt{n(n-1)}}&\frac{1}{\sqrt{n(n-1)}}&\frac{1}{\sqrt{n(n-1)}}&\cdots&\frac{-(n-1)}{\sqrt{n(n-1)}}\end{pmatrix}.$$

Then $AA^\tau = I$ (the identity matrix) and

$$ADA^\tau = \sigma^2 \begin{pmatrix} 1+(n-1)\rho & 0 & \cdots & 0 \\ 0 & 1-\rho & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & \cdots & 1-\rho \end{pmatrix}.$$

Let $Y = AX$. Then Y is normally distributed with $E(Y) = AE(X) = (\sqrt{n}\mu, 0, ..., 0)$ and $\text{Var}(Y) = ADA^\tau$, i.e., the components of Y are independent. Let Y_i be the ith component of Y. Then, $Y_1 = \sqrt{n}\bar{X}$ and $\sum_{i=1}^n Y_i^2 = Y^\tau Y = X^\tau A^\tau AX = X^\tau X = \sum_{i=1}^n X_i^2$. Hence $\bar{X} = Y_1/\sqrt{n}$ and $W = \sum_{i=1}^n (X_i - \bar{X})^2 = \sum_{i=1}^n X_i^2 - n\bar{X}^2 = \sum_{i=1}^n Y_i^2 - Y_1^2 = \sum_{i=2}^n Y_i^2$. Since Y_i's are independent, $\bar{X}$ and W are independent.

Since Y_1 has distribution $N(\sqrt{n}\mu, [1+(n-1)\rho]\sigma^2)$, $\bar{X} = Y_1/\sqrt{n}$ has distribution $N\left(\mu, \frac{1+(n-1)\rho}{n}\sigma^2\right)$. Since $Y_2, ..., Y_n$ are independent and identically distributed as $N(0, (1-\rho)\sigma^2)$, $W/[(1-\rho)\sigma^2] = \sum_{i=2}^n Y_i^2/[(1-\rho)\sigma^2]$ has the χ^2_{n-1} distribution. ∎

Exercise 6. Let $(X_1, ..., X_n)$ be a random sample from the uniform distribution on the interval $[0,1]$ and let $R = X_{(n)} - X_{(1)}$, where $X_{(i)}$ is the ith order statistic. Derive the Lebesgue density of R and show that the limiting distribution of $2n(1-R)$ is the chi-square distribution χ^2_4.
Solution. The joint Lebesgue density of $X_{(1)}$ and $X_{(n)}$ is

$$f(x,y) = \begin{cases} n(n-1)(y-x)^{n-2} & 0 < x < y < 1 \\ 0 & \text{otherwise} \end{cases}$$

(see, e.g., Example 2.9 in Shao, 2003). Then, the joint Lebesgue density of R and $X_{(n)}$ is

$$g(x,y) = \begin{cases} n(n-1)x^{n-2} & 0 < x < y < 1 \\ 0 & \text{otherwise} \end{cases}$$

and, when $0 < x < 1$, the Lebesgue density of R is

$$\int g(x,y)dy = \int_x^1 n(n-1)y^{n-2}ds = n(n-1)x^{n-2}(1-x)$$

for $0 < x < 1$. Consequently, the Lebesgue density of $2n(1-R)$ is

$$h_n(x) = \begin{cases} \frac{n-1}{4n}x\left(1-\frac{x}{2n}\right)^{n-2} & 0 < x < 2n \\ 0 & \text{otherwise.} \end{cases}$$

Since $\lim_n \left(1-\frac{x}{2n}\right)^{n-2} = e^{-x/2}$, $\lim_n h_n(x) = 4^{-1}xe^{x/2}I_{(0,\infty)}(x)$, which is the Lebesgue density of the χ^2_4 distribution. By Scheffé's theorem (e.g., Proposition 1.18 in Shao, 2003), the limiting distribution of $2n(1-R)$ is the χ^2_4 distribution. ∎

Exercise 7. Let $(X_1, ..., X_n)$ be a random sample from the exponential distribution with Lebesgue density $\theta^{-1}e^{(a-x)/\theta}I_{(0,\infty)}(x)$, where $a \in \mathcal{R}$ and $\theta > 0$ are parameters. Let $X_{(1)} \leq \cdots \leq X_{(n)}$ be order statistics, $X_{(0)} = 0$, and $Z_i = X_{(i)} - X_{(i-1)}$, $i = 1, ..., n$. Show that
(i) $Z_1, ..., Z_n$ are independent and $2(n-i+1)Z_i/\theta$ has the χ^2_2 distribution;
(ii) $2[\sum_{i=1}^r X_{(i)} + (n-r)X_{(r)} - na]/\theta$ has the χ^2_{2r} distribution, $r = 1, ..., n$;
(iii) $X_{(1)}$ and Y are independent and $(X_{(1)} - a)/Y$ has the Lebesgue density $n\left(1 + \frac{nt}{n-1}\right)^{-n} I_{(0,\infty)}(t)$, where $Y = (n-1)^{-1}\sum_{i=1}^n (X_i - X_{(1)})$.

Solution. If we can prove the result for the case of $a = 0$ and $\theta = 1$, then the result for the general case follows by considering the transformation $(X_i - a)/\theta$, $i = 1, ..., n$. Hence, we assume that $a = 0$ and $\theta = 1$.
(i) The joint Lebesgue density of $X_{(1)}, ..., X_{(n)}$ is

$$f(x_1, ..., x_n) = \begin{cases} n!e^{-x_1-\cdots-x_n} & 0 < x_1 < \cdots < x_n \\ 0 & \text{otherwise.} \end{cases}$$

Then the joint Lebesgue density of Z_i, $i = 1, ..., n$, is

$$g(x_1, ..., x_n) = \begin{cases} n!e^{-nx_1-\cdots-(n-i+1)x_i-\cdots-x_n} & x_i > 0,\ i = 1, ..., n, \\ 0 & \text{otherwise.} \end{cases}$$

Hence $Z_1, ..., Z_n$ are independent and, for each i, the Lebesgue density of $2Z_i$ is $(n-i+1)e^{-(n-i+1)x_i}I_{(0,\infty)}(x_i)$. Then the density of $2(n-i+1)Z_i$ is $2^{-1}e^{-x_i/2}I_{(0,\infty)}(x_i)$, which is the density of the χ^2_2 distribution.
(ii) For $r = 1, ..., n$,

$$\sum_{i=1}^r X_{(i)} + (n-r)X_{(r)} = \sum_{i=1}^r (n-i+1)Z_i.$$

From (i), $Z_1, ..., Z_n$ are independent and $2(n-i+1)Z_i$ has the χ^2_2 distribution. Hence $2\sum_{i=1}^r X_{(i)} + (n-r)X_{(r)}$ has the χ^2_{2r} distribution for any r.
(iii) Note that

$$Y = \frac{1}{n-1}\sum_{i=2}^n (X_{(i)} - X_{(1)}) = \frac{1}{n-1}\sum_{i=2}^n (n-i+1)Z_i.$$

From the result in (i), Y and $X_{(1)}$ are independent and $2(n-1)Y$ has the $\chi^2_{2(n-1)}$ distribution. Hence the Lebesgue density of Y is $f_Y(y) = \frac{(n-1)^n}{(n-1)!}y^{n-2}e^{-(n-1)y}I_{(0,\infty)}(y)$. Note that the Lebesgue density of $X_{(1)}$ is

$f_{X_{(1)}}(x) = ne^{-nx}I_{(0,\infty)}(x)$. Hence, for $t > 0$, the density of the ratio $X_{(1)}/Y$ is (e.g., Example 1.15 in Shao, 2003)

$$\begin{aligned} f(t) &= \int |x| f_Y(x) f_{X_{(1)}}(tx)dx \\ &= \int_0^\infty \frac{n(n-1)^n}{(n-1)!} x^{n-1} e^{-(n+nt-1)x} dx \\ &= n\left(1+\frac{nt}{n-1}\right)^{-n} \int_0^\infty \frac{(n+nt-1)^n}{(n-1)!} x^{n-1} e^{-(n+nt-1)x} dx \\ &= n\left(1+\frac{nt}{n-1}\right)^{-n}. \ \blacksquare \end{aligned}$$

Exercise 8 (#2.19). Let $(X_1, ..., X_n)$ be a random sample from the gamma distribution with shape parameter α and scale parameter γ_x and let $(Y_1, ..., Y_n)$ be a random sample from the gamma distribution with shape parameter α and scale parameter γ_y. Assume that X_i's and Y_i's are independent. Derive the distribution of the statistic $\bar{X}/\bar{Y}$, where $\bar{X}$ and $\bar{Y}$ are the sample means based on X_i's and Y_i's, respectively.
Solution. From the property of the gamma distribution, $n\bar{X}$ has the gamma distribution with shape parameter $n\alpha$ and scale parameter γ_x and $n\bar{Y}$ has the gamma distribution with shape parameter $n\alpha$ and scale parameter γ_y. Since $\bar{X}$ and $\bar{Y}$ are independent, the Lebesgue density of the ratio $\bar{X}/\bar{Y}$ is, for $t > 0$,

$$\begin{aligned} f(t) &= \frac{1}{[\Gamma(n\alpha)]^2(\gamma_x\gamma_y)^{n\alpha}} \int_0^\infty (tx)^{n\alpha-1} e^{-tx/\gamma_x} x^{n\alpha} e^{-x/\gamma_y} dx \\ &= \frac{\Gamma(2n\alpha)t^{n\alpha-1}}{[\Gamma(n\alpha)]^2(\gamma_x\gamma_y)^{n\alpha}} \left(\frac{t}{\gamma_x}+\frac{1}{\gamma_y}\right)^{-2n\alpha}. \ \blacksquare \end{aligned}$$

Exercise 9 (#2.22). Let (Y_i, Z_i), $i = 1, ..., n$, be independent and identically distributed random 2-vectors. The sample correlation coefficient is defined to be

$$T = \frac{1}{(n-1)S_Y S_Z} \sum_{i=1}^n (Y_i - \bar{Y})(Z_i - \bar{Z}),$$

where $\bar{Y} = n^{-1}\sum_{i=1}^n Y_i$, $\bar{Z} = n^{-1}\sum_{i=1}^n Z_i$, $S_Y^2 = (n-1)^{-1}\sum_{i=1}^n (Y_i - \bar{Y})^2$, and $S_Z^2 = (n-1)^{-1}\sum_{i=1}^n (Z_i - \bar{Z})^2$.
(i) Assume that $E|Y_i|^4 < \infty$ and $E|Z_i|^4 < \infty$. Show that

$$\sqrt{n}(T - \rho) \to_d N(0, c^2),$$

where ρ is the correlation coefficient between Y_1 and Z_1 and c is a constant. Identify c in terms of moments of (Y_1, Z_1).
(ii) Assume that Y_i and Z_i are independently distributed as $N(\mu_1, \sigma_1^2)$ and $N(\mu_2, \sigma_2^2)$, respectively. Show that T has the Lebesgue density

$$\frac{\Gamma\left(\frac{n-1}{2}\right)}{\sqrt{\pi}\Gamma\left(\frac{n-2}{2}\right)}(1-t^2)^{(n-4)/2}I_{(-1,1)}(t).$$

(iii) Under the conditions of part (ii), show that the result in (i) is the same as that obtained by applying Scheffé's theorem to the density of $\sqrt{n}T$.
Solution. (i) Consider first the special case of $EY_1 = EZ_1 = 0$ and $\text{Var}(Y_1) = \text{Var}(Z_1) = 1$. Let $W_i = \left(Y_i, Z_i, Y_i^2, Z_i^2, Y_iZ_i\right)$ and $\bar{W} = n^{-1}\sum_{i=1}^n W_i$. Since $W_1, ..., W_n$ are independent and identically distributed and $\text{Var}(W_1)$ is finite under the assumption of $E|Y_1|^4 < \infty$ and $E|Z_1|^4 < \infty$, by the central limit theorem, $\sqrt{n}(\bar{W} - \theta) \to_d N_5(0, \Sigma)$, where $\theta = (0, 0, 1, 1, \rho)$ and

$$\Sigma = \begin{pmatrix} 1 & \rho & E(Y_1^3) & E(Y_1Z_1^2) & E(Y_1^2Z_1) \\ \rho & 1 & E(Y_1^2Z_1) & E(Z_1^3) & E(Y_1Z_1^2) \\ E(Y_1^3) & E(Y_1^2Z_1) & E(Y_1^4)-1 & E(Y_1^2Z_1^2)-1 & E(Y_1^3Z_1)-\rho \\ E(Y_1Z_1^2) & E(Z_1^3) & E(Y_1^2Z_1^2)-1 & E(Z_1^4)-1 & E(Y_1Z_1^3)-\rho \\ E(Y_1^2Z_1) & E(Y_1Z_1^2) & E(Y_1^3Z_1)-\rho & E(Y_1Z_1^3)-\rho & E(Y_1^2Z_1^2)-\rho^2 \end{pmatrix}.$$

Define

$$h(x_1, x_2, x_3, x_4, x_5) = \frac{x_5 - x_1x_2}{\sqrt{(x_3 - x_1^2)(x_4 - x_2^2)}}.$$

Then $T = h(\bar{W})$ and $\rho = h(\theta)$. By the δ-method (e.g., Theorem 1.12 in Shao, 2003), $\sqrt{n}[h(\bar{W}) - h(\theta)] \to_d N(0, c^2)$, where $c^2 = \xi^\tau \Sigma \xi$ and $\xi = \frac{\partial h(w)}{\partial w}\Big|_{w=\theta} = (0, 0, -\rho/2, -\rho/2, 1)$. Hence

$$\begin{aligned} c^2 = {} & \rho^2[E(Y_1^4) + E(Z_1^4) + 2E(Y_1^2Z_1^2)]/4 \\ & - \rho[E(Y_1^3Z_1) + E(Y_1Z_1^3)] + E(Y_1^2Z_1^2). \end{aligned}$$

The result for the general case can be obtained by considering the transformation $(Y_i - EY_i)/\sqrt{\text{Var}(Y_i)}$ and $(Z_i - EZ_i)/\sqrt{\text{Var}(Z_i)}$. The value of c^2 is then given by the previous expression with Y_1 and Z_1 replaced by $(Y_1 - EY_1)/\sqrt{\text{Var}(Y_1)}$ and $(Z_1 - EZ_1)/\sqrt{\text{Var}(Z_1)}$, respectively.
(ii) We only need to consider the case of $\mu_1 = \mu_2 = 0$ and $\sigma_1^2 = \sigma_2^2 = 1$. Let $Y = (Y_1, ..., Y_n)$, $Z = (Z_1, ..., Z_n)$, and A_Z be the n-vector whose ith component is $(Z_i - \bar{Z})/(\sqrt{n-1}S_Z)$. Note that

$$(n-1)S_Y^2 - (A_Z^\tau Y)^2 = Y^\tau B_Z Y$$

with $B_Z = I_n - n^{-1}JJ^\tau - A_ZA_Z^\tau$, where I_n is the identity matrix of order n and J is the n-vector of 1's. Since $A_Z^\tau A_Z = 1$ and $J^\tau A_Z = 0$, $B_ZA_Z = 0$,

$B_Z^2 = B_Z$ and $\text{tr}(B_Z) = n-2$. Consequently, when Z is considered to be a fixed vector, $Y^\tau B_Z Y$ and $A_Z^\tau Y$ are independent, $A_Z^\tau Y$ is distributed as $N(0,1)$, $Y^\tau B_Z Y$ has the χ^2_{n-2} distribution, and $\sqrt{n-2}A_Z^\tau Y/\sqrt{Y^\tau B_Z Y}$ has the t-distribution t_{n-2}. Since $T = A_Z^\tau Y/(\sqrt{n-1}S_Y)$,

$$\begin{aligned} P(T \le t) &= E[P(T \le t|Z)] \\ &= E\left[P\left(\frac{A_Z^\tau Y}{\sqrt{Y^\tau B_Z Y + (A_Z^\tau Y)^2}} \le t \middle| Z\right)\right] \\ &= E\left[P\left(\frac{A_Z^\tau Y}{\sqrt{Y^\tau B_Z Y}} \le \frac{t}{\sqrt{1-t^2}} \middle| Z\right)\right] \\ &= E\left[P\left(t_{n-2} \le \frac{t\sqrt{n-2}}{\sqrt{1-t^2}}\right)\right] \\ &= \frac{\Gamma(\frac{n-1}{2})}{\sqrt{(n-2)\pi}\Gamma(\frac{n-2}{2})}\int_0^{\frac{t\sqrt{n-2}}{\sqrt{1-t^2}}}\left(1+\frac{x^2}{n-2}\right)^{-(n-1)/2}dx, \end{aligned}$$

where t_{n-2} denotes a random variable having the t-distribution t_{n-2} and the third equality follows from the fact that $\frac{a}{\sqrt{a^2+b^2}} \le t$ if and only if $\frac{a}{\sqrt{b^2}} \le \frac{t}{\sqrt{1-t^2}}$ for real numbers a and b and $t \in (0,1)$. Thus, T has Lebesgue density

$$\frac{d}{dt}P(T \le t) = \frac{\Gamma(\frac{n-1}{2})}{\sqrt{\pi}\Gamma(\frac{n-2}{2})}(1-t^2)^{(n-4)/2}I_{(-1,1)}(t).$$

(iii) Under the conditions of part (ii), $\rho = 0$ and, from the result in (i), $c = 1$ and $\sqrt{n}T \to_d N(0,1)$. From the result in (ii), $\sqrt{n}T$ has Lebesgue density

$$\frac{\Gamma(\frac{n-1}{2})}{\sqrt{n\pi}\Gamma(\frac{n-2}{2})}\left(1-\frac{t^2}{n}\right)^{(n-4)/2} I_{(-\sqrt{n},\sqrt{n})}(t) \to \frac{1}{\sqrt{2\pi}}e^{-t^2/2}$$

by Stirling's formula. By Scheffé's Theorem, $\sqrt{n}T \to_d N(0,1)$. ∎

Exercise 10 (#2.23). Let $X_1,...,X_n$ be independent and identically distributed random variables with $EX_1^4 < \infty$, $T = (Y,Z)$, and $T_1 = Y/\sqrt{Z}$, where $Y = n^{-1}\sum_{i=1}^n |X_i|$ and $Z = n^{-1}\sum_{i=1}^n X_i^2$.
(i) Show that $\sqrt{n}(T-\theta) \to_d N_2(0,\Sigma)$ and $\sqrt{n}(T_1-\vartheta) \to_d N(0,c^2)$. Identify θ, Σ, ϑ, and c^2 in terms of moments of X_1.
(ii) Repeat (i) when X_1 has the normal distribution $N(0,\sigma^2)$.
(iii) Repeat (i) when X_1 has Lebesgue density $(2\sigma)^{-1}e^{-|x|/\sigma}$.
Solution. (i) Define $\theta_j = E|X_1|^j$, $j = 1,2,3,4$, and $W_i = (|X_i|, X_i^2)$, $i = 1,...,n$. Then $T = n^{-1}\sum_{i=1}^n W_i$. Let $\theta = EW_1 = (\theta_1,\theta_2)$. By the central limit theorem, $\sqrt{n}(T-\theta) \to_d N_2(0,\Sigma)$, where

$$\Sigma = \begin{pmatrix} \text{Var}(|X_1|) & \text{Cov}(|X_1|,X_1^2) \\ \text{Cov}(|X_1|,X_1^2) & \text{Var}(X_1^2) \end{pmatrix} = \begin{pmatrix} \theta_2-\theta_1^2 & \theta_3-\theta_1\theta_2 \\ \theta_3-\theta_1\theta_2 & \theta_4-\theta_2^2 \end{pmatrix}.$$

Let $g(y,z) = y/\sqrt{z}$. Then $T_1 = g(T)$, $g(\theta) = \theta_1/\sqrt{\theta_2}$, $\frac{\partial g}{\partial y}\big|_{(y,z)=\theta} = 1/\sqrt{\theta_2}$, and $\frac{\partial g}{\partial z}\big|_{(y,z)=\theta} = -\theta_1/(2\theta_2^{3/2})$. Then, by the δ-method, $\sqrt{n}(T_1 - \vartheta) \to_d N(0, c^2)$ with $\vartheta = \theta_1/\sqrt{\theta_2}$ and

$$c^2 = 1 + \frac{\theta_1^2\theta_4}{4\theta_2^3} - \frac{\theta_1\theta_3}{\theta_2^2} - \frac{\theta_1^2}{4\theta_3}.$$

(ii) We only need to calculate θ_j. When X_1 is distributed as $N(0,\sigma^2)$, a direct calculation shows that $\theta_1 = \sqrt{2}\sigma/\sqrt{\pi}$, $\theta_2 = \sigma^2$, $\theta_3 = 2\sqrt{2}\sigma^3/\sqrt{\pi}$, and $\theta_4 = 3\sigma^4$.
(iii) Note that $|X_1|$ has the exponential distribution with Lebesgue density $\sigma^{-1}e^{-x/\sigma}I_{(0,\infty)}(x)$. Hence, $\theta_j = \sigma^j j!$. ∎

Exercise 11 (#2.25). Let X be a sample from $P \in \mathcal{P}$, where $\mathcal{P}$ is a family of distributions on the Borel σ-field on $\mathcal{R}^n$. Show that if $T(X)$ is a sufficient statistic for $P \in \mathcal{P}$ and $T = \psi(S)$, where ψ is measurable and $S(X)$ is another statistic, then $S(X)$ is sufficient for $P \in \mathcal{P}$.
Solution. Assume first that all P in $\mathcal{P}$ are dominated by a σ-finite measure ν. Then, by the factorization theorem (e.g., Theorem 2.2 in Shao, 2003),

$$\frac{dP}{d\nu}(x) = g_P(T(x))h(x),$$

where h is a Borel function of x (not depending on P) and $g_P(t)$ is a Borel function of t. If $T = \psi(S)$, then

$$\frac{dP}{d\nu}(x) = g_P(\psi(S(x)))h(x)$$

and, by the factorization theorem again, $S(X)$ is sufficient for $P \in \mathcal{P}$.

Consider the general case. Suppose that $S(X)$ is not sufficient for $P \in \mathcal{P}$. By definition, there exist at least two measures $P_1 \in \mathcal{P}$ and $P_2 \in \mathcal{P}$ such that the conditional distributions of X given $S(X)$ under P_1 and P_2 are different. Let $\mathcal{P}_0 = \{P_1, P_2\}$, which is a sub-family of $\mathcal{P}$. Since $T(X)$ is sufficient for $P \in \mathcal{P}$, it is also sufficient for $P \in \mathcal{P}_0$. Since all P in $\mathcal{P}_0$ are dominated by the measure $P_1 + P_2$, by the previously proved result, $S(X)$ is sufficient for $P \in \mathcal{P}_0$. Hence, the conditional distributions of X given $S(X)$ under P_1 and P_2 are the same. This contradiction proves that $S(X)$ is sufficient for $P \in \mathcal{P}$. ∎

Exercise 12. Let $\mathcal{P} = \{f_\theta : \theta \in \Theta\}$, where f_θ's are probability densities, $f_\theta(x) > 0$ for all $x \in \mathcal{R}$ and, for any $\theta \in \Theta$, $f_\theta(x)$ is continuous in x. Let X_1 and X_2 be independent and identically distributed as f_θ. Show that if $X_1 + X_2$ is sufficient for θ, then $\mathcal{P}$ is an exponential family indexed by θ.
Solution. The joint density of X_1 and X_2 is $f_\theta(x_1)f_\theta(x_2)$. By the factorization theorem, there exist functions $g_\theta(t)$ and $h(x_1, x_2)$ such that

$$f_\theta(x_1)f_\theta(x_2) = g_\theta(x_1 + x_2)h(x_1, x_2).$$

Then

$$\log f_\theta(x_1) + \log f_\theta(x_2) = g(x_1 + x_2, \theta) + h_1(x_1, x_2),$$

where $g(t,\theta) = \log g_\theta(t)$ and $h_1(x_1,x_2) = \log h(x_1,x_2)$. Let $\theta_0 \in \Theta$ and $r(x,\theta) = \log f_\theta(x) - \log f_{\theta_0}(x)$ and $q(x,\theta) = g(x,\theta) - g(x,\theta_0)$. Then

$$\begin{aligned} q(x_1 + x_2, \theta) &= \log f_\theta(x_1) + \log f_\theta(x_2) + h_1(x_1,x_2) \\ &\quad - \log f_{\theta_0}(x_1) - \log f_{\theta_0}(x_2) - h_1(x_1,x_2) \\ &= r(x_1,\theta) + r(x_2,\theta). \end{aligned}$$

Consequently,

$$r(x_1 + x_2, \theta) + r(0,\theta) = q(x_1 + x_2, \theta) = r(x_1,\theta) + r(x_2,\theta)$$

for any x_1, x_2, and θ. Let $s(x,\theta) = r(x,\theta) - r(0,\theta)$. Then

$$s(x_1,\theta) + s(x_2,\theta) = s(x_1 + x_2, \theta)$$

for any x_1, x_2, and θ. Hence,

$$s(n,\theta) = ns(1,\theta) \quad n = 0, \pm 1, \pm 2, ...$$

For any rational number $\frac{n}{m}$ (n and m are integers and $m \neq 0$),

$$s\left(\tfrac{n}{m},\theta\right) = ns\left(\tfrac{1}{m},\theta\right) = \tfrac{m}{m} ns\left(\tfrac{1}{m},\theta\right) = \tfrac{n}{m} s\left(\tfrac{m}{m},\theta\right) = \tfrac{n}{m} s\left(1,\theta\right).$$

Hence $s(x,\theta) = xs(1,\theta)$ for any rational x. From the continuity of f_θ, we conclude that $s(x,\theta) = xs(1,\theta)$ for any $x \in \mathcal{R}$, i.e.,

$$r(x,\theta) = s(1,\theta)x + r(0,\theta)$$

any $x \in \mathcal{R}$. Then, for any x and θ,

$$\begin{aligned} f_\theta(x) &= \exp\{r(x,\theta) + \log f_{\theta_0}(x)\} \\ &= \exp\{s(1,\theta)x + r(0,\theta) + \log f_{\theta_0}(x)\} \\ &= \exp\{\eta(\theta)x - \xi(\theta)\}h(x), \end{aligned}$$

where $\eta(\theta) = s(1,\theta)$, $\xi(\theta) = -r(0,\theta)$, and $h(x) = f_{\theta_0}(x)$. This shows that $\mathcal{P}$ is an exponential family indexed by θ. ∎

Exercise 13 (#2.30). Let X and Y be two random variables such that Y has the binomial distribution with size N and probability π and, given $Y = y$, X has the binomial distribution with size y and probability p.
(i) Suppose that $p \in (0,1)$ and $\pi \in (0,1)$ are unknown and N is known. Show that (X,Y) is minimal sufficient for (p,π).
(ii) Suppose that π and N are known and $p \in (0,1)$ is unknown. Show

whether X is sufficient for p and whether Y is sufficient for p.
Solution. (i) Let $A = \{(x,y) : x = 0,1,...,y,\ y = 0,1,...,N\}$. The joint probability density of (X,Y) with respect to the counting measure is

$$\binom{N}{y}\pi^y(1-\pi)^{N-y}\binom{y}{x}p^x(1-p)^{y-x}I_A$$
$$= \exp\left\{x\log\frac{p}{1-p} + y\log\frac{\pi(1-p)}{1-\pi} + N\log(1-\pi)\right\}\binom{N}{y}\binom{y}{x}I_A.$$

Hence, (X,Y) has a distribution from an exponential family of full rank ($0 < p < 1$ and $0 < \pi < 1$). This implies that (X,Y) is minimal sufficient for (p,π).
(ii) The joint probability density of (X,Y) can be written as

$$\exp\left\{x\log\frac{p}{1-p} + y\log(1-p)\right\}\pi^y(1-\pi)^{N-y}\binom{N}{y}\binom{y}{x}I_A.$$

This is from an exponential family not of full rank. Let $p_0 = \frac{1}{2}$, $p_1 = \frac{1}{3}$, $p_2 = \frac{2}{3}$, and $\eta(p) = (\log\frac{p}{1-p}, \log(1-p))$. Then, two vectors in $\mathcal{R}^2$, $\eta(p_1) - \eta(p_0) = (-\log 2, 2\log 2 - \log 3)$ and $\eta(p_2) - \eta(p_0) = (\log 2, \log 2 - \log 3)$, are linearly independent. By the properties of exponential families (e.g., Example 2.14 in Shao, 2003), (X,Y) is minimal sufficient for p. Thus, neither X nor Y is sufficient for p. ∎

Exercise 14 (#2.34). Let $X_1,...,X_n$ be independent and identically distributed random variables having the Lebesgue density

$$\exp\left\{-\left(\tfrac{x-\mu}{\sigma}\right)^4 - \xi(\theta)\right\},$$

where $\theta = (\mu,\sigma) \in \Theta = \mathcal{R} \times (0,\infty)$. Show that $\mathcal{P} = \{P_\theta : \theta \in \Theta\}$ is an exponential family, where P_θ is the joint distribution of $X_1,...,X_n$, and that the statistic $T = \left(\sum_{i=1}^n X_i, \sum_{i=1}^n X_i^2, \sum_{i=1}^n X_i^3, \sum_{i=1}^n X_i^4\right)$ is minimal sufficient for $\theta \in \Theta$.
Solution. Let $T(x) = (\sum_{i=1}^n x_i, \sum_{i=1}^n x_i^2, \sum_{i=1}^n x_i^3, \sum_{i=3}^n x_i^4)$ for any $x = (x_1,...,x_n)$ and let $\eta(\theta) = \sigma^{-4}(-4\mu^3, 6\mu^2, -4\mu, 1)$. The joint density of $(X_1,...,X_n)$ is

$$f_\theta(x) = \exp\left\{[\eta(\theta)]^\tau T(x) - n\mu^4/\sigma^4 - n\xi(\theta)\right\},$$

which belongs to an exponential family. For any two sample points $x = (x_1,...,x_n)$ and $y = (y_1,...,y_n)$,

$$\frac{f_\theta(x)}{f_\theta(y)} = \exp\left\{-\frac{1}{\sigma^4}\left[\left(\sum_{i=1}^n x_i^4 - \sum_{i=1}^n y_i^4\right) - 4\mu\left(\sum_{i=1}^n x_i^3 - \sum_{i=1}^n y_i^3\right)\right.\right.$$
$$\left.\left. +\ 6\mu^2\left(\sum_{i=1}^n x_i^2 - \sum_{i=1}^n y_i^2\right) - 4\mu^3\left(\sum_{i=1}^n x_i - \sum_{i=1}^n y_i\right)\right]\right\},$$

which is free of parameter (μ, σ) if and only if $T(x) = T(y)$. By Theorem 2.3(iii) in Shao (2003), $T(X)$ is minimal sufficient for θ. ∎

Exercise 15 (#2.35). Let $(X_1, ..., X_n)$ be a random sample of random variables having the Lebesgue density $f_\theta(x) = (2\theta)^{-1}\big[I_{(0,\theta)}(x) + I_{(2\theta,3\theta)}(x)\big]$. Find a minimal sufficient statistic for $\theta \in (0, \infty)$.

Solution. We use the idea of Theorem 2.3(i)-(ii) in Shao (2003). Let $\Theta_r = \{\theta_1, \theta_2, ...\}$ be the set of positive rational numbers, $\mathcal{P}_0 = \{g_\theta : \theta \in \Theta_r\}$, and $\mathcal{P} = \{g_\theta : \theta > 0\}$, where $g_\theta(x) = \prod_{i=1}^n f_\theta(x_i)$ for $x = (x_1, ..., x_n)$. Then $\mathcal{P}_0 \subset \mathcal{P}$ and a.s. $\mathcal{P}_0$ implies a.s. $\mathcal{P}$ (i.e., if an event A satisfying $P(A) = 0$ for all $P \in \mathcal{P}_0$, then $P(A) = 0$ for all $P \in \mathcal{P}$). Let $\{c_i\}$ be a sequence of positive numbers satisfying $\sum_{i=1}^\infty c_i = 1$ and $g_\infty(x) = \sum_{i=1}^\infty c_i g_{\theta_i}(x)$. Define $T = (T_1, T_2, ...)$ with $T_i(x) = g_{\theta_i}(x)/g_\infty(x)$. By Theorem 2.3(ii) in Shao (2003), T is minimal sufficient for $\theta \in \Theta_0$ (or $P \in \mathcal{P}_0$). For any $\theta > 0$, there is a sequence $\{\theta_{i_k}\} \subset \{\theta_i\}$ such that $\lim_k \theta_{i_k} = \theta$. Then

$$g_\theta(x) = \lim_k g_{\theta_{i_k}}(x) = \lim_k T_{i_k}(x) g_\infty(x)$$

holds for all $x \in C$ with $P(C) = 1$ for all $P \in \mathcal{P}$. By the factorization theorem, T is sufficient for $\theta > 0$ (or $P \in \mathcal{P}$). By Theorem 2.3(i) in Shao (2003), T is minimal sufficient for $\theta > 0$. ∎

Exercise 16 (#2.36). Let $(X_1, ..., X_n)$ be a random sample of random variables having the Cauchy distribution with location parameter μ and scale parameter σ, where $\mu \in \mathcal{R}$ and $\sigma > 0$ are unknown parameters. Show that the vector of order statistics is minimal sufficient for (μ, σ).

Solution. The joint Lebesgue density of $(X_1, ..., X_n)$ is

$$f_{\mu,\sigma}(x) = \frac{\sigma^n}{\pi^n} \prod_{i=1}^n \frac{1}{\sigma^2 + (x_i - \mu)^2}, \quad x = (x_1, ..., x_n).$$

For any $x = (x_1, ..., x_n)$ and $y = (y_1, ..., y_n)$, suppose that

$$\frac{f_{\mu,\sigma}(x)}{f_{\mu,\sigma}(y)} = \psi(x, y)$$

holds for any μ and σ, where ψ does not depend on (μ, σ). Let $\sigma = 1$. Then we must have

$$\prod_{i=1}^n \left[1 + (y_i - \mu)^2\right] = \psi(x, y) \prod_{i=1}^n \left[1 + (x_i - \mu)^2\right]$$

for all μ. Both sides of the above identity can be viewed as polynomials of degree $2n$ in μ. Comparison of the coefficients to the highest terms gives

$\psi(x, y) = 1$. Thus,

$$\prod_{i=1}^{n} \left[1 + (y_i - \mu)^2\right] = \prod_{i=1}^{n} \left[1 + (x_i - \mu)^2\right]$$

for all μ. As a polynomial of μ, the left-hand side of the above identity has $2n$ complex roots $x_i \pm \sqrt{-1}$, $i = 1, ..., n$, while the right-hand side of the above identity has $2n$ complex roots $y_i \pm \sqrt{-1}$, $i = 1, ..., n$. By the unique factorization theorem for the entire functions in complex analysis, we conclude that the two sets of roots must agree. This means that the ordered values of x_i's are the same as the ordered values of y_i's. By Theorem 2.3(iii) in Shao (2003), the order statistics of $X_1, ..., X_n$ is minimal sufficient for (μ, σ). ∎

Exercise 17 (#2.40). Let $(X_1, ..., X_n)$, $n \geq 2$, be a random sample from a distribution having Lebesgue density $f_{\theta,j}$, where $\theta > 0$, $j = 1, 2$, $f_{\theta,1}$ is the density of $N(0, \theta^2)$, and $f_{\theta,2}(x) = (2\theta)^{-1} e^{-|x|/\theta}$. Show that $T = (T_1, T_2)$ is minimal sufficient for (θ, j), where $T_1 = \sum_{i=1}^{n} X_i^2$ and $T_2 = \sum_{i=1}^{n} |X_i|$.
Solution A. Let P be the joint distribution of $X_1, ..., X_n$. By the factorization theorem, T is sufficient for (θ, j). Let $\mathcal{P} = \{P : \theta > 0, j = 1, 2\}$, $\mathcal{P}_1 = \{P : \theta > 0, j = 1\}$, and $\mathcal{P}_2 = \{P : \theta > 0, j = 2\}$. Let S be a statistic sufficient for $P \in \mathcal{P}$. Then S is sufficient for $P \in \mathcal{P}_j$, $j = 1, 2$. Note that $\mathcal{P}_1$ is an exponential family with T_1 as a minimal sufficient statistic. Hence, there exists a Borel function ψ_1 such that $T_1 = \psi_1(S)$ a.s. $\mathcal{P}_1$. Since all densities in $\mathcal{P}$ are dominated by those in $\mathcal{P}_1$, we conclude that $T_1 = \psi_1(S)$ a.s. $\mathcal{P}$. Similarly, $\mathcal{P}_2$ is an exponential family with T_2 as a minimal sufficient statistic and, thus, there exists a Borel function ψ_2 such that $T_2 = \psi_2(S)$ a.s. $\mathcal{P}$. This proves that $T = (\psi_1(S), \psi_2(S))$ a.s. $\mathcal{P}$. Hence T is minimal sufficient for (θ, j).
Solution B. Let P be the joint distribution of $X_1, ..., X_n$. The Lebesgue density of P can be written as

$$\exp\left\{-\frac{I_{\{1\}}(j)}{2\theta^2} T_1 - \frac{I_{\{2\}}(j)}{\theta} T_2\right\} \left[\frac{I_{\{1\}}(j)}{(2\pi\theta^2)^{n/2}} + \frac{I_{\{2\}}(j)}{(2\theta)^n}\right].$$

Hence $\mathcal{P} = \{P : \theta > 0, j = 1, 2\}$ is an exponential family. Let

$$\eta(\theta, j) = -\left(\frac{I_{\{1\}}(j)}{2\theta^2}, \frac{I_{\{2\}}(j)}{\theta}\right).$$

Note that $\eta(1, 1) = (-\frac{1}{2}, 0)$, $\eta(2^{-1/2}, 1) = (-1, 0)$, and $\eta(1, 2) = (0, -1)$. Then, $\eta(2^{-1/2}, 1) - \eta(1, 1) = (-\frac{1}{2}, 0)$ and $\eta(1, 2) - \eta(1, 1) = (\frac{1}{2}, -1)$ are two linearly independent vectors in $\mathcal{R}^2$. Hence $T = (T_1, T_2)$ is minimal sufficient for (θ, j) (e.g., Example 2.14 in Shao, 2003). ∎

Exercise 18 (#2.41). Let $(X_1, ..., X_n)$, $n \geq 2$, be a random sample from a distribution with discrete probability density $f_{\theta,j}$, where $\theta \in (0,1)$, $j = 1, 2$, $f_{\theta,1}$ is the Poisson distribution with mean θ, and $f_{\theta,2}$ is the binomial distribution with size 1 and probability θ.
(i) Show that $T = \sum_{i=1}^n X_i$ is not sufficient for (θ, j).
(ii) Find a two-dimensional minimal sufficient statistic for (θ, j).
Solution. (i) To show that T is not sufficient for (θ, j), it suffices to show that, for some $x \leq t$, $P(X_n = x|T = t)$ for $j = 1$ is different from $P(X_n = x|T = t)$ for $j = 2$. When $j = 1$,

$$P(X_n = x|T = t) = \binom{t}{x} \frac{(n-1)^{t-x}}{n^t} > 0,$$

whereas when $j = 2$, $P(X_n = x|T = t) = 0$ as long as $x > 1$.
(ii) Let $g_{\theta,j}$ be the joint probability density of $X_1, ..., X_n$. Let $\mathcal{P}_0 = \{g_{\frac{1}{4},1}, g_{\frac{1}{2},1}, g_{\frac{1}{2},2}\}$. Then, a.s. $\mathcal{P}_0$ implies a.s. $\mathcal{P}$. By Theorem 2.3(ii) in Shao (2003), the two-dimensional statistic

$$S = \left(\frac{g_{\frac{1}{2},1}}{g_{\frac{1}{4},1}}, \frac{g_{\frac{1}{2},2}}{g_{\frac{1}{4},1}} \right) = \left(e^{n/4} 2^{-T}, e^{n/2} W 2^{T-n} \right)$$

is minimal sufficient for the family $\mathcal{P}_0$, where

$$W = \begin{cases} 1 & X_i = 0 \text{ or } 1, i = 1, ..., n, \\ 0 & \text{otherwise.} \end{cases}$$

Since there is a one-to-one transformation between S and (T, W), we conclude that (T, W) is minimal sufficient for the family $\mathcal{P}_0$. For any $x = (x_1, ..., x_n)$, the joint density of $X_1, ..., X_n$ is

$$e^{n\theta I_{\{1\}}(j)} (1-\theta)^{n I_{\{2\}}(j)} W^{I_{\{2\}}(j)} e^{T[I_{\{1\}}(j) \log \theta + I_{\{2\}}(j) \log \frac{\theta}{(1-\theta)}]} \prod_{i=1}^n \frac{1}{x_i!}.$$

Hence, by the factorization theorem, (T, W) is sufficient for (θ, j). By Theorem 2.3(i) in Shao (2003), (T, W) is minimal sufficient for (θ, j). ∎

Exercise 19 (#2.44). Let $(X_1, ..., X_n)$ be a random sample from a distribution on $\mathcal{R}$ having the Lebesgue density $\theta^{-1} e^{-(x-\theta)/\theta} I_{(\theta,\infty)}(x)$, where $\theta > 0$ is an unknown parameter.
(i) Find a statistic that is minimal sufficient for θ.
(ii) Show whether the minimal sufficient statistic in (i) is complete.
Solution. (i) Let $T(x) = \sum_{i=1}^n x_i$ and $W(x) = \min_{1 \leq i \leq n} x_i$, where $x = (x_1, ..., x_n)$. The joint density of $X = (X_1, ..., X_n)$ is

$$f_\theta(x) = \frac{e^n}{\theta^n} e^{-T(x)/\theta} I_{(\theta,\infty)}(W(x)).$$

For $x = (x_1, ..., x_n)$ and $y = (y_1, ..., y_n)$,

$$\frac{f_\theta(x)}{f_\theta(y)} = e^{[T(y)-T(x)]/\theta} \frac{I_{(\theta,\infty)}(W(x))}{I_{(\theta,\infty)}(W(y))}$$

is free of θ if and only if $T(x) = T(y)$ and $W(x) = W(y)$. Hence, the two-dimensional statistic $(T(X), W(X))$ is minimal sufficient for θ.
(ii) A direct calculation shows that, for any θ, $E[T(X)] = 2n\theta$ and $E[W(X)] = (1 + n^{-1})\theta$. Hence $E[(2n)^{-1}T - (1 + n^{-1})^{-1}W(X)] = 0$ for any θ and $(2n)^{-1}T - (1 + n^{-1})^{-1}W(X)$ is not a constant. Thus, (T, W) is not complete. ∎

Exercise 20 (#2.48). Let T be a complete (or boundedly complete) and sufficient statistic. Suppose that there is a minimal sufficient statistic S. Show that T is minimal sufficient and S is complete (or boundedly complete).
Solution. We prove the case when T is complete. The case in which T is boundedly complete is similar. Since S is minimal sufficient and T is sufficient, there exists a Borel function h such that $S = h(T)$ a.s. Since h cannot be a constant function and T is complete, we conclude that S is complete. Consider $T - E(T|S) = T - E[T|h(T)]$, which is a Borel function of T and hence can be denoted as $g(T)$. Note that $E[g(T)] = 0$. By the completeness of T, $g(T) = 0$ a.s., that is, $T = E(T|S)$ a.s. This means that T is also a function of S and, therefore, T is minimal sufficient. ∎

Exercise 21 (#2.53). Let X be a discrete random variable with probability density

$$f_\theta(x) = \begin{cases} \theta & x = 0 \\ (1-\theta)^2\theta^{x-1} & x = 1, 2, ... \\ 0 & \text{otherwise,} \end{cases}$$

where $\theta \in (0, 1)$. Show that X is boundedly complete, but not complete.
Solution. Consider any Borel function $h(x)$ such that

$$E[h(X)] = h(0)\theta + \sum_{x=1}^{\infty} h(x)(1-\theta)^2\theta^{x-1} = 0$$

for any $\theta \in (0, 1)$. Rewriting the left-hand side of the above equation in the ascending order of the powers of θ, we obtain that

$$h(1) + \sum_{x=1}^{\infty} [h(x-1) - 2h(x) + h(x+1)]\,\theta^x = 0$$

for any $\theta \in (0, 1)$. Comparing the coefficients of both sides, we obtain that $h(1) = 0$ and $h(x-1) - h(x) = h(x) - h(x+1)$. Therefore, $h(x) = (1-x)h(0)$

for $x = 1, 2, \ldots$. This function is bounded if and only if $h(0) = 0$. If $h(x)$ is assumed to be bounded, then $h(0) = 0$ and, hence, $h(x) \equiv 0$. This means that X is boundedly complete. For $h(x) = 1 - x$, $E[h(X)] = 0$ for any θ but $h(X) \neq 0$. Therefore, X is not complete. ∎

Exercise 22. Let X be a discrete random variable with

$$P_\theta(X = x) = \frac{\binom{\theta}{x}\binom{N-\theta}{n-x}}{\binom{N}{n}}, \quad x = 0, 1, 2, \ldots, \min\{\theta, n\}, n - x \leq N - \theta,$$

where n and N are positive integers, $N \geq n$, and $\theta = 0, 1, \ldots, N$. Show that X is complete.

Solution. Let $g(x)$ be a function of $x \in \{0, 1, \ldots, n\}$. Assume $E_\theta[g(X)] = 0$ for any θ, where E_θ is the expectation with respect to P_θ. When $\theta = 0$, $P_0(X = x) = 1$ if $x = 0$ and $E_0[g(X)] = g(0)$. Thus, $g(0) = 0$. When $\theta = 1$, $P_1(X \geq 2) = 0$ and

$$E_1[g(X)] = g(0)P_1(X = 0) + g(1)P_1(X = 1) = g(1)\frac{\binom{N-1}{n-1}}{\binom{N}{n}}.$$

Since $E_1[g(X)] = 0$, we obtain that $g(1) = 0$. Similarly, we can show that $g(2) = \cdots = g(n) = 0$. Hence X is complete. ∎

Exercise 23. Let X be a random variable having the uniform distribution on the interval $(\theta, \theta + 1)$, $\theta \in \mathcal{R}$. Show that X is not complete.

Solution. Consider $g(X) = \cos(2\pi X)$. Then $g(X) \neq 0$ but

$$E[g(X)] = \int_\theta^{\theta+1} \cos(2\pi x)dx = \frac{\sin(2\pi(\theta + 1)) - \sin(2\pi\theta)}{2\pi} = 0$$

for any θ. Hence X is not complete. ∎

Exercise 24 (#2.57). Let $(X_1, \ldots, X_n)$ be a random sample from the $N(\theta, \theta^2)$ distribution, where $\theta > 0$ is a parameter. Find a minimal sufficient statistic for θ and show whether it is complete.

Solution. The joint Lebesgue density of $X_1, \ldots, X_n$ is

$$\frac{1}{(2\pi\theta^2)^n} \exp\left\{-\frac{1}{2\theta^2}\sum_{i=1}^n x_i^2 + \frac{1}{\theta}\sum_{i=1}^n x_i - \frac{1}{2}\right\}.$$

Let

$$\eta(\theta) = \left(-\frac{1}{2\theta^2}, \frac{1}{\theta}\right).$$

Then $\eta(\frac{1}{2}) - \eta(1) = (-\frac{3}{2}, 1)$ and $\eta(\frac{1}{\sqrt{2}}) - \eta(1) = (-\frac{1}{2}, \sqrt{2})$ are linearly independent vectors in $\mathcal{R}^2$. Hence $T = (\sum_{i=1}^n X_i^2, \sum_{i=1}^n X_i)$ is minimal sufficient for θ. Note that

$$E\left(\sum_{i=1}^n X_i^2\right) = nEX_1^2 = 2n\theta^2$$

and

$$E\left(\sum_{i=1}^n X_i\right)^2 = n\theta^2 + (n\theta)^2 = (n+n^2)\theta^2.$$

Let $h(t_1,t_2) = \frac{1}{2n}t_1 - \frac{1}{n(n+1)}t_2^2$. Then $h(t_1,t_2) \neq 0$ but $E[h(T)] = 0$ for any θ. Hence T is not complete. ∎

Exercise 25 (#2.56). Suppose that $(X_1,Y_1),...,(X_n,Y_n)$ are independent and identically distributed random 2-vectors and X_i and Y_i are independently distributed as $N(\mu,\sigma_X^2)$ and $N(\mu,\sigma_Y^2)$, respectively, with $\theta = (\mu,\sigma_X^2,\sigma_Y^2) \in \mathcal{R}\times(0,\infty)\times(0,\infty)$. Let $\bar{X}$ and S_X^2 be the sample mean and variance for X_i's and $\bar{Y}$ and S_Y^2 be the sample mean and variance for Y_i's. Show that $T = (\bar{X},\bar{Y},S_X^2,S_Y^2)$ is minimal sufficient for θ but T is not boundedly complete.
Solution. Let

$$\eta = \left(-\frac{1}{2\sigma_X^2}, \frac{\mu}{\sigma_X^2}, -\frac{1}{2\sigma_Y^2}, \frac{\mu}{\sigma_Y^2}\right)$$

and

$$S = \left(\sum_{i=1}^n X_i^2, \sum_{i=1}^n Y_i^2, \sum_{i=1}^n X_i, \sum_{i=1}^n Y_i\right).$$

Then the joint Lebesgue density of $(X_1,Y_1),...,(X_n,Y_n)$ is

$$\frac{1}{(2\pi)^n}\exp\left\{\eta^\tau S - \frac{n\mu^2}{2\sigma_X^2} - \frac{n\mu^2}{2\sigma_Y^2} - n\log(\sigma_X\sigma_Y)\right\}.$$

Since the parameter space $\{\eta : \mu \in \mathcal{R}, \sigma_X^2 > 0, \sigma_Y^2 > 0\}$ is a three-dimensional curved hyper-surface in $\mathcal{R}^4$, we conclude that S is minimal sufficient. Note that there is a one-to-one correspondence between T and S. Hence T is also minimal sufficient.

To show that T is not boundedly complete, consider $h(T) = I_{\{\bar{X}>\bar{Y}\}} - \frac{1}{2}$. Then $|h(T)| \leq 0.5$ and $E[h(T)] = 0$ for any η, but $h(T) \neq 0$. Hence T is not boundedly complete. ∎

Exercise 26 (#2.58). Suppose that $(X_1,Y_1),...,(X_n,Y_n)$ are independent and identically distributed random 2-vectors having the normal distribution with $EX_1 = EY_1 = 0$, $\text{Var}(X_1) = \text{Var}(Y_1) = 1$, and $\text{Cov}(X_1,Y_1) = \theta \in (-1,1)$.
(i) Find a minimal sufficient statistic for θ.
(ii) Show whether the minimal sufficient statistic in (i) is complete or not.

(iii) Prove that $T_1 = \sum_{i=1}^n X_i^2$ and $T_2 = \sum_{i=1}^n Y_i^2$ are both ancillary but (T_1, T_2) is not ancillary.
Solution. (i) The joint Lebesgue density of $(X_1, Y_1), ..., (X_n, Y_n)$ is

$$\left(\frac{1}{2\pi\sqrt{1-\theta^2}}\right)^n \exp\left\{-\frac{1}{1-\theta^2}\sum_{i=1}^n (x_i^2 + y_i^2) + \frac{2\theta}{1-\theta^2}\sum_{i=1}^n x_i y_i\right\}.$$

Let

$$\eta = \left(-\frac{1}{1-\theta^2}, \frac{2\theta}{1-\theta^2}\right).$$

The parameter space $\{\eta : -1 < \theta < 1\}$ is a curve in $\mathcal{R}^2$. Therefore, $(\sum_{i=1}^n (X_i^2 + Y_i^2), \sum_{i=1}^n X_i Y_i)$ is minimal sufficient.
(ii) Note that $E[\sum_{i=1}^n (X_i^2 + Y_i^2)] - 2n = 0$, but $\sum_{i=1}^n (X_i^2 + Y_i^2) - 2n \neq 0$. Therefore, the minimal sufficient statistic is not complete.
(iii) Both T_1 and T_2 have the chi-square distribution χ_n^2, which does not depend on θ. Hence both T_1 and T_2 are ancillary. Note that

$$\begin{aligned} E(T_1 T_2) &= E\left(\sum_{i=1}^n X_i^2\right)\left(\sum_{j=1}^n Y_j^2\right) \\ &= E\left(\sum_{i=1}^n X_i^2 Y_i^2\right) + E\left(\sum_{i\neq j} X_i^2 Y_j^2\right) \\ &= nE(X_1^2 Y_1^2) + n(n-1)E(X_1^2)E(Y_1^2) \\ &= n(1 + 2\theta^2) + 2n(n-1), \end{aligned}$$

which depends on θ. Therefore the distribution of (T_1, T_2) depends on θ and (T_1, T_2) is not ancillary. ∎

Exercise 27 (#2.59). Let $(X_1, ..., X_n)$, $n > 2$, be a random sample from the exponential distribution on (a, ∞) with scale parameter θ. Show that
(i) $\sum_{i=1}^n (X_i - X_{(1)})$ and $X_{(1)}$ are independent for any (a, θ), where $X_{(j)}$ is the jth order statistic;
(ii) $Z_i = (X_{(n)} - X_{(i)})/(X_{(n)} - X_{(n-1)})$, $i = 1, ..., n-2$, are independent of $(X_{(1)}, \sum_{i=1}^n (X_i - X_{(1)}))$.
Solution: (i) Let θ be arbitrarily fixed. Since the joint density of $X_1, ..., X_n$ is

$$\theta^{-n} e^{na/\theta} \exp\left\{-\frac{1}{\theta}\sum_{i=1}^n x_i\right\} I_{(a,\infty)}(x_{(1)}),$$

where only a is considered as an unknown parameter, we conclude that $X_{(1)}$ is sufficient for a. Note that $\frac{n}{\theta} e^{-n(x-a)/\theta} I_{(a,\infty)}(x)$ is the Lebesgue density for $X_{(1)}$. For any Borel function g,

$$E[g(X_{(1)})] = \frac{n}{\theta}\int_a^\infty g(x)e^{-n(x-a)/\theta}dx = 0$$

for any a is equivalent to

$$\int_a^\infty g(x)e^{-nx/\theta}dx = 0$$

for any a, which implies $g(x) = 0$ a.e. with respect to Lebesgue measure. Hence, for any fixed θ, $X_{(1)}$ is sufficient and complete for a. The Lebesgue density of $X_i - a$ is $\theta^{-1}e^{-x/\theta}I_{(0,\infty)}(x)$, which does not depend on a. Therefore, for any fixed θ, $\sum_{i=1}^n(X_i - X_{(1)}) = \sum_{i=1}^n[(X_i - a) - (X_{(1)} - a)]$ is ancillary. By Basu's theorem (e.g., Theorem 2.4 in Shao, 2003), $\sum_{i=1}^n(X_i - X_{(1)})$ and $X_{(1)}$ are independent for any fixed θ. Since θ is arbitrary, we conclude that $\sum_{i=1}^n(X_i - X_{(1)})$ and $X_{(1)}$ are independent for any (a, θ).
(ii) From Example 5.14 in Lehmann (1983, p. 47), $(X_{(1)}, \sum_{i=1}^n(X_i - X_{(1)}))$ is sufficient and complete for (a, θ). Note that $(X_i - a)/\theta$ has Lebesgue density $e^{-x}I_{(0,\infty)}(x)$, which does not depend on (a, θ). Since

$$Z_i = \frac{X_{(n)} - X_{(i)}}{X_{(n)} - X_{(n-1)}} = \frac{\frac{X_{(n)}-a}{\theta} - \frac{X_{(i)}-a}{\theta}}{\frac{X_{(n)}-a}{\theta} - \frac{X_{(n-1)}-a}{\theta}},$$

the statistic $(Z_1, ..., Z_{n-2})$ is ancillary. By Basu's Theorem, $(Z_1, ..., Z_{n-2})$ is independent of $\left(X_{(1)}, \sum_{i=1}^n(X_i - X_{(1)})\right)$. ∎

Exercise 28 (#2.61). Let $(X_1, ..., X_n)$, $n > 2$, be a random sample of random variables having the uniform distribution on the interval $[a, b]$, where $-\infty < a < b < \infty$. Show that $Z_i = (X_{(i)} - X_{(1)})/(X_{(n)} - X_{(1)})$, $i = 2, ..., n-1$, are independent of $(X_{(1)}, X_{(n)})$ for any a and b, where $X_{(j)}$ is the jth order statistic.
Solution. Note that $(X_i - a)/(b - a)$ has the uniform distribution on the interval $[0, 1]$, which does not depend on any (a, b). Since

$$Z_i = \frac{X_{(i)} - X_{(1)}}{X_{(n)} - X_{(1)}} = \frac{\frac{X_{(i)}-a}{b-a} - \frac{X_{(1)}-a}{b-a}}{\frac{X_{(n)}-a}{b-a} - \frac{X_{(1)}-a}{b-a}},$$

the statistic $(Z_2, ..., Z_{n-1})$ is ancillary. By Basu's Theorem, the result follows if $(X_{(1)}, X_{(n)})$ is sufficient and complete for (a, b). The joint Lebesgue density of $X_1, ..., X_n$ is $(b - a)^{-n}I_{\{a<x_{(1)}<x_{(n)}<b\}}$. By the factorization theorem, $(X_{(1)}, X_{(n)})$ is sufficient for (a, b). The joint Lebesgue density of $(X_{(1)}, X_{(n)})$ is

$$\frac{n(n-1)}{(b-a)^n}(y - x)^{n-2}I_{\{a<x<y<b\}}.$$

For any Borel function $g(x,y)$, $E[g(X_{(1)},X_{(n)})]=0$ for any $a<b$ implies that

$$\int_{a<x<y<b} g(x,y)(y-x)^{n-2}dxdy = 0$$

for any $a<b$. Hence $g(x,y)(y-x)^{n-2}=0$ a.e. m^2, where m^2 is the Lebesgue measure on $\mathcal{R}^2$. Since $(y-x)^{n-2}\neq 0$ a.e. m^2, we conclude that $g(x,y)=0$ a.e. m^2. Hence, $(X_{(1)},X_{(n)})$ is complete. ∎

Exercise 29 (#2.62). Let $(X_1,...,X_n)$, $n>2$, be a random sample from a distribution P on $\mathcal{R}$ with $EX_1^2<\infty$, $\bar{X}$ be the sample mean, $X_{(j)}$ be the jth order statistic, and $T=(X_{(1)}+X_{(n)})/2$. Consider the estimation of a parameter $\theta\in\mathcal{R}$ under the squared error loss.
(i) Show that $\bar{X}$ is better than T if $P=N(\theta,\sigma^2)$, $\theta\in\mathcal{R}$, $\sigma>0$.
(ii) Show that T is better than $\bar{X}$ if P is the uniform distribution on the interval $(\theta-\frac{1}{2},\theta+\frac{1}{2})$, $\theta\in\mathcal{R}$.
(iii) Find a family $\mathcal{P}$ for which neither $\bar{X}$ nor T is better than the other.
Solution. (i) Since $\bar{X}$ is complete and sufficient for θ and $T-\bar{X}$ is ancillary to θ, by Basu's theorem, $T-\bar{X}$ and $\bar{X}$ are independent. Then

$$R_T(\theta) = E[(T-\bar{X})+(\bar{X}-\theta)]^2 = E(T-\bar{X})^2 + R_{\bar{X}}(\theta) > R_{\bar{X}}(\theta),$$

where the last inequality follows from the fact that $T\neq\bar{X}$ a.s. Therefore $\bar{X}$ is better.
(ii) Let $W=\frac{X_{(1)}-\theta+X_{(n)}-\theta}{2}$. Then the Lebesgue density of W is

$$f(w) = \begin{cases} n2^{n-1}\left(w+\frac{1}{2}\right)^{n-1} & -\frac{1}{2}<w<0 \\ n2^{n-1}\left(\frac{1}{2}-w\right)^{n-1} & 0<w<\frac{1}{2} \\ 0 & \text{otherwise.} \end{cases}$$

Therefore $ET=EW+\theta=\theta$ and

$$R_T(\theta) = \mathrm{Var}(T) = \mathrm{Var}(W) = \frac{1}{2(n+1)(n+2)}.$$

On the other hand,

$$R_{\bar{X}}(\theta) = \mathrm{Var}(\bar{X}) = \frac{\mathrm{Var}(X_1)}{n} = \frac{1}{12n}.$$

Hence, when $n>2$, $R_T(\theta)<R_{\bar{X}}(\theta)$.
(iii) Consider the family $\mathcal{P}=\mathcal{P}_1\cup\mathcal{P}_2$, where $\mathcal{P}_1$ is the family in part (i) and $\mathcal{P}_2$ is the family in part (ii). When $P\in\mathcal{P}_1$, $\bar{X}$ is better than T. When $P\in\mathcal{P}_2$, T is better than $\bar{X}$. Therefore, neither of them is better than the other for $P\in\mathcal{P}$. ∎

Exercise 30 (#2.64). Let $(X_1,...,X_n)$ be a random sample of binary random variables with $P(X_i=1)=\theta\in(0,1)$. Consider estimating θ with

the squared error loss. Calculate the risks of the following estimators:
(i) the nonrandomized estimators $\bar{X}$ (the sample mean) and

$$T_0(X) = \begin{cases} 0 & \text{if more than half of } X_i\text{'s are 0} \\ 1 & \text{if more than half of } X_i\text{'s are 1} \\ \frac{1}{2} & \text{if exactly half of } X_i\text{'s are 0;} \end{cases}$$

(ii) the randomized estimators

$$T_1(X) = \begin{cases} \bar{X} & \text{with probability } \frac{1}{2} \\ T_0 & \text{with probability } \frac{1}{2} \end{cases}$$

and

$$T_2(X) = \begin{cases} \bar{X} & \text{with probability } \bar{X} \\ \frac{1}{2} & \text{with probability } 1 - \bar{X}. \end{cases}$$

Solution. (i) Note that

$$\begin{aligned} R_{T_0}(\theta) &= E(T_0 - \theta)^2 \\ &= \theta^2 P(\bar{X} < 0.5) + (1-\theta)^2 P(\bar{X} > 0.5) + (0.5 - \theta)^2 P(\bar{X} = 0.5). \end{aligned}$$

When $n = 2k$,

$$P(\bar{X} < 0.5) = \sum_{j=1}^{k-1} \binom{2k}{j} \theta^j (1-\theta)^{2k-j},$$

$$P(\bar{X} > 0.5) = \sum_{j=k+1}^{2k} \binom{2k}{j} \theta^j (1-\theta)^{2k-j},$$

and

$$P(\bar{X} = 0.5) = \binom{2k}{k} \theta^k (1-\theta)^k.$$

When $n = 2k+1$,

$$P(\bar{X} < 0.5) = \sum_{j=0}^{k} \binom{2k+1}{j} \theta^j (1-\theta)^{2k+1-j},$$

$$P(\bar{X} > 0.5) = \sum_{j=k+1}^{2k+1} \binom{2k+1}{j} \theta^j (1-\theta)^{2k+1-j},$$

and $P(\bar{X} = 0.5) = 0$.
(ii) A direct calculation shows that

$$\begin{aligned} R_{T_1}(\theta) &= E(T_1 - \theta)^2 \\ &= \frac{1}{2} E(\bar{X} - \theta)^2 + \frac{1}{2} E(T_0 - \theta)^2 \\ &= \frac{\theta(1-\theta)}{2n} + \frac{1}{2} R_{T_0}(\theta), \end{aligned}$$

where $R_{T_0}(\theta)$ is given in part (i), and

$$\begin{aligned}
R_{T_2}(\theta) &= E(T_2 - \theta)^2 \\
&= E\left[\bar{X}(\bar{X}-\theta)^2 + \left(\frac{1}{2}-\theta\right)^2 (1-\bar{X})\right] \\
&= E(\bar{X}-\theta)^3 + \theta E(\bar{X}-\theta)^2 + \left(\frac{1}{2}-\theta\right)^2 (1-\theta) \\
&= \frac{1}{n^3}\sum_{i=1}^n\sum_{j=1}^n\sum_{k=1}^n E(X_i-\theta)(X_j-\theta)(X_k-\theta) \\
&\quad + \frac{\theta^2(1-\theta)}{n} + \left(\frac{1}{2}-\theta\right)^2 (1-\theta) \\
&= \frac{E(X_1-\theta)^3}{n^2} + \frac{\theta^2(1-\theta)}{n} + \left(\frac{1}{2}-\theta\right)^2 (1-\theta) \\
&= \frac{\theta(1-\theta)^3 - \theta^3(1-\theta)}{n^2} + \frac{\theta^2(1-\theta)}{n} + \left(\frac{1}{2}-\theta\right)^2 (1-\theta),
\end{aligned}$$

where the fourth equality follows from $E(\bar{X}-\theta)^2 = \mathrm{Var}(\bar{X}) = \theta(1-\theta)/n$ and the fifth equality follows from the fact that $E(X_i-\theta)(X_j-\theta)(X_k-\theta) \neq 0$ if and only if $i = j = k$. ∎

Exercise 31 (#2.66). Consider the estimation of an unknown parameter $\theta \geq 0$ under the squared error loss. Show that if T and U are two estimators such that $T \leq U$ and $R_T(P) < R_U(P)$, then $R_{T_+}(P) < R_{U_+}(P)$, where $R_T(P)$ is the risk of an estimator T and T_+ denotes the positive part of T.
Solution. Note that $T = T_+ - T_-$, where $T_- = \max\{-T, 0\}$ is the negative part of T, and $T_+T_- = 0$. Then

$$\begin{aligned}
R_T(P) &= E(T-\theta)^2 \\
&= E(T_+ - T_- - \theta)^2 \\
&= E(T_+ - \theta)^2 + E(T_-^2) + 2\theta E(T_-) - 2E(T_+T_-) \\
&= R_{T_+}(P) + E(T_-^2) + 2\theta E(T_-).
\end{aligned}$$

Similarly,

$$R_U(P) = R_{U_+}(P) + E(U_-^2) + 2\theta E(U_-).$$

Since $T \leq U$, $T_- \geq U_-$. Also, $\theta \geq 0$. Hence,

$$E(T_-^2) + 2\theta E(T_-) \geq E(U_-^2) + 2\theta E(U_-).$$

Since $R_T(P) < R_U(P)$, we must have $R_{T_+}(P) < R_{U_+}(P)$. ∎

Exercise 32. Consider the estimation of an unknown parameter $\theta \in \mathcal{R}$ under the squared error loss. Show that if T and U are two estimators such that $P(\theta - t < T < \theta + t) \geq P(\theta - t < U < \theta + t)$ for any $t > 0$, then $R_T(P) \leq R_U(P)$.

Solution. From the condition,

$$P((T-\theta)^2 > s) \leq P((U-\theta)^2 > s)$$

for any $s > 0$. Hence,

$$\begin{aligned} R_T(P) &= E(T-\theta)^2 \\ &= \int_0^\infty P((T-\theta)^2 > s)ds \\ &\leq \int_0^\infty P((U-\theta)^2 > s)ds \\ &= E(U-\theta)^2 \\ &= R_U(P). \quad \blacksquare \end{aligned}$$

Exercise 33 (#2.67). Let $(X_1, ..., X_n)$ be a random sample from the exponential distribution on $(0, \infty)$ with scale parameter $\theta \in (0, \infty)$. Consider the hypotheses $H_0 : \theta \leq \theta_0$ versus $H_1 : \theta > \theta_0$, where $\theta_0 > 0$ is a fixed constant. Obtain the risk function (in terms of θ) of the test rule $T_c = I_{(c,\infty)}(\bar{X})$ under the 0-1 loss, where $\bar{X}$ is the sample mean and $c > 0$ is a constant.

Solution. Let $L(\theta, a)$ be the loss function. Then $L(\theta, 1) = 0$ when $\theta > \theta_0$, $L(\theta, 1) = 1$ when $\theta \leq \theta_0$, $L(\theta, 0) = 0$ when $\theta \leq \theta_0$, and $L(\theta, 0) = 1$ when $\theta > \theta_0$. Hence,

$$\begin{aligned} R_{T_c}(\theta) &= E[L(\theta, I_{(c,\infty)}(\bar{X}))] \\ &= E\left[L(\theta,1)I_{(c,\infty)}(\bar{X}) + L(\theta,0)I_{(0,c]}(\bar{X})\right] \\ &= L(\theta,1)P(\bar{X} > c) + L(\theta,0)P(\bar{X} \leq c) \\ &= \begin{cases} P(\bar{X} > c) & \theta \leq \theta_0 \\ P(\bar{X} \leq c) & \theta > \theta_0. \end{cases} \end{aligned}$$

Since $n\bar{X}$ has the gamma distribution with shape parameter n and scale parameter θ,

$$P(\bar{X} > c) = \frac{1}{\theta^n (n-1)!} \int_{nc}^\infty x^{n-1} e^{-x/\theta} dx. \quad \blacksquare$$

Exercise 34 (#2.71). Consider an estimation problem with a parametric family $\mathcal{P} = \{P_\theta : \theta \in \Theta\}$ and the squared error loss. If $\theta_0 \in \Theta$ satisfies that $P_\theta \ll P_{\theta_0}$ for any $\theta \in \Theta$, show that the estimator $T \equiv \theta_0$ is admissible.

Solution. Note that the risk $R_T(\theta) = 0$ when $\theta = \theta_0$. Suppose that U is an estimator of θ and $R_U(\theta) = E(U-\theta)^2 \leq R_T(\theta)$ for all θ. Then $R_U(\theta_0) = 0$, i.e., $E(U-\theta_0)^2 = 0$ under P_{θ_0}. Therefore, $U = \theta_0$ a.s. P_{θ_0}. Since $P_\theta \ll P_{\theta_0}$ for any θ, we conclude that $U = \theta_0$ a.s. $\mathcal{P}$. Hence $U = T$ a.s. $\mathcal{P}$. Thus, T is admissible. ∎

Exercise 35 (#2.73). Let $(X_1, ..., X_n)$ be a random sample of random variables with $EX_1^2 < \infty$. Consider estimating $\mu = EX_1$ under the squared error loss. Show that
(i) any estimator of the form $a\bar{X} + b$ is inadmissible, where $\bar{X}$ is the sample mean, a and b are constants, and $a > 1$;
(ii) any estimator of the form $\bar{X} + b$ is inadmissible, where $b \neq 0$ is a constant.
Solution. (i) Note that

$$\begin{aligned} R_{a\bar{X}+b}(P) &= E(a\bar{X} + b - \mu)^2 \\ &= a^2\text{Var}(\bar{X}) + (a\mu + b - \mu)^2 \\ &\geq a^2\text{Var}(\bar{X}) \\ &= a^2 R_{\bar{X}}(P) \\ &> R_{\bar{X}}(P) \end{aligned}$$

when $a > 1$. Hence $\bar{X}$ is better than $a\bar{X} + b$ with $a > 1$.
(ii) For $b \neq 0$,

$$R_{\bar{X}+b}(P) = E(\bar{X} + b - \mu)^2 = \text{Var}(\bar{X}) + b^2 > \text{Var}(\bar{X}) = R_{\bar{X}}(P).$$

Hence $\bar{X}$ is better than $\bar{X} + b$ with $b \neq 0$. ∎

Exercise 36 (#2.74). Consider an estimation problem with $\vartheta \in [c, d] \subset \mathcal{R}$, where c and d are known. Suppose that the action space contains $[c, d]$ and the loss function is $L(|\vartheta - a|)$, where $L(\cdot)$ is an increasing function on $[0, \infty)$. Show that any decision rule T with $P(T(X) \notin [c, d]) > 0$ for some $P \in \mathcal{P}$ is inadmissible.
Solution. Consider the decision rule

$$T_1 = cI_{(-\infty,c)}(T) + TI_{[c,d]}(T) + dI_{(d,\infty)}(T).$$

Then $|T_1 - \vartheta| \leq |T - \vartheta|$ and, since L is an increasing function,

$$R_{T_1}(P) = E[L(|T_1 - \vartheta|)] \leq E[L(|T - \vartheta|)] = R_T(P)$$

for any $P \in \mathcal{P}$. Since

$$P(|T_1(X) - \vartheta| < |T(X) - \vartheta|) = P(T(X) \notin [a, b]) > 0$$

holds for some $P_* \in \mathcal{P}$,

$$R_{T_1}(P_*) < R_T(P_*).$$

Hence T_1 is better than T and T is inadmissible. ∎

Exercise 37 (#2.75). Let X be a sample from $P \in \mathcal{P}$, $\delta_0(X)$ be a nonrandomized rule in a decision problem with $\mathcal{R}^k$ as the action space, and T be a sufficient statistic for $P \in \mathcal{P}$. Show that if $E[I_A(\delta_0(X))|T]$ is a nonrandomized rule, i.e., $E[I_A(\delta_0(X))|T] = I_A(h(T))$ for any Borel $A \subset \mathcal{R}^k$, where h is a Borel function, then $\delta_0(X) = h(T(X))$ a.s. P.
Solution. From the assumption,

$$E\left[\sum_{i=1}^{n} c_i I_{A_i}(\delta_0(X))\middle| T\right] = \sum_{i=1}^{n} c_i I_{A_i}(h(T))$$

for any positive integer n, constants $c_1, ..., c_n$, and Borel sets $A_1, ..., A_n$. Using the results in Exercise 39 of Chapter 1, we conclude that for any bounded continuous function f, $E[f(\delta_0(X))|T] = f(h(T))$ a.s. P. Then, by the result in Exercise 45 of Chapter 1, $\delta_0(X) = h(T)$ a.s. P. ∎

Exercise 38 (#2.76). Let X be a sample from $P \in \mathcal{P}$, $\delta_0(X)$ be a decision rule (which may be randomized) in a problem with $\mathcal{R}^k$ as the action space, and T be a sufficient statistic for $P \in \mathcal{P}$. For any Borel $A \subset \mathcal{R}^k$, define

$$\delta_1(T, A) = E[\delta_0(X, A)|T].$$

Let $L(P, a)$ be a loss function. Show that

$$\int_{\mathcal{R}^k} L(P, a) d\delta_1(X, a) = E\left[\int_{\mathcal{R}^k} L(P, a) d\delta_0(X, a)\middle| T\right] \quad \text{a.s. } P.$$

Solution. If L is a simple function (a linear combination of indicator functions), then the result follows from the definition of δ_1. For nonnegative L, it is the limit of a sequence of nonnegative increasing simple functions. Then the result follows from the result for simple L and the monotone convergence theorem for conditional expectations (Exercise 38 in Chapter 1). ∎

Exercise 39 (#2.80). Let $X_1, ..., X_n$ be random variables with a finite common mean $\mu = EX_i$ and finite variances. Consider the estimation of μ under the squared error loss.
(i) Show that there is no optimal rule in $\Im$ if $\Im$ contains all possible estimators.
(ii) Find an optimal rule in

$$\Im_2 = \left\{\sum_{i=1}^{n} c_i X_i : c_i \in \mathcal{R}, \sum_{i=1}^{n} c_i = 1\right\}$$

if $\mathrm{Var}(X_i) = \sigma^2/a_i$ with an unknown σ^2 and known a_i, $i = 1, ..., n$.
(iii) Find an optimal rule in $\Im_2$ if $X_1, ..., X_n$ are identically distributed but

are correlated with correlation coefficient ρ.

Solution. (i) Suppose that there exists an optimal rule T^*. Let P_1 and P_2 be two possible distributions of $X = (X_1, ..., X_n)$ such that $\mu = \mu_j$ under P_j and $\mu_1 \neq \mu_2$. Let $R_T(P)$ be the risk of T. For $T_1(X) \equiv \mu_1$, $R_{T_1}(P_1) = 0$. Since T^* is better than T_1, $R_{T^*}(P_1) \leq R_{T_1}(P_1) = 0$ and, hence, $T^* \equiv \mu_1$ a.s. P_1. Let $\bar{P} = (P_1 + P_2)/2$. If X has distribution $\bar{P}$, then $\mu = (\mu_1 + \mu_2)/2$. Let $T_0(X) \equiv (\mu_1 + \mu_2)/2$. Then $R_{T_0}(\bar{P}) = 0$. Since T^* is better than T_0, $R_{T^*}(\bar{P}) = 0$ and, hence, $T^* \equiv (\mu_1 + \mu_2)/2$ a.s. $\bar{P}$, which implies that $T^* \equiv (\mu_1 + \mu_2)/2$ a.s. P_1 since $P_1 \ll \bar{P}$. This is impossible since $\mu_1 \neq (\mu_1 + \mu_2)/2$.

(ii) Let $T = \sum_{i=1}^n c_i X_i$ and $T^* = \sum_{i=1}^n a_i X_i / \sum_{i=1}^n a_i$. Then

$$\begin{aligned} R_{T^*}(P) &= \mathrm{Var}(T^*) \\ &= \mathrm{Var}\left(\sum_{i=1}^n a_i X_i\right) \Big/ \left(\sum_{i=1}^n a_i\right)^2 \\ &= \sum_{i=1}^n a_i^2 \mathrm{Var}(X_i) \Big/ \left(\sum_{i=1}^n a_i\right)^2 \\ &= \sum_{i=1}^n a_i \sigma^2 \Big/ \left(\sum_{i=1}^n a_i\right)^2 \\ &= \sigma^2 \Big/ \left(\sum_{i=1}^n a_i\right). \end{aligned}$$

By the Cauchy-Schwarz inequality,

$$\left(\sum_{i=1}^n a_i\right)\left(\sum_{i=1}^n \frac{c_i^2}{a_i}\right) \geq \left(\sum_{i=1}^n c_i\right)^2 = 1.$$

Hence,

$$R_{T^*}(P) \leq \sigma^2 \sum_{i=1}^n \frac{c_i^2}{a_i} = \mathrm{Var}\left(\sum_{i=1}^n c_i X_i\right) = \mathrm{Var}(T) = R_T(P).$$

Therefore T^* is optimal.

(iii) For any $T = \sum_{i=1}^n c_i X_i$,

$$\begin{aligned} R_T(P) &= \mathrm{Var}(T) \\ &= \sum_{i=1}^n c_i^2 \sigma^2 + \sum_{i \neq j} c_i c_j \rho \sigma^2 \\ &= \sum_{i=1}^n c_i^2 \sigma^2 + \rho\sigma^2 \left[\left(\sum_{i=1}^n c_i\right)^2 - \sum_{i=1}^n c_i^2\right] \end{aligned}$$

$$
\begin{aligned}
&= \sigma^2 \left[(1-\rho) \sum_{i=1}^n c_i^2 + \rho \right] \\
&\geq \sigma^2 \left[(1-\rho) \left(\sum_{i=1}^n c_i \right)^2 \Big/ n + \rho \right] \\
&= \sigma^2 \left[(1-\rho)/n + \rho \right] \\
&= \mathrm{Var}(\bar{X}),
\end{aligned}
$$

where the last equality follows from the Cauchy–Schwarz inequality

$$\left(\sum_{i=1}^n c_i \right)^2 \leq n \sum_{i=1}^n c_i^2.$$

Hence, the sample mean $\bar{X}$ is optimal. ∎

Exercise 40 (#2.83). Let X be a discrete random variable with

$$P(X=-1) = p, \quad P(X=k) = (1-p)^2 p^k, \quad k = 0, 1, 2, ...,$$

where $p \in (0,1)$ is unknown. Show that
(i) $U(X)$ is an unbiased estimator of 0 if and only if $U(k) = ak$ for all $k = -1, 0, 1, 2, ...$ and some a;
(ii) $T_0(X) = I_{\{0\}}(X)$ is unbiased for $(1-p)^2$ and, under the squared error loss, T_0 is an optimal rule in $\Im$, where $\Im$ is the class of all unbiased estimators of $(1-p)^2$;
(iii) $T_0(X) = I_{\{-1\}}(X)$ is unbiased for p and, under the squared error loss, there is no optimal rule in $\Im$, where $\Im$ is the class of all unbiased estimators of p.

Solution. (i) If $U(X)$ is unbiased for 0, then

$$
\begin{aligned}
E[U(X)] &= U(-1)p + \sum_{k=0}^\infty U(k)(1-p)^2 p^k \\
&= \sum_{k=0}^\infty U(k)p^k - 2\sum_{k=0}^\infty U(k)p^{k+1} + U(-1)p + \sum_{k=0}^\infty U(k)p^{k+2} \\
&= U(0) + \sum_{k=-1}^\infty U(k+2)p^{k+2} - 2\sum_{k=-1}^\infty U(k+1)p^{k+2} \\
&\quad + \sum_{k=-1}^\infty U(k)p^{k+2} \\
&= \sum_{k=-1}^\infty [U(k) - 2U(k+1) + U(k+2)]p^{k+2} \\
&= 0
\end{aligned}
$$

for all p, which implies $U(0) = 0$ and $U(k) - 2U(k+1) + U(k+2) = 0$ for $k = -1, 0, 1, 2, ...$, or equivalently, $U(k) = ak$, where $a = U(1)$.
(ii) Since

$$E[T_0(X)] = P(X = 0) = (1-p)^2,$$

T_0 is unbiased. Let T be another unbiased estimator of $(1-p)^2$. Then $T(X) - T_0(X)$ is unbiased for 0 and, by the result in (i), $T(X) = T_0(X) + aX$ for some a. Then,

$$\begin{aligned} R_T(p) &= E[T_0(X) + aX - (1-p)^2]^2 \\ &= E(T_0 + aX)^2 + (1-p)^4 - 2(1-p)^2 E[T_0(X) + aX] \\ &= E(T_0 + aX)^2 - (1-p)^4 \\ &= a^2 P(X = -1) + P(X = 0) + a^2 \sum_{k=1}^{\infty} k^2 P(X = k) - (1-p)^4 \\ &\geq P(X = 0) - (1-p)^4 \\ &= \text{Var}(T_0). \end{aligned}$$

Hence T_0 is a optimal rule in $\Im$.
(iii) Since

$$E[T_0(X)] = P(X = -1) = p,$$

T_0 is unbiased. Let T be another unbiased estimator of p. Then $T(X) = T_0(X) + aX$ for some a and

$$\begin{aligned} R_T(p) &= E(T_0 + aX)^2 - p^2 \\ &= (1-a)^2 p + a^2 \sum_{k=0}^{\infty} k^2 (1-p) p^2 - p^2, \end{aligned}$$

which is a quadratic function in a with minimum

$$a = \left[1 + (1-p) \sum_{k=1}^{\infty} k^2 p^{k-1} \right]^{-1}$$

depending on p. Therefore, there is no optimal rule in $\Im$. ∎

Exercise 41. Let X be a random sample from a population and θ be an unknown parameter. Suppose that there are $k+1$ estimators of θ, $T_1, ..., T_{k+1}$, such that $ET_i = \theta + \sum_{j=1}^{k} c_{i,j} b_j(\theta)$, $i = 1, ..., k+1$, where $c_{i,j}$'s are constants and $b_j(\theta)$ are functions of θ. Suppose that the determinant

$$C = \begin{vmatrix} 1 & 1 & \cdots & 1 \\ c_{1,1} & c_{2,1} & \cdots & c_{k+1,1} \\ \cdots & \cdots & \cdots & \cdots \\ c_{1,k} & c_{2,k} & \cdots & c_{k+1,k} \end{vmatrix} \neq 0.$$

Show that

$$T^* = \frac{1}{C}\begin{vmatrix} T_1 & T_2 & \cdots & T_{k+1} \\ c_{1,1} & c_{2,1} & \cdots & c_{k+1,1} \\ \cdots & \cdots & \cdots & \cdots \\ c_{1,k} & c_{2,k} & \cdots & c_{k+1,k} \end{vmatrix}$$

is an unbiased estimator of θ.

Solution. From the properties of a determinant,

$$\begin{aligned} ET^* &= \frac{1}{C}\begin{vmatrix} ET_1 & ET_2 & \cdots & ET_{k+1} \\ c_{1,1} & c_{2,1} & \cdots & c_{k+1,1} \\ \cdots & \cdots & \cdots & \cdots \\ c_{1,k} & c_{2,k} & \cdots & c_{k+1,k} \end{vmatrix} \\ &= \frac{1}{C}\begin{vmatrix} \theta + \sum_{j=1}^k c_{1,j}b_j(\theta) & \cdots & \theta + \sum_{j=1}^k c_{k+1,j}b_j(\theta) \\ c_{1,1} & \cdots & c_{k+1,1} \\ \cdots & \cdots & \cdots \\ c_{1,k} & \cdots & c_{k+1,k} \end{vmatrix} \\ &= \frac{\theta}{C}\begin{vmatrix} 1 & \cdots & 1 \\ c_{1,1} & \cdots & c_{k+1,1} \\ \cdots & \cdots & \cdots \\ c_{1,k} & \cdots & c_{k+1,k} \end{vmatrix} \\ &\quad + \frac{1}{C}\begin{vmatrix} \sum_{j=1}^k c_{1,j}b_j(\theta) & \cdots & \sum_{j=1}^k c_{k+1,j}b_j(\theta) \\ c_{1,1} & \cdots & c_{k+1,1} \\ \cdots & \cdots & \cdots \\ c_{1,k} & \cdots & c_{k+1,k} \end{vmatrix} \\ &= \theta, \end{aligned}$$

where the last equality follows from the fact that the last determinant is 0 because its first row is a linear combination of its other k rows. ∎

Exercise 42 (#2.84). Let X be a random variable having the binomial distribution with size n and probability $p \in (0,1)$. Show that there is no unbiased estimator of p^{-1}.

Solution. Suppose that $T(X)$ is an unbiased estimator of p^{-1}. Then

$$E[T(X)] = \sum_{k=0}^{n} \binom{n}{k} T(k) p^k (1-p)^{n-k} = \frac{1}{p}$$

for all p. However,

$$\sum_{k=0}^{n} \binom{n}{k} T(k) p^k (1-p)^{n-k} \le \sum_{k=0}^{n} \binom{n}{k} T(k) < \infty$$

for any p but p^{-1} diverges to ∞ as $p \to 0$. This is impossible. Hence, there is no unbiased estimator of p^{-1}. ∎

Exercise 43 (#2.85). Let $X = (X_1, ..., X_n)$ be a random sample from $N(\theta, 1)$, where $\theta = 0$ or 1. Consider the estimation of θ with action space $\{0, 1\}$, i.e., the range of any estimator is $\{0, 1\}$.
(i) Show that there does not exist any unbiased estimator of θ.
(ii) Find an estimator $\hat{\theta}$ of θ that is approximately unbiased, that is, $\lim_n E(\hat{\theta}) = \theta$.
Solution. (i) Since the action space is $\{0, 1\}$, any randomized estimator $\hat{\theta}$ can be written as $T(X)$, where T is Borel, $0 \leq T(X) \leq 1$, and

$$\hat{\theta} = \begin{cases} 1 & \text{with probability } T(X) \\ 0 & \text{with probability } 1 - T(X). \end{cases}$$

Then $E(\hat{\theta}) = E[T(X)]$. If $\hat{\theta}$ is unbiased, then $E[T(X)] = \theta$ for $\theta = 0, 1$. This implies that, when $\theta = 0$, $T(X) = 0$ a.e. Lebesgue measure, whereas when $\theta = 1$, $T(X) = 1$ a.e. Lebesgue measure. This is impossible. Hence there does not exist any unbiased estimator of θ.
(ii) Consider $\hat{\theta} = I_{(n^{-1/4}, \infty)}(|\bar{X}|)$, where $\bar{X}$ is the sample mean. Since $\bar{X}$ is distributed as $N(\theta, n^{-1})$,

$$E(\hat{\theta}) = P(|\bar{X}| > n^{-1/4}) = 1 - \Phi\left(n^{1/4} - \theta\sqrt{n}\right) + \Phi\left(-n^{1/4} - \theta\sqrt{n}\right),$$

where Φ is the cumulative distribution function of $N(0, 1)$. Hence, when $\theta = 0$, $\lim_n E(\hat{\theta}) = 1 - \Phi(\infty) + \Phi(-\infty) = 0$ and, when $\theta = 1$, $\lim_n E(\hat{\theta}) = 1 - \Phi(-\infty) + \Phi(-\infty) = 1$. ∎

Exercise 44 (#2.92(c)). Let X be a sample from P_θ, where $\theta \in \Theta \subset \mathcal{R}$. Consider the estimation of θ under the absolute error loss function $|a - \theta|$. Let Π be a given distribution on Θ with finite mean. Find a Bayes rule.
Solution. Let $P_{\theta|X}$ be the posterior distribution of θ and P_X be the marginal of X. By Fubini's theorem,

$$\int \int |\theta| dP_{\theta|X} dP_X = \int \int |\theta| dP_\theta d\Pi = \int |\theta| d\Pi < \infty.$$

Hence, for almost all X, $\int |\theta| dP_{\theta|X} < \infty$. From Exercise 11 in Chapter 1, if m_X is a median of $P_{\theta|X}$, then

$$\int |\theta - m_X| dP_{\theta|X} \leq \int |\theta - a| dP_{\theta|X} \quad \text{for almost all } X$$

holds for any a. Hence, $E|\theta - m_X| \leq E|\theta - T(X)|$ for any other estimator $T(X)$. This shows that m_X is a Bayes rule. ∎

Exercise 45 (#2.93). Let X be a sample having a probability density $f_j(x)$ with respect to a σ-finite measure ν, where j is unknown and $j \in \{1, ..., J\}$ with a known integer $J \geq 2$. Consider a decision problem in which the action space is $\{1, ..., J\}$ and the loss function is

$$L(j, a) = \begin{cases} 0 & \text{if } a = j \\ 1 & \text{if } a \neq j. \end{cases}$$

(i) Obtain the risk of a decision rule (which may be randomized).
(ii) Let Π be a prior probability measure on $\{1, ..., J\}$ with $\Pi(\{j\}) = \pi_j$, $j = 1, ..., J$. Obtain the Bayes risk of a decision rule.
(iii) Obtain a Bayes rule under the prior Π in (ii).
(iv) Assume that $J = 2$, $\pi_1 = \pi_2 = 0.5$, and $f_j(x) = \phi(x - \mu_j)$, where $\phi(x)$ is the Lebesgue density of the standard normal distribution and μ_j, $j = 1, 2$, are known constants. Obtain the Bayes rule in (iii).
(v) Obtain a minimax rule when $J = 2$.
Solution. (i) Let δ be a randomized decision rule. For any X, let $\delta(X, j)$ be the probability of taking action j under the rule δ. Let E_j be the expectation taking under f_j. Then

$$R_\delta(j) = E_j \left[\sum_{k=1}^{J} L(j, k)\delta(X, k) \right] = \sum_{k \neq j} E_j[\delta(X, k)] = 1 - E_j[\delta(X, j)],$$

since $\sum_{k=1}^{J} \delta(X, k) = 1$.
(ii) The Bayes risk of a decision rule δ is

$$r_\delta = \sum_{j=1}^{J} \pi_j R_\delta(j) = 1 - \sum_{j=1}^{J} \pi_j E_j[\delta(X, j)].$$

(iii) Let δ^* be a rule satisfying $\delta^*(X, j) = 1$ if and only if $\pi_j f_j(X) = g(X)$, where $g(X) = \max_{1 \leq k \leq J} \pi_k f_k(X)$. Then δ_* is a Bayes rule, since, for any rule δ,

$$\begin{aligned} r_\delta &= 1 - \sum_{j=1}^{J} \int \pi_j \delta(x, j) f_j(x) d\nu \\ &\geq 1 - \sum_{j=1}^{J} \int \delta(x, j) g(x) d\nu \\ &= 1 - \int g(x) d\nu \\ &= 1 - \sum_{j=1}^{J} \int_{g(x) = \pi_j f_j(x)} \pi_j f_j(x) d\nu \\ &= r_{\delta^*}. \end{aligned}$$

(iv) From the result in (iii), the Bayes rule $\delta^*(X,j) = 1$ if and only if $\phi(x-\mu_j) > \phi(x-\mu_k)$, $k \neq j$. Since $\phi(x-\mu_j) = e^{-(x-\mu_j)^2/2}/\sqrt{2\pi}$, we can obtain a nonrandomized Bayes rule that takes action 1 if and only if $|X-\mu_1| < |X-\mu_2|$.
(v) Let c be a positive constant and consider a rule δ_c such that $\delta_c(X,1) = 1$ if $f_1(X) > cf_2(X)$, $\delta_c(X,2) = 1$ if $f_1(X) < cf_2(X)$, and $\delta_c(X,1) = \gamma$ if $f_1(X) = cf_2(X)$. Since $\delta_c(X,j) = 1$ if and only if $\pi_j f_j(X) = \max_k \pi_k f_k(X)$, where $\pi_1 = 1/(c+1)$ and $\pi_2 = c/(c+1)$, it follows from part (iii) of the solution that δ_c is a Bayes rule. Let P_j be the probability corresponding to f_j. The risk of δ_c is $P_1(f_1(X) \leq cf_2(X)) - \gamma P_1(f_1(x) = cf_2(X))$ when $j = 1$ and $1 - P_2(f_1(X) \leq cf_2(X)) + \gamma P_2(f_1(x) = cf_2(X))$ when $j = 2$. Let $\psi(c) = P_1(f_1(X) \leq cf_2(X)) + P_2(f_1(X) \leq cf_2(X)) - 1$. Then ψ is nondecreasing in c, $\psi(0) = -1$, $\lim_{c\to\infty}\psi(c) = 1$, and $\psi(c) - \psi(c-) = P_1(f_1(X) = cf_2(X)) + P_2(f_1(X) = cf_2(X))$. Let $c_* = \inf\{c : \psi(c) \geq 0\}$. If $\psi(c_*) = \psi(c_*-)$, we set $\gamma = 0$; otherwise, we set $\gamma = \psi(c_*)/[\psi(c_*)-\psi(c_*-)]$. Then, the risk of δ_{c_*} is a constant. For any rule δ, $\sup_j R_\delta(j) \geq r_\delta \geq r_{\delta_{c_*}} = R_{\delta_{c_*}}(j) = \sup_j R_{\delta_{c_*}}(j)$. Hence, δ_{c_*} is a minimax rule. ∎

Exercise 46 (#2.94). Let $\hat\theta$ be an unbiased estimator of an unknown $\theta \in \mathcal{R}$.
(i) Under the squared error loss, show that the estimator $\hat\theta + c$ is not minimax unless $\sup_\theta R_T(\theta) = \infty$ for any estimator T, where $c \neq 0$ is a known constant.
(ii) Under the squared error loss, show that the estimator $c\hat\theta$ is not minimax unless $\sup_\theta R_T(\theta) = \infty$ for any estimator T, where $c \in (0,1)$ is a known constant.
(iii) Consider the loss function $L(\theta,a) = (a-\theta)^2/\theta^2$ (assuming $\theta \neq 0$). Show that $\hat\theta$ is not minimax unless $\sup_\theta R_T(\theta) = \infty$ for any T.
Solution. (i) Under the squared error loss, the risk of $\hat\theta + c$ is

$$R_{\hat\theta+c}(P) = E(\hat\theta + c - \theta)^2 = c^2 + \mathrm{Var}(\hat\theta) = c^2 + R_{\hat\theta}(P).$$

Then

$$\sup_P R_{\hat\theta+c}(P) = c^2 + \sup_P R_{\hat\theta}(P)$$

and either $\sup_P R_{\hat\theta+c}(P) = \infty$ or $\sup_P R_{\hat\theta+c}(P) > \sup_P R_{\hat\theta}(P)$. Hence, the only case where $\hat\theta + c$ is minimax is when $\sup_P R_T(P) = \infty$ for any estimator T.
(ii) Under the squared error loss, the risk of $c\hat\theta$ is

$$R_{c\hat\theta}(P) = E(c\hat\theta - \theta)^2 = (1-c)^2\theta^2 + c^2\mathrm{Var}(\hat\theta) = (1-c)^2\theta^2 + c^2 R_{\hat\theta}(P).$$

Then, $\sup_P R_{\hat\theta+c}(P) = \infty$ and the only case where $\hat\theta + c$ is minimax is when $\sup_P R_T(P) = \infty$ for any estimator T.

(iii) Under the given loss function, the risk of $c\hat{\theta}$ is

$$R_{c\hat{\theta}}(P) = (1-c)^2 + c^2 R_{\hat{\theta}}(P).$$

If $\sup_P R_{\hat{\theta}}(P) = \infty$, then the result follows. Assume $\xi = \sup_P R_{\hat{\theta}}(P) < \infty$. Let $c = \xi/(\xi+1)$. Then

$$\sup_P R_{c\hat{\theta}}(P) = (1-c)^2 + c^2\xi = \frac{\xi^2}{(\xi+1)^2} + \frac{\xi}{(\xi+1)^2} = \frac{\xi}{\xi+1} < \xi.$$

Hence $\hat{\theta}$ is not minimax. ∎

Exercise 47 (#2.96). Let X be an observation from the binomial distribution with size n and probability $\theta \in (0,1)$, where n is a known integer ≥ 2. Consider testing hypotheses $H_0 : \theta \leq \theta_0$ versus $H_1 : \theta > \theta_0$, where $\theta_0 \in (0,1)$ is a fixed value. Let $\Im = \{T_j : j = 0, 1, ..., n-1\}$ be a class of nonrandomized decision rules, where $T_j(X) = 1$ (rejecting H_0) if and only if $X \geq j+1$. Consider the 0-1 loss function.
(i) When the uniform distribution on $(0,1)$ is used as the prior, show that the Bayes rule within the class $\Im$ is $T_{j^*}(X)$, where j^* is the largest integer in $\{0, 1, ..., n-1\}$ such that $B_{j+1,n-j+1}(\theta_0) \geq \frac{1}{2}$ and $B_{a,b}(\cdot)$ denotes the cumulative distribution function of the beta distribution with parameter (a, b).
(ii) Derive a minimax rule over the class $\Im$.
Solution. (i) Let P_θ be the probability law of X. Under the 0-1 loss, the risk of T_j is

$$\begin{aligned} R_{T_j}(\theta) &= P_\theta(X > j) I_{(0,\theta_0]}(\theta) + P_\theta(X \leq j) I_{(\theta_0,1)}(\theta) \\ &= \sum_{k=j+1}^{n} \binom{n}{k} \theta^k (1-\theta)^{n-k} I_{(0,\theta_0]}(\theta) + \sum_{k=0}^{j} \binom{n}{k} \theta^k (1-\theta)^{n-k} I_{(\theta_0,1)}(\theta). \end{aligned}$$

Hence, the Bayes risk of T_j is

$$\begin{aligned} r_{T_j} &= \sum_{k=j+1}^{n} \binom{n}{k} \int_0^{\theta_0} \theta^k (1-\theta)^{n-k} d\theta + \sum_{k=0}^{j} \binom{n}{k} \int_{\theta_0}^{1} \theta^k (1-\theta)^{n-k} d\theta \\ &= \sum_{k=j+1}^{n} B_{k+1,n-k+1}(\theta_0) + \sum_{k=0}^{j} [1 - B_{k+1,n-k+1}(\theta_0)]. \end{aligned}$$

Then, for $j = 1, ..., n-1$,

$$r_{T_{j-1}} - r_{T_j} = 2B_{j+1,n-j+1}(\theta_0) - 1.$$

The family $\{B_{\beta+1,n-\beta+1}(y) : \beta > 0\}$ is an exponential family having monotone likelihood ratio in $\log y - \log(1-y)$. By Lemma 6.3 in Shao (2003), if

Y has distribution $B_{\beta+1,n-\beta+1}$, then $P(Y \le t) = P(\log Y - \log(1-Y) \le \log t - \log(1-t))$ is decreasing in β for any fixed $t \in (0,1)$. This shows that $B_{j+1,n-j+1}(\theta_0)$ is decreasing in j. Hence, if j^* is the largest integer j such that $B_{j+1,n-j+1}(\theta_0) \ge \frac{1}{2}$, then

$$r_{T_{j-1}} - r_{T_j} \ge 0 \quad j = 1, ..., j^*$$

and

$$r_{T_{j-1}} - r_{T_j} \le 0 \quad j = j^* + 1, ..., n-1.$$

Consequently,

$$r_{T_{j^*}} = \min_{j=0,1,...,n-1} r_{T_j}.$$

This shows that T_{j^*} is the Bayes rule over the class $\Im$.
(ii) Again, by Lemma 6.3 in Shao (2003), $P_\theta(X \le j)$ is decreasing in θ and $P_\theta(X > j)$ is increasing in θ. Hence,

$$\sup_{\theta\in(0,1)} R_{T_j}(\theta) = P_{\theta_0}(X > j) = \sum_{k=j+1}^{n} \binom{n}{k} \theta_0^k (1-\theta_0)^{n-k}.$$

Then, the minimax rule over the class $\Im$ is T_{n-1}. ∎

Exercise 48 (#2.99). Let $(X_1, ..., X_n)$ be a random sample from the Cauchy distribution with location parameter $\mu \in \mathcal{R}$ and a known scale parameter $\sigma > 0$. Consider the hypotheses $H_0 : \mu \le \mu_0$ versus $H_1 : \mu > \mu_0$, where μ_0 is a fixed constant. Calculate the size of the nonrandomized test $T_c(X) = I_{(c,\infty)}(\bar{X})$, where c is a fixed constant; find a c_α such that T_{c_α} has size $\alpha \in (0,1)$; and find the p-value for T_{c_α}.
Solution: Note that $\bar{X}$ has the same distribution as X_1. Hence, the size of $T_c(X)$ is

$$\begin{aligned}
\sup_{\mu\le\mu_0} E(T_c(X)) &= \sup_{\mu\le\mu_0} P(\bar{X} > c) \\
&= \sup_{\mu\le\mu_0} P\left(\frac{\bar{X}-\mu}{\sigma} > \frac{c-\mu}{\sigma}\right) \\
&= \sup_{\mu\le\mu_0} \left[\frac{1}{2} - \frac{1}{\pi}\arctan\left(\frac{c-\mu}{\sigma}\right)\right] \\
&= \frac{1}{2} - \frac{1}{\pi}\arctan\left(\frac{c-\mu_0}{\sigma}\right).
\end{aligned}$$

Therefore, if $c_\alpha = \mu_0 + \sigma\tan\left(\pi(\frac{1}{2} - \alpha)\right)$, then the size of $T_{c_\alpha}(X)$ is exactly α. Note that

$$\alpha = \frac{1}{2} - \frac{1}{\pi}\arctan\left(\frac{c_\alpha - \mu_0}{\sigma}\right) > \frac{1}{2} - \frac{1}{\pi}\arctan\left(\frac{\bar{X} - \mu_0}{\sigma}\right)$$

if and only if $\bar{X} > c_\alpha$ (i.e., $T_{c_\alpha}(X) = 1$). Hence, the p-value of $T_{c_\alpha}(X)$ is

$$\inf\{\alpha | T_{c_\alpha}(X) = 1\} = \frac{1}{2} - \frac{1}{\pi} \arctan\left(\frac{\bar{X} - \mu_0}{\sigma}\right). \blacksquare$$

Exercise 49 (#2.101). Let $(X_1, ..., X_n)$ be a random sample from the exponential distribution on (a, ∞) with scale parameter θ, where $a \in \mathcal{R}$ and $\theta > 0$ are unknown parameters. Let $\alpha \in (0, 1)$ be given.
(i) Using $T_1(X) = \sum_{i=1}^n (X_i - X_{(1)})$, where $X_{(1)}$ is the smallest order statistic, construct a confidence interval for θ with confidence coefficient $1 - \alpha$ and find the expected interval length.
(ii) Using $T_1(X)$ and $T_2(X) = X_{(1)}$, construct a confidence interval for a with confidence coefficient $1 - \alpha$ and find the expected interval length.
(iii) Construct a confidence set for the two-dimensional parameter (a, θ) with confidence coefficient $1 - \alpha$.
Solution. (i) Let $W = T_1(X)/\theta$. Then W has the gamma distribution with shape parameter $n - 1$ and scale parameter 1. Let $c_1 < c_2$ such that $P(c_1 < W < c_2) = 1 - \alpha$. Then c_1 and c_2 can be chosen so that they do not depend on unknown parameters. A confidence interval for θ with confidence coefficient $1 - \alpha$ is

$$\left(\frac{T_1(X)}{c_2}, \frac{T_1(X)}{c_1}\right).$$

Its expected length is

$$\left(\frac{1}{c_1} - \frac{1}{c_2}\right) E(T_1) = \left(\frac{1}{c_1} - \frac{1}{c_2}\right)(n - 1)\theta.$$

(ii) Using the result in Exercise 7(iii), $[T_2(X) - a]/T_1(X)$ has the Lebesgue density $n\left(1 + \frac{nt}{n-1}\right)^{-n} I_{(0,\infty)}(t)$, which does not depend on any unknown parameter. Choose two constants $0 < c_1 < c_2$ such that

$$\int_{c_1}^{c_2} n\left(1 + \frac{nt}{n - 1}\right)^{-n} dt = 1 - \alpha.$$

Then a confidence interval for a with confidence coefficient $1 - \alpha$ is

$$(T_2 - c_2 T_1, T_2 - c_1 T_1).$$

Its expected length is

$$E[(c_2 - c_1)T_1] = (c_2 - c_1)(n - 1)\theta.$$

(iii) Let $0 < a_1 < a_2$ be constants such that

$$P(a_1 < W < a_2) = \sqrt{1 - \alpha}$$

and let $0 < b_1 < b_2$ be constants such that

$$P\left(b_1 < \frac{T_2(X) - a}{\theta} < b_2\right) = e^{-nb_1} - e^{-nb_2} = \sqrt{1-\alpha}.$$

Consider the region

$$C(X) = \left\{(\theta, a)\colon \ \frac{T_1(X)}{a_2} < \theta < \frac{T_1(X)}{a_1}, T_2(X) - b_2\theta < a < T_2(X) - b_1\theta\right\}.$$

By the result in Exercise 7(iii), $T_1(X)$ and $T_2(X)$ are independent. Hence

$$\begin{aligned} P\big((a,\theta) \in C(X)\big) &= P\left(a_1 < \frac{T_1(X)}{\theta} < a_2,\ b_1 < \frac{T_2(X) - a}{\theta} < b_2\right) \\ &= P\left(a_1 < \frac{T_1(X)}{\theta} < a_2\right) P\left(b_1 < \frac{T_2(X) - a}{\theta} < b_2\right) \\ &= \sqrt{1-\alpha}\sqrt{1-\alpha} \\ &= 1 - \alpha. \end{aligned}$$

Hence, $C(X)$ is a confidence region for (a, θ) with confidence coefficient $1 - \alpha$. ∎

Exercise 50 (#2.104). Let $(X_1, ..., X_n)$ be a random sample from the uniform distribution on the interval $(\theta - \frac{1}{2}, \theta + \frac{1}{2})$, where $\theta \in \mathcal{R}$ is unknown. Let $X_{(j)}$ be the jth order statistic. Show that $(X_{(1)} + X_{(n)})/2$ is strongly consistent for θ and also consistent in mean squared error.

Solution. (i) For any $\epsilon > 0$,

$$\begin{aligned} P\left(|X_{(1)} - (\theta - \tfrac{1}{2})| > \epsilon\right) &= P\left(X_{(1)} > \epsilon + (\theta - \tfrac{1}{2})\right) \\ &= \left[P\left(X_1 > \epsilon + \theta - \tfrac{1}{2}\right)\right]^n \\ &= (1-\epsilon)^n \end{aligned}$$

and

$$\begin{aligned} P\left(|X_{(n)} - (\theta + \tfrac{1}{2})| > \epsilon\right) &= P\left(X_{(n)} < (\theta + \tfrac{1}{2}) - \epsilon\right) \\ &= \left[P\left(X_1 < \theta + \tfrac{1}{2} - \epsilon\right)\right]^n \\ &= (1-\epsilon)^n. \end{aligned}$$

Since $\sum_{i=1}^n (1-\epsilon)^n < \infty$, we conclude that $\lim_n X_{(1)} = \theta - \frac{1}{2}$ a.s. and $\lim_n X_{(n)} = \theta + \frac{1}{2}$ a.s. Hence $\lim_n (X_{(1)} + X_{(n)})/2 = \theta$ a.s.

(ii) A direct calculation shows that

$$E[X_{(n)} - (\theta + \tfrac{1}{2})] = n\int_0^1 x^n dx - 1 = -\frac{1}{n+1}$$

and

$$E[X_{(1)} - (\theta - \tfrac{1}{2})] = n\int_0^1 x(1-x)^{n-1}dx = \frac{1}{n+1}.$$

Hence $(X_{(1)} + X_{(n)})/2$ is unbiased for θ. Note that

$$\begin{aligned}
\mathrm{Var}(X_{(n)}) &= \mathrm{Var}\big(X_{(n)} - (\theta - \tfrac{1}{2})\big) \\
&= E[X_{(n)} - (\theta - \tfrac{1}{2})]^2 - [EX_{(n)} - (\theta - \tfrac{1}{2})]^2 \\
&= n\int_0^1 x^{n+1}dx - \left[\theta + \tfrac{1}{2} - \frac{1}{n+1} - (\theta - \tfrac{1}{2})\right]^2 \\
&= \frac{n}{n+2} - \left(\frac{n}{n+1}\right)^2 \\
&\to 0
\end{aligned}$$

as $n \to \infty$. Similarly, $\lim_n \mathrm{Var}(X_{(1)}) = 0$. By the Cauchy-Schwarz inequality, $[\mathrm{Cov}(X_{(1)}, X_{(n)})]^2 \leq \mathrm{Var}(X_{(1)})\mathrm{Var}(X_{(n)})$. Thus, $\lim_n \mathrm{Cov}(X_{(1)}, X_{(n)}) = 0$ and, consequently, $\lim_n E[(X_{(1)} + X_{(n)})/2 - \theta]^2 = \lim_n 4^{-1}[\mathrm{Var}(X_{(1)} + \mathrm{Var}(X_{(n)}) + 2\mathrm{Cov}(X_{(1)}, X_{(n)})] = 0$. ∎

Exercise 51 (#2.105). Let $(X_1, ..., X_n)$ be a random sample from a population with the Lebesgue density $f_\theta(x) = 2^{-1}(1 + \theta x)I_{(-1,1)}(x)$, where $\theta \in (-1, 1)$ is an unknown parameter. Find an estimator of θ that is strongly consistent and consistent in mean squared error.

Solution. By the strong law of large numbers, the sample mean $\bar{X}$ is strongly consistent for

$$EX_1 = \frac{1}{2}\int_{-1}^1 x(1+\theta x)dx = \frac{\theta}{2}\int_{-1}^1 x^2 dx = \frac{\theta}{3}.$$

Hence $3\bar{X}$ is a strongly consistent estimator of θ. Since $3\bar{X}$ is unbiased for θ and $\mathrm{Var}(3\bar{X}) = 9\mathrm{Var}(X_1)/n$, where

$$\mathrm{Var}(X_1) = EX_1^2 - (EX_1)^2 = \frac{1}{2}\int_{-1}^1 x^2(1+\theta x)dx - \frac{\theta^2}{9} = \frac{1}{2} - \frac{\theta^2}{9},$$

we conclude that $3\bar{X}$ is consistent in mean squared error. ∎

Exercise 52 (#2.106). Let $X_1, ..., X_n$ be a random sample. Suppose that T_n is an unbiased estimator of ϑ based on $X_1, ..., X_n$ such that for any n, $\mathrm{Var}(T_n) < \infty$ and $\mathrm{Var}(T_n) \leq \mathrm{Var}(U_n)$ for any other unbiased estimator U_n of ϑ based on $X_1, ..., X_n$. Show that T_n is consistent in mean squared error.

Solution. Let $U_n = n^{-1}\sum_{i=1}^n T_1(X_i)$. Then U_n is unbiased for ϑ since $T_1(X_1)$ is unbiased for ϑ. By the assumption, $\mathrm{Var}(T_n) \leq \mathrm{Var}(U_n)$. Hence $\lim_n \mathrm{Var}(T_n) = 0$ since $\lim_n \mathrm{Var}(U_n) = \lim_n \mathrm{Var}\big(T_1(X_1)\big)/n = 0$. ∎

Exercise 53 (#2.111). Let $X_1, ..., X_n$ be a random sample from P with unknown mean $\mu \in \mathcal{R}$ and variance $\sigma^2 > 0$, and let $g(\mu) = 0$ if $\mu \neq 0$ and $g(0) = 1$. Find a consistent estimator of $g(\mu)$.
Solution. Consider the estimator $T(X) = I_{(0,n^{-1/4})}(|\bar{X}|)$, where $\bar{X}$ is the sample mean. Note that $T = 0$ or 1. Hence, we only need to show that $\lim_n P(T = 1) = 1$ when $g(\mu) = 1$ (i.e., $\mu = 0$) and $\lim_n P(T = 1) = 0$ when $g(\mu) = 0$ (i.e., $\mu \neq 0$). If $\mu = 0$, by the central limit theorem, $\sqrt{n}\bar{X} \to_d N(0, \sigma^2)$ and, thus

$$\lim_n P(T(X) = 1) = \lim_n P(\sqrt{n}|\bar{X}| < n^{1/4}) = \lim_n \Phi(n^{1/4}) = 1,$$

where Φ is the cumulative distribution function of $N(0,1)$. If $\mu \neq 0$, then by the law of large numbers, $|\bar{X}| \to_p |\mu| > 0$ and, hence, $n^{-1/4}/|\bar{X}| \to_p 0$. Then

$$\lim_n P(T(X) = 1) = \lim_n P(1 < n^{-1/4}/|\bar{X}|) = 0. \quad ∎$$

Exercise 54 (#2.115). Let $(X_1, ..., X_n)$ be a random sample of random variables from a population P with $EX_1^2 < \infty$ and $\bar{X}$ be the sample mean. Consider the estimation of $\mu = EX_1$.
(i) Let $T_n = \bar{X} + \xi_n/\sqrt{n}$, where ξ_n is a random variable satisfying $\xi_n = 0$ with probability $1 - n^{-1}$ and $\xi_n = n^{3/2}$ with probability n^{-1}. Show that the bias of T_n is not the same as the asymptotic bias of T_n for any P.
(ii) Let $T_n = \bar{X} + \eta_n/\sqrt{n}$, where η_n is a random variable that is independent of $X_1, ..., X_n$ and equals 0 with probability $1 - 2n^{-1}$ and $\pm\sqrt{n}$ with probability n^{-1}. Show that the asymptotic mean squared error of T_n, the asymptotic mean squared error of $\bar{X}$, and the mean squared error of $\bar{X}$ are the same, but the mean squared error of T_n is larger than the mean squared error of $\bar{X}$ for any P.
Note. The asymptotic bias and mean squared error are defined according to Definitions 2.11 and 2.12 in Shao (2003).
Solution. (i) Since $E(\xi_n) = n^{3/2}n^{-1} = n^{1/2}$, $E(T_n) = E(\bar{X}) + n^{-1/2}E(\xi_n) = \mu + 1$. This means that the bias of T_n is 1. Since $\xi_n \to_p 0$ and $\bar{X} \to_p \mu$, $T_n \to_p \mu$. Thus, the asymptotic bias of T_n is 0.
(ii) Since $\eta_n \to_p 0$ and $\sqrt{n}(\bar{X} - \mu) \to_d N(0, \sigma^2)$, where $\sigma^2 = \text{Var}(X_1)$, by Slutsky's theorem, $\sqrt{n}(T_n - \mu) = \sqrt{n}(\bar{X} - \mu) + \eta_n \to_d N(0, \sigma^2)$. Hence, the asymptotic mean squared error of T_n is the same as that of $\bar{X}$ and is equal to σ^2/n, which is the mean squared error of $\bar{X}$. Since $E(\eta_n) = 0$, $E(T_n) = E(\bar{X}) = \mu$ and the mean squared error of T_n is

$$\text{Var}(T_n) = \text{Var}(\bar{X}) + \text{Var}(\eta_n/\sqrt{n}) = \frac{\sigma^2}{n} + \frac{2}{n} > \frac{\sigma^2}{n},$$

which is the mean squared error of $\bar{X}$. ∎

Exercise 55 (#2.116(b)). Let $(X_1, ..., X_n)$ be a random sample of random variables with finite $\theta = EX_1$ and $\text{Var}(X_1) = \theta$, where $\theta > 0$ is unknown. Consider the estimation of $\sqrt{\theta}$. Let $T_{1n} = \sqrt{\bar{X}}$ and $T_{2n} = \bar{X}/S$, where $\bar{X}$ and S^2 are the sample mean and sample variance. Obtain the asymptotic relative efficiency of T_{1n} with respect to T_{2n}.

Solution. Since $\sqrt{n}(\bar{X} - \theta) \to_d N(0, \theta)$, by the δ-method with $g(t) = \sqrt{t}$ and $g'(t) = (2\sqrt{t})^{-1}$, $\sqrt{n}(\sqrt{\bar{X}} - \sqrt{\theta}) \to_d N(0, \frac{1}{4})$. From Example 2.8 in Shao (2003),

$$\sqrt{n}(\bar{X} - \theta, S^2 - \theta) \to_d N_2(0, \Sigma),$$

where

$$\Sigma = \begin{pmatrix} \theta & \mu_3 \\ \mu_3 & \mu_4 - \theta^2 \end{pmatrix}$$

and $\mu_k = E(X_1 - \theta)^k$, $k = 3, 4$. By the δ-method with $g(x, y) = x/\sqrt{y}$, $\partial g/\partial x = 1/\sqrt{y}$ and $\partial g/\partial y = -x/(2y^{3/2})$, we obtain that

$$\sqrt{n}(T_{2n} - \sqrt{\theta}) \to_d N\left(0, \theta^{-1}[\theta^2 - \mu_3 + (\mu_4 - \theta^2)/4]\right).$$

Hence, the asymptotic relative efficiency of T_{1n} with respect to T_{2n} is $4\theta - 4\theta^{-1}\mu_3 + \theta^{-1}(\mu_4 - \theta^2)$. ∎

Exercise 56 (#2.118). Let $(X_1, ..., X_n)$ be a random sample from the $N(0, \sigma^2)$ distribution with an unknown $\sigma > 0$. Consider the estimation of σ. Find the asymptotic relative efficiency of $T_{1n} = \sqrt{\pi/2}\sum_{i=1}^n |X_i|/n$ with respect to $T_{2n} = (\sum_{i=1}^n X_i^2/n)^{1/2}$.

Solution. Since $E(\sqrt{\pi/2}|X_1|) = \sigma$ and $\text{Var}(\sqrt{\pi/2}|X_1|) = (\frac{\pi}{2} - 1)\sigma^2$, by the central limit theorem, we obtain that

$$\sqrt{n}(T_{1n} - \sigma) \to_d N\left(0, \left(\tfrac{\pi}{2} - 1\right)\sigma^2\right).$$

Since $EX_1^2 = \sigma^2$ and $\text{Var}(X_1) = 2\sigma^4$, $\sqrt{n}(n^{-1}\sum_{i=1}^n X_i^2 - \sigma^2) \to_d N(0, 2\sigma^4)$. By the δ-method with $g(t) = \sqrt{t}$ and $g'(t) = (2\sqrt{t})^{-1}$, we obtain that

$$\sqrt{n}(T_{2n} - \sigma) \to_d N\left(0, \tfrac{1}{2}\sigma^2\right).$$

Hence, the asymptotic relative efficiency of T_{1n} with respect to T_{2n} is equal to $\frac{1}{2}/(\frac{\pi}{2} - 1) = (\pi - 2)^{-1}$. ∎

Exercise 57 (#2.121). Let $X_1, ..., X_n$ be a random sample of random variables with $EX_i = \mu$, $\text{Var}(X_i) = 1$, and $EX_i^4 < \infty$. Let $T_{1n} = n^{-1}\sum_{i=1}^n X_i^2 - 1$ and $T_{2n} = \bar{X}^2 - n^{-1}$ be estimators of μ^2, where $\bar{X}$ is the sample mean.

(i) Find the asymptotic relative efficiency of T_{1n} with respect to T_{2n}.

(ii) Show that the asymptotic relative efficiency of T_{1n} with respect to T_{2n} is no larger than 1 if the distribution of $X_i - \mu$ is symmetric about 0 and

$\mu \neq 0$.
(iii) Find a distribution P for which the asymptotic relative efficiency of T_{1n} with respect to T_{2n} is larger than 1.
Solution. (i) Since $EX_1^2 = \text{Var}(X_1) + \mu^2 = 1 + \mu^2$, by applying the central limit theorem to $\{X_i^2\}$ we obtain that

$$\sqrt{n}(T_{1n} - \mu^2) = \sqrt{n}\left[\frac{1}{n}\sum_{i=1}^{n} X_i^2 - (1+\mu^2)\right] \to_d N(0,\gamma),$$

where $\gamma = \text{Var}(X_1^2)$. Also, by the central limit theorem, $\sqrt{n}(\bar{X} - \mu) \to_d N(0,1)$. When $\mu \neq 0$, by the δ-method and Slutsky's theorem,

$$\sqrt{n}(T_{2n} - \mu^2) = \sqrt{n}(\bar{X}^2 - \mu^2) - \frac{1}{\sqrt{n}} \to_d N(0, 4\mu^2).$$

When $\mu = 0$, $\sqrt{n}\bar{X} \to_d N(0,1)$ and, thus,

$$n(T_{2n} - \mu^2) = n\bar{X}^2 - 1 = (\sqrt{n}\bar{X})^2 - 1 \to_d W - 1,$$

where W has the chi-square distribution χ_1^2. Note that $E(W-1) = 0$ and $\text{Var}(W-1) = 2$. Therefore, the asymptotic relative efficiency of T_{1n} with respect to T_{2n} is equal to

$$e = \begin{cases} \frac{4\mu^2}{\text{Var}(X_1^2)} & \mu \neq 0 \\ \frac{2}{n\text{Var}(X_1^2)} & \mu = 0. \end{cases}$$

(ii) If the distribution of $X_1 - \mu$ is symmetric about 0, then $E(X_1 - \mu)^3 = 0$ and, thus,

$$\begin{aligned} \text{Var}(X_1^2) &= EX_1^4 - (EX_1^2)^2 \\ &= E[(X_1 - \mu) + \mu]^4 - (1+\mu^2)^2 \\ &= E(X_1-\mu)^4 + 4\mu E(X_1-\mu)^3 + 6\mu^2 E(X_1-\mu)^2 \\ &\quad + 4\mu^3 E(X_1 - \mu) + \mu^4 - (1 + 2\mu^2 + \mu^4) \\ &= E(X_1-\mu)^4 + 4\mu^2 - 1 \\ &\geq 4\mu^2, \end{aligned}$$

where the inequality follows from the Jensen's inequality $E(X_1 - \mu)^4 \geq [E(X_1 - \mu)^2]^2 = 1$. Therefore, when $\mu \neq 0$, the asymptotic relative efficiency $e \leq 1$.
(iii) Let the common distribution of X_i be the distribution of $Y/\sqrt{p(1-p)}$, where Y is a binary random variable with $P(Y=1) = p$ and $P(Y=0) = 1-p$. Then $EX_i = \sqrt{p/(1-p)} = \mu$, $\text{Var}(X_1) = 1$, and $EX_1^4 < \infty$. Note that

$$\text{Var}(X_1^2) = \text{Var}(Y^2)/[p^2(1-p)^2] = \text{Var}(Y)/[p^2(1-p)^2] = [p(1-p)]^{-1}.$$

Then the asymptotic relative efficiency is $e = 4\mu^2/\text{Var}(X_1^2) = 4p^2$, which is larger than 1 if $p \in (1/2, 1)$. ∎

Exercise 58 (#2.119). Let $(X_1, ..., X_n)$ be a random sample of random variables with unknown mean $\mu \in \mathcal{R}$, unknown variance $\sigma^2 > 0$, and $EX_1^4 < \infty$. Consider the estimation of μ^2 and the following three estimators: $T_{1n} = \bar{X}^2$, $T_{2n} = \bar{X}^2 - S^2/n$, $T_{3n} = \max\{0, T_{2n}\}$, where $\bar{X}$ and S^2 are the sample mean and variance.
(i) Show that the asymptotic mean squared errors of T_{jn}, $j = 1, 2, 3$, are the same when $\mu \neq 0$.
(ii) When $\mu = 0$, obtain the asymptotic relative efficiency of T_{2n} with respect to T_{1n} and the asymptotic relative efficiency of T_{3n} with respect to T_{2n}. Find out which estimator is asymptotically more efficient.
Solution. (i) By the central limit theorem and the δ-method,

$$\sqrt{n}(\bar{X}^2 - \mu^2) \to_d N(0, 4\mu^2\sigma^2).$$

By the law of large numbers, $S^2 \to_p \sigma^2$ and, hence, $S^2/\sqrt{n} \to_p 0$. By Slutsky's theorem,

$$\sqrt{n}(T_{2n} - \mu^2) = \sqrt{n}\bar{X}^2 - S^2/\sqrt{n} \to_d N(0, 4\mu^2\sigma^2).$$

This shows that, when $\mu \neq 0$, the asymptotic mean squared error of T_{2n} is the same as that of $T_{1n} = \bar{X}^2$. When $\mu \neq 0$, $\bar{X}^2 \to_p \mu^2 > 0$. Hence

$$\lim_n P(T_{2n} \neq T_{3n}) = \lim_n P(T_{2n} < 0) = \lim_n P(\bar{X}^2 < S^2/n) = 0,$$

since $S^2/n \to_p 0$. Therefore, the limiting distribution of $\sqrt{n}(T_{3n} - \mu^2)$ is the same as that of $\sqrt{n}(T_{2n} - \mu^2)$.
(ii) Assume $\mu = 0$. From $\sqrt{n}\bar{X} \to_d N(0, \sigma^2)$, we conclude that $n\bar{X}^2 \to_d \sigma^2 W$, where W has the chi-square distribution χ_1^2. Since $\mu = 0$, this shows that $n(T_{1n} - \mu^2) \to_d \sigma^2 W$ and, hence, the asymptotic mean squared error of T_{1n} is $\sigma^4 EW^2/n^2 = 3\sigma^4/n^2$. On the other hand, by Slutsky's theorem, $n(T_{2n} - \mu^2) = n\bar{X} - S^2 \to_p \sigma^2 W - \sigma^2$, since $S^2 \to_p \sigma^2$. Hence, the asymptotic mean squared error of T_{2n} is $\sigma^4 E(W-1)^2/n^2 = \sigma^4 \text{Var}(W)/n^2 = 2\sigma^4/n^2$. The asymptotic relative efficiency of T_{2n} with respect to T_{1n} is $3/2$. Hence T_{2n} is asymptotically more efficient than T_{1n}. Note that

$$n(T_{3n} - \mu^2) = n\max\{0, T_{2n}\} = \max\{0, nT_{2n}\} \to_d \max\{0, \sigma^2(W-1)\},$$

since $\max\{0, t\}$ is a continuous function of t. Then the asymptotic mean squared error of T_{3n} is $\sigma^4 E(\max\{0, W-1\})^2/n^2$ and The asymptotic relative efficiency of T_{3n} with respect to T_{2n} is $E(W-1)^2/E(\max\{0, W-1\})^2$. Since

$$E(\max\{0, W-1\})^2 = E[(W-1)^2 I_{\{W>1\}}] < E(W-1)^2,$$

we conclude that T_{3n} is asymptotically more efficient than T_{jn}, $j = 1, 2$. ∎

Exercise 59. Let $(X_1, ..., X_n)$ be a random sample from the exponential distribution $\theta^{-1}e^{-x/\theta}I_{(0,\infty)}(x)$, where $\theta \in (0, \infty)$. Consider the hypotheses $H_0 : \theta \leq \theta_0$ versus $H_1 : \theta > \theta_0$, where $\theta_0 > 0$ is a fixed constant. Let $T_c = I_{(c,\infty)}(\bar{X})$, where $\bar{X}$ is the sample mean.
(i) For any given level of significance $\alpha \in (0, 1)$, find a $c_{n,\alpha}$ such that the test $T_{c_{n,\alpha}}$ has size α and show that $T_{c_{n,\alpha}}$ is a consistent test, i.e., the power of $T_{c_{n,\alpha}}$ converges to 1 as $n \to \infty$ for any $\theta > \theta_0$.
(ii) Find a sequence $\{b_n\}$ such that the test T_{b_n} is consistent and the size of T_{b_n} converges to 0 as $n \to \infty$.
Solution. (i) Note that $\bar{X}/\theta$ has the gamma distribution with shape parameter n and scale parameter θ/n. Let $G_{n,\theta}$ denote the cumulative distribution function of this distribution and $c_{n,\alpha}$ be the constant satisfying $G_{n,\theta_0}(c_{n,\alpha}) = 1 - \alpha$. Then,

$$\sup_{\theta \leq \theta_0} P(T_{c_{n,\alpha}} = 1) = \sup_{\theta \leq \theta_0} [1 - G_{n,\theta}(c_{n,\alpha})] = 1 - G_{n,\theta_0}(c_{n,\alpha}) = \alpha,$$

i.e., the size of $T_{c_{n,\alpha}}$ is α.

Since the power of $T_{c_{n,\alpha}}$ is $P(T_{c_{n,\alpha}} = 1) = P(\bar{X} > c_{n,\alpha})$ for $\theta > \theta_0$ and, by the law of large numbers, $\bar{X} \to_p \theta$, the consistency of the test $T_{c_{n,\alpha}}$ follows if we can show that $\lim_n c_{n,\alpha} = \theta_0$. By the central limit theorem, $\sqrt{n}(\bar{X} - \theta) \to_d N(0, \theta^2)$. Hence, $\sqrt{n}(\frac{\bar{X}}{\theta} - 1) \to_d N(0, 1)$. By Pólya's theorem (e.g., Proposition 1.16 in Shao, 2003),

$$\lim_n \sup_t \left| P\left(\sqrt{n} \left(\tfrac{\bar{X}}{\theta} - 1 \right) \leq t \right) - \Phi(t) \right| = 0,$$

where Φ is the cumulative distribution function of the standard normal distribution. When $\theta = \theta_0$,

$$\alpha = P(\bar{X} \geq c_{n,\alpha}) = P\left(\sqrt{n} \left(\tfrac{\bar{X}}{\theta_0} - 1 \right) \geq \sqrt{n} \left(\tfrac{c_{n,\alpha}}{\theta_0} - 1 \right) \right).$$

Hence

$$\lim_n \Phi\left(\sqrt{n} \left(\tfrac{c_{n,\alpha}}{\theta_0} - 1 \right) \right) = 1 - \alpha,$$

which implies $\lim_n \sqrt{n}(\frac{c_{n,\alpha}}{\theta_0} - 1) = \Phi^{-1}(1 - \alpha)$ and, thus, $\lim_n c_{n,\alpha} = \theta_0$.
(ii) Let $\{a_n\}$ be a sequence of positive numbers such that $\lim_n a_n = 0$ and $\lim_n \sqrt{n}a_n = \infty$. Let $\alpha_n = 1 - \Phi(\sqrt{n}a_n)$ and $b_n = c_{n,\alpha_n}$, where $c_{n,\alpha}$ is defined in the proof of part (i). From the proof of part (i), the size of T_{b_n} is α_n, which converges to 0 as $n \to \infty$ since $\lim_n \sqrt{n}a_n = \infty$.

Using the same argument as that in the proof of part (i), we can show that

$$\lim_n \left| 1 - \alpha_n - \Phi\left(\sqrt{n} \left(\tfrac{c_{n,\alpha_n}}{\theta_0} - 1 \right) \right) \right| = 0,$$

which implies that

$$\lim_n \frac{\sqrt{n}}{\Phi^{-1}(1-\alpha_n)}\left(\frac{c_{n,\alpha_n}}{\theta_0}-1\right)=1.$$

Since $1-\alpha_n = \Phi(\sqrt{n}a_n)$, this implies that $\lim_n c_{n,\alpha_n} = \theta_0$. Since $b_n = c_{n,\alpha_n}$, the test T_{b_n} is consistent. ∎

Exercise 60 (#2.130). Let (Y_i, Z_i), $i = 1, ..., n$, be a random sample from a bivariate normal distribution and let ρ be the correlation coefficient between Y_1 and Z_1. Construct a confidence interval for ρ that has asymptotic significance level $1-\alpha$, based on the sample correlation coefficient

$$\hat{\rho} = \frac{1}{(n-1)\sqrt{S_Y^2 S_Z^2}}\sum_{i=1}^n (Y_i-\bar{Y})(Z_i-\bar{Z}),$$

where $\bar{Y}=n^{-1}\sum_{i=1}^n Y_i$, $\bar{Z}=n^{-1}\sum_{i=1}^n Z_i$, $S_Y^2=(n-1)^{-1}\sum_{i=1}^n(Y_i-\bar{Y})^2$, and $S_Z^2=(n-1)^{-1}\sum_{i=1}^n(Z_i-\bar{Z})^2$.

Solution. Assume first that $EY_1 = EZ_1 = 0$ and $\text{Var}(Y_1) = \text{Var}(Z_1) = 1$. From Exercise 9, $\sqrt{n}(\hat{\rho}-\rho) \to_d N(0, c^2)$ with

$$\begin{aligned} c^2 &= \rho^2[E(Y_1^4) + E(Z_1^4) + 2E(Y_1^2 Z_1^2)]/4 \\ &\quad - \rho[E(Y_1^3 Z_1) + E(Y_1 Z_1^3)] + E(Y_1^2 Z_1^2). \end{aligned}$$

We now derive the value of c^2. Under the normality assumption, $E(Y_1^4) = E(Z_1^4) = 3$. Let $U = Y_1 + Z_1$ and $V = Y_1 - Z_1$. Then U is distributed as $N(0, 2(1+\rho))$, V is distributed as $N(0, 2(1-\rho))$ and U and V are independent, since $\text{Cov}(U, V) = E(UV) = E(Y_1^2 - Z_1^2) = 0$. Note that $Y_1 = (U+V)/2$ and $Z_1 = (U-V)/2$. Then,

$$\begin{aligned} E(Y_1^2 Z_1^2) &= \frac{E\left[(U+V)^2(U-V)^2\right]}{16} \\ &= \frac{E\left(U^4+V^4-2U^2V^2\right)}{16} \\ &= \frac{EU^4+EV^4-2EU^2EV^2}{16} \\ &= \frac{3[2(1+\rho)]^2+3[2(1-\rho)]^2-2[2(1+\rho)][2(1-\rho)]}{16} \\ &= \frac{3[(1+\rho)^2+(1-\rho)^2]-2(1-\rho^2)}{4} \\ &= \frac{3(2+2\rho^2)-2+2\rho^2}{4} \\ &= 1+2\rho^2 \end{aligned}$$

and

$$\begin{aligned} E(Y_1^3 Z_1) &= \frac{E[(U+V)^3(U-V)]}{16} \\ &= \frac{E[(U+V)^3 U] - E[(U+V)^3 V]}{16} \\ &= \frac{EU^4 + 3E(U^2V^2) - EV^4 - 3E(U^2V^2)}{16} \\ &= \frac{3[2(1+\rho)]^2 - 3[2(1-\rho)]^2}{16} \\ &= \frac{3(1+\rho)^2 - 3(1-\rho)^2}{4} \\ &= 3\rho. \end{aligned}$$

By symmetry, $E(Y_1 Z_1^3) = 3\rho$. Using these results, we obtain that

$$\begin{aligned} c^2 &= \rho^2[3+3+2(1+2\rho^2)]/4 - 2\rho(3\rho) + 1 + 2\rho^2 \\ &= \rho^2(2+\rho^2) - 6\rho^2 + 1 + 2\rho^2 \\ &= \rho^4 - 2\rho^2 + 1 \\ &= (1-\rho^2)^2. \end{aligned}$$

In general, the distribution of $\hat{\rho}$ does not depend on the parameter vector $(EY_1, EZ_1, \text{Var}(Y_1), \text{Var}(Z_1))$, which can be shown by considering the transformation $(Y_i - EY_i)/\sqrt{\text{Var}(Y_i)}$ and $(Z_i - EZ_i)/\sqrt{\text{Var}(Z_i)}$. Hence,

$$\sqrt{n}(\hat{\rho} - \rho) \to_d N(0, (1-\rho)^2)$$

always holds, which implies that $\hat{\rho} \to_p \rho$. By Slutsky's theorem,

$$\frac{\sqrt{n}(\hat{\rho} - \rho)}{1 - \hat{\rho}^2} \to_d N(0, 1).$$

Hence

$$\lim_n P\left(-z_{\alpha/2} \leq \frac{\sqrt{n}(\hat{\rho} - \rho)}{1 - \hat{\rho}^2} \leq z_{\alpha/2}\right) = 1 - \alpha,$$

where z_a is the $(1-a)$th quantile of the standard normal distribution. Thus, a confidence interval for ρ that has asymptotic significance level $1 - \alpha$ is

$$\left[\, \hat{\rho} - (1-\hat{\rho}^2) z_{\alpha/2}/\sqrt{n}, \hat{\rho} + (1-\hat{\rho}^2) z_{\alpha/2}/\sqrt{n} \,\right]. \ \blacksquare$$

Chapter 3

Unbiased Estimation

Exercise 1. Let X be a sample from $P \in \mathcal{P}$ and θ be a parameter. Show that if both $T_1(X)$ and $T_2(X)$ are UMVUE's (uniformly minimum variance unbiased estimators) of θ with finite variances, then $T_1(X) = T_2(X)$ a.s. P for any $P \in \mathcal{P}$.
Solution. Since both T_1 and T_2 are unbiased, $T_1 - T_2$ is unbiased for 0. By the necessary and sufficient condition for UMVUE (e.g., Theorem 3.2 in Shao, 2003),

$$E[T_1(T_1 - T_2)] = 0 \quad \text{and} \quad E[T_2(T_1 - T_2)] = 0$$

for any P. Then, for any $P \in \mathcal{P}$,

$$E(T_1 - T_2)^2 = E[T_1(T_1 - T_2)] - E[T_2(T_1 - T_2)] = 0,$$

which implies that $T_1 = T_2$ a.s. P. ∎

Exercise 2 (#3.1). Let $(X_1, ..., X_n)$ be a sample of binary random variables with $P(X_i = 1) = p \in (0, 1)$.
(i) Find the UMVUE of p^m, where m is a positive integer and $m \leq n$.
(ii) Find the UMVUE of $P(X_1 + \cdots + X_m = k)$, where m and k are positive integers and $k \leq m \leq n$.
(iii) Find the UMVUE of $P(X_1 + \cdots + X_{n-1} > X_n)$.
Solution. (i) Let $T = \sum_{i=1}^n X_i$. Then T is a complete and sufficient statistic for p. By Lehmann-Scheffé's theorem (e.g., Theorem 3.1 in Shao, 2003), the UMVUE should be $h_m(T)$ with a Borel h_m satisfying $E[h_m(T)] = p^m$. We now try to find such a function h_m. Note that T has the binomial distribution with size n and probability p. Hence

$$E[h_m(T)] = \sum_{k=0}^{n} \binom{n}{k} h_m(k) p^k (1-p)^{n-k}.$$

Setting $E[h_m(T)] = p^m$, we obtain that

$$\sum_{k=0}^{n} \binom{n}{k} h_m(k) p^{k-m} (1-p)^{n-m-(k-m)} = 1$$

for all p. If $m < k$, $p^{k-m} \to \infty$ as $p \to 0$. Hence, we must have $h_m(k) = 0$ for $k = 0, 1, ..., m-1$. Then

$$\sum_{k=m}^{n} \binom{n}{k} h_m(k) p^{k-m} (1-p)^{n-m-(k-m)} = 1$$

for all p. On the other hand, from the property of a binomial distribution,

$$\sum_{k=m}^{n} \binom{n-m}{k-m} p^{k-m} (1-p)^{n-m-(k-m)} = 1$$

for all p. Hence, $\binom{n}{k} h_m(k) = \binom{n-m}{k-m}$ for $k = m, ..., n$. The UMVUE of p^m is

$$h_m(T) = \begin{cases} \frac{\binom{n-m}{T-m}}{\binom{n}{T}} & T = m, ..., n \\ 0 & T = 0, 1, ..., m-1. \end{cases}$$

(ii) Note that

$$\begin{aligned} P(X_1 + \cdots + X_m = k) &= \binom{m}{k} p^k (1-p)^{m-k} \\ &= \binom{m}{k} p^k \sum_{j=0}^{m-k} \binom{m-k}{j} (-1)^j p^j \\ &= \binom{m}{k} \sum_{j=0}^{m-k} \binom{m-k}{j} (-1)^j p^{j+k}. \end{aligned}$$

By the result in part (i), the UMVUE of p^{j+k} is $h_{j+k}(T)$, where the function h_{j+k} is given in part (i) of the solution, $j = 0, 1, ..., m-k$. By Corollary 3.1 in Shao (2003), the UMVUE of $P(X_1 + \cdots + X_m = k)$ is

$$\binom{m}{k} \sum_{j=0}^{m-k} \binom{m-k}{j} (-1)^j h_{j+k}(T).$$

(iii) Let $S_{n-1} = X_1 + \cdots + X_{n-1}$. Then S_{n-1} and X_n are independent and S_{n-1} has the binomial distribution with size $n-1$ and probability p.

Hence,

$$\begin{aligned}
P(S_{n-1} > X_n) &= P(X_n = 0)P(S_{n-1} > 0) + P(X_n = 1)P(S_{n-1} > 1) \\
&= P(S_{n-1} > 0) - P(X_n = 1)P(S_{n-1} = 1) \\
&= 1 - (1-p)^{n-1} - (n-1)p^2(1-p)^{n-2} \\
&= \sum_{j=1}^{n-1} \binom{n-1}{j} (-1)^{j+1} p^j - (n-1) \sum_{j=0}^{n-2} \binom{n-2}{j} (-1)^j p^{j+2} \\
&= \sum_{j=1}^{n} c_j p^j,
\end{aligned}$$

where $c_1 = n - 1$, $c_n = (-1)^{n+1}(n-1)$, and

$$c_j = (-1)^{j+1} \left[\binom{n-1}{j} + (n-1)\binom{n-2}{j-2} \right], \quad j = 2, ..., n-1.$$

The UMVUE of $P(S_{n-1} > X_n)$ is $\sum_{j=1}^{n} c_j h_j(T)$ with h_j defined in part (i) of the solution. ∎

Exercise 3 (#3.2). Let $(X_1, ..., X_n)$ be a random sample from $N(\mu, \sigma^2)$ with an unknown $\mu \in \mathcal{R}$ and a known $\sigma^2 > 0$.
(i) Find the UMVUE's of μ^3 and μ^4.
(ii) Find the UMVUE's of $P(X_1 \leq t)$ and $\frac{d}{dt}P(X_1 \leq t)$ with a fixed $t \in \mathcal{R}$.
Solution. (i) Let $\bar{X}$ be the sample mean, which is complete and sufficient for μ. Since

$$0 = E(\bar{X} - \mu)^3 = E(\bar{X}^3 - 3\mu\bar{X}^2 + 3\mu^2\bar{X} - \mu^3) = E(\bar{X}^3) - 3\mu\sigma^2/n - \mu^3,$$

we obtain that

$$E[\bar{X}^3 - (3\sigma^2/n)\bar{X}] = E(\bar{X}^3) - 3\mu\sigma^2/n = \mu^3$$

for all μ. By Lehmann-Scheffé's theorem, the UMVUE of μ^3 is $\bar{X}^3 - (3\sigma^2/n)\bar{X}$. Similarly,

$$\begin{aligned}
3\sigma^4 &= E(\bar{X} - \mu)^4 \\
&= E[\bar{X}(\bar{X} - \mu)^3] \\
&= E[\bar{X}^4 - 3\mu\bar{X}^3 + 3\mu^2\bar{X}^2 - \mu^3\bar{X}] \\
&= E(\bar{X}^4) - 3\mu(3\mu\sigma^2/n + \mu^3) + 3\mu^2(\sigma^2/n + \mu^2) - \mu^4 \\
&= E(\bar{X}^4) - 6\mu^2\sigma^2/n - 4\mu^4 \\
&= E(\bar{X}^4) - (6\sigma^2/n)E(\bar{X}^2 - \sigma^2/n) - 4\mu^4.
\end{aligned}$$

Hence, the UMVUE of μ^4 is $[\bar{X}^4 - (6\sigma^2/n)(\bar{X}^2 - \sigma^2/n) - 3\sigma^4]/4$.
(ii) Since $E[P(X_1 \leq t|\bar{X})] = P(X_1 \leq t)$, the UMVUE of $P(X_1 \leq t)$ is

$P(X_1 \le t|\bar{X})$. From the properties of normal distributions, $(X_1, \bar{X})$ is bivariate normal with mean (μ, μ) and covariance matrix

$$\sigma^2 \begin{pmatrix} 1 & n^{-1} \\ n^{-1} & n^{-1} \end{pmatrix}.$$

Consequently, the conditional distribution of X_1 given $\bar{X}$ is the normal distribution $N(\bar{X}, (1-n^{-1})\sigma^2)$. Then, the UMVUE of $P(X_1 \le t)$ is

$$\Phi\left(\frac{t-\bar{X}}{\sigma\sqrt{1-n^{-1}}}\right),$$

where Φ is the cumulative distribution function of $N(0,1)$. By the dominated convergence theorem,

$$\frac{d}{dt}P(X_1 \le t) = \frac{d}{dt}E\left[\Phi\left(\frac{t-\bar{X}}{\sigma\sqrt{1-n^{-1}}}\right)\right] = E\left[\frac{d}{dt}\Phi\left(\frac{t-\bar{X}}{\sigma\sqrt{1-n^{-1}}}\right)\right].$$

Hence, the UMVUE of $\frac{d}{dt}P(X_1 \le t)$ is

$$\frac{d}{dt}\Phi\left(\frac{t-\bar{X}}{\sigma\sqrt{1-n^{-1}}}\right) = \frac{1}{\sigma\sqrt{1-n^{-1}}}\Phi'\left(\frac{t-\bar{X}}{\sigma\sqrt{1-n^{-1}}}\right). \quad \blacksquare$$

Exercise 4 (#3.4). Let $(X_1, ..., X_m)$ be a random sample from $N(\mu_x, \sigma_x^2)$ and let $Y_1, ..., Y_n$ be a random sample from $N(\mu_y, \sigma_y^2)$. Assume that X_i's and Y_j's are independent.
(i) Assume that $\mu_x \in \mathcal{R}$, $\mu_y \in \mathcal{R}$, $\sigma_x^2 > 0$, and $\sigma_y^2 > 0$. Find the UMVUE's of $\mu_x - \mu_y$ and $(\sigma_x/\sigma_y)^r$, where $r > 0$ and $r < n$.
(ii) Assume that $\mu_x \in \mathcal{R}$, $\mu_y \in \mathcal{R}$, and $\sigma_x^2 = \sigma_y^2 > 0$. Find the UMVUE's of σ_x^2 and $(\mu_x - \mu_y)/\sigma_x$.
(iii) Assume that $\mu_x = \mu_y \in \mathcal{R}$, $\sigma_x^2 > 0$, $\sigma_y^2 > 0$, and $\sigma_x^2/\sigma_y^2 = \gamma$ is known. Find the UMVUE of μ_x.
(iv) Assume that $\mu_x = \mu_y \in \mathcal{R}$, $\sigma_x^2 > 0$, and $\sigma_y^2 > 0$. Show that a UMVUE of μ_x does not exist.
(v) Assume that $\mu_x \in \mathcal{R}$, $\mu_y \in \mathcal{R}$, $\sigma_x^2 > 0$, and $\sigma_y^2 > 0$. Find the UMVUE of $P(X_1 \le Y_1)$.
(vi) Repeat (v) under the assumption that $\sigma_x = \sigma_y$.
Solution: (i) The complete and sufficient statistic for $(\mu_x, \mu_y, \sigma_x^2, \sigma_y^2)$ is $(\bar{X}, \bar{Y}, S_X^2, S_Y^2)$, where $\bar{X}$ and S_X^2 are the sample mean and variance based on X_i's and $\bar{Y}$ and S_Y^2 are the sample mean and variance based on Y_i's. Therefore $\bar{X} - \bar{Y}$ is the UMVUE of $\mu_x - \mu_y$. A direct calculation shows that

$$E(S_X^r) = \sigma_x^r/\kappa_{m-1,r},$$

where

$$\kappa_{m,r} = \frac{m^{r/2}\Gamma(\frac{m}{2})}{2^{r/2}\Gamma(\frac{m+r}{2})}.$$

Hence, the UMVUE of σ_x^r is $\kappa_{m-1,r}S_X^r$. Similarly, the UMVUE of σ_y^{-r} is $\kappa_{n-1,-r}S_Y^{-r}$. Since S_X and S_Y are independent, the UMVUE of $(\sigma_x/\sigma_y)^r$ is $\kappa_{m-1,r}\kappa_{n-1,-r}S_X^r S_Y^{-r}$.
(ii) The complete and sufficient statistic for $(\mu_x, \mu_y, \sigma_x^2)$ is $(\bar{X}, \bar{Y}, S^2)$, where

$$S^2 = \frac{1}{m+n-2}\left[\sum_{i=1}^{m}(X_i - \bar{X})^2 + \sum_{j=1}^{n}(Y_j - \bar{Y})^2\right].$$

Since $(m+n-2)S^2/\sigma_x^2$ has the chi-square distribution χ^2_{m+n-2}, the UMVUE of σ_x^2 is S^2 and the UMVUE of σ_x^{-1} is $\kappa_{m+n-2,-1}S^{-1}$. Since $\bar{X} - \bar{Y}$ and S^2 are independent, $\kappa_{m+n-2,-1}(\bar{X} - \bar{Y})/S$ is the UMVUE of $(\mu_x - \mu_y)/\sigma_x$.
(iii) The joint distribution of X_i's and Y_j's is from an exponential family with $(m\bar{X} + \gamma n\bar{Y}, \sum_{i=1}^m X_i^2 + \gamma\sum_{j=1}^n Y_j^2)$ as the complete and sufficient statistic for (μ_x, σ_x^2). Hence, the UMVUE of μ_x is $(m\bar{X} + \gamma n\bar{Y})/(m + \gamma n)$.
(iv) Let $\mathcal{P}$ be the family of all possible distributions of $(X_1, ..., X_m, Y_1, ..., Y_n)$ and $\mathcal{P}_\gamma$ be the sub-family of $\mathcal{P}$ with $\sigma_x^2 = \gamma\sigma_y^2$. Suppose that T is a UMVUE of μ_x. By the result in (iii), $T_\gamma = (m\bar{X} + \gamma n\bar{Y})/(m + \gamma n)$ is a UMVUE of μ_x when $\mathcal{P}_\gamma$ is considered as the family of distributions for $(X_1, ..., X_m, Y_1, ..., Y_n)$. Since $E(T - T_\gamma) = 0$ for any $P \in \mathcal{P}$ and T is a UMVUE, $E[T(T - T_\gamma)] = 0$ for any $P \in \mathcal{P}$. Similarly, $E[T_\gamma(T - T_\gamma)] = 0$ for any $P \in \mathcal{P}_\gamma$. Then, $E(T - T_\gamma)^2 = 0$ for any $P \in \mathcal{P}_\gamma$ and, thus, $T = T_\gamma$ a.s. $\mathcal{P}_\gamma$. Since a.s. $\mathcal{P}_\gamma$ implies a.s. $\mathcal{P}$, $T = T_\gamma$ a.s. $\mathcal{P}$ for any $\gamma > 0$. This shows that T depends on $\gamma = \sigma_x^2/\sigma_y^2$, which is impossible.
(v) Since $U = (\bar{X}, \bar{Y}, S_X^2, S_Y^2)$ is complete and sufficient for $(\mu_x, \mu_y, \sigma_x^2, \sigma_y^2)$, $P(X_1 \leq Y_1|U)$ is UMVUE for $P(X_1 \leq Y_1)$. Note that

$$P\big(X_1 \leq t, Y_1 \leq v|U = (\bar{x}, \bar{y}, s_x^2, s_y^2)\big) = P\left(Z \leq \frac{t - \bar{x}}{s_x}, W \leq \frac{v - \bar{y}}{s_y}\right),$$

where $Z = (X_1 - \bar{X})/S_X$ and $W = (Y_1 - \bar{Y})/S_Y$. From Example 3.4 in Shao (2003), Z has Lebesgue density $f_m(z)$ and W has Lebesgue density $f_n(w)$, where

$$f_k(z) = \frac{\sqrt{k}\Gamma\left(\frac{k-1}{2}\right)}{\sqrt{\pi}(k-1)\Gamma\left(\frac{k-2}{2}\right)}\left[1 - \frac{kz^2}{(k-1)^2}\right]^{(k/2)-2} I_{(0,(k-1)/\sqrt{k})}(|z|).$$

Since Z and W are independent, the conditional density of (X_1, Y_1) given U is

$$\frac{1}{S_X}f_m\left(\frac{t - \bar{X}}{S_X}\right)\frac{1}{S_Y}f_n\left(\frac{v - \bar{Y}}{S_Y}\right).$$

Hence, the UMVUE is

$$P(X_1 \leq Y_1|U) = \frac{1}{S_X S_Y} \int_{-\infty}^{0} \int_{-\infty}^{\infty} f_m\left(\frac{v - \bar{X}}{S_X}\right) f_n\left(\frac{t - v - \bar{Y}}{S_Y}\right) dv.$$

(vi) In this case, $U = (\bar{X}, \bar{Y}, S^2)$ with S^2 defined in (ii) is complete and sufficient for $(\mu_x, \mu_y, \sigma_x^2)$. Similar to part (v) of the solution, we have

$$P(X_1 \leq Y_1|U = u) = P\left(\frac{(X_1 - \bar{X}) - (Y_1 - \bar{Y})}{\sqrt{m + n - 2S}} \leq r\right),$$

where r is the observed value of $R = -(\bar{X} - \bar{Y})/(\sqrt{m + n - 2}S)$. If we denote the Lebesgue density of $T = [(X_1 - \bar{X}) - (Y_1 - \bar{Y})]/(\sqrt{m + n - 2}S)$ by $f(t)$, then the UMVUE of $P(X_1 \leq Y_1)$ is $\int_{-\infty}^{R} f(t)dt$. To determine f, we consider the orthogonal transformation

$$(Z_1, ..., Z_{m+n})^\tau = A(X_1, ..., X_m, Y_1, ..., Y_n)^\tau,$$

where A is an orthogonal matrix of order $m + n$ whose first three rows are

$$(m^{-1/2} J_m, 0J_n),$$

$$(0J_m, n^{-1/2} J_n),$$

and

$$(2 - m^{-1} - n^{-1})^{-1/2}\big(1 - m^{-1}, -m^{-1} J_{m-1}, n^{-1} - 1, n^{-1} J_{n-1}\big),$$

and J_k denotes a row of 1's with dimension k. Then $Z_1 = \sqrt{m}\bar{X}$, $Z_2 = \sqrt{n}\bar{Y}$, $Z_3 = (2 - m^{-1} - n^{-1})^{-1}[(X_1 - \bar{X}) - (Y_1 - \bar{Y})]$, $(m + n - 2)S^2 = \sum_{i=3}^{m+n} Z_3^2$, and Z_i, $i = 3, ..., m + n$, are independent and identically distributed as $N(0, \sigma_x^2)$. Note that

$$T = \frac{\sqrt{2 - m^{-1} - n^{-1}} Z_3}{\sqrt{Z_3^2 + Z_4^2 + \cdots + Z_{m+n}^2}}.$$

Then, a direct calculation shows that

$$f(t) = c_{m,n}\left(1 - \frac{t^2}{2 - m^{-1} - n^{-1}}\right)^{(m+n-5)/2} I_{(0, \sqrt{2-m^{-1}-n^{-1}})}(|t|),$$

where

$$c_{m,n} = \frac{\Gamma(\frac{m+n-2}{2})}{\sqrt{\pi(2 - m^{-1} - n^{-1})}\Gamma(\frac{m+n-3}{2})}. \blacksquare$$

Exercise 5 (#3.5). Let $(X_1, ..., X_n)$, $n > 2$, be a random sample from the uniform distribution on the interval $(\theta_1 - \theta_2, \theta_1 + \theta_2)$, where $\theta_1 \in \mathcal{R}$ and $\theta_2 > 0$. Find the UMVUE's of θ_j, $j = 1, 2$, and θ_1/θ_2.

Solution. Let $X_{(j)}$ be the jth order statistic. Then $(X_{(1)}, X_{(n)})$ is complete and sufficient for (θ_1, θ_2). Hence, it suffices to find a function of $(X_{(1)}, X_{(n)})$ that is unbiased for the parameter of interest. Let $Y_i = [X_i - (\theta_1 - \theta_2)]/(2\theta_2)$, $i = 1, ..., n$. Then Y_i's are independent and identically distributed as the uniform distribution on the interval $(0, 1)$. Let $Y_{(j)}$ be the jth order statistic of Y_i's. Then,

$$\begin{aligned} E(X_{(n)}) &= 2\theta_2 E(Y_{(n)}) + \theta_1 - \theta_2 \\ &= 2\theta_2 n \int_0^1 y^n dy + \theta_1 - \theta_2 \\ &= \frac{2\theta_2 n}{n+1} + \theta_1 - \theta_2 \end{aligned}$$

and

$$\begin{aligned} E(X_{(1)}) &= 2\theta_2 E(Y_{(1)}) + \theta_1 - \theta_2 \\ &= 2\theta_2 n \int_0^1 y(1-y)^{n-1} dy + \theta_1 - \theta_2 \\ &= -\frac{2\theta_2 n}{n+1} + \theta_1 + \theta_2. \end{aligned}$$

Hence, $E(X_{(n)} + X_{(1)})/2 = \theta_1$ and $E(X_{(n)} - X_{(1)}) = 2\theta_2(n-1)/(n+1)$. Therefore, the UMVUE's of θ_1 and θ_2 are, respectively, $(X_{(n)} + X_{(1)})/2$ and $(n+1)(X_{(n)} + X_{(1)})/[2(n-1)]$. Furthermore,

$$\begin{aligned} E\left(\frac{X_{(n)} + X_{(1)}}{X_{(n)} - X_{(1)}}\right) &= E\left(\frac{Y_{(n)} + Y_{(1)}}{Y_{(n)} - Y_{(1)}}\right) + \frac{\theta_1 - \theta_2}{\theta_2} E\left(\frac{1}{Y_{(n)} - Y_{(1)}}\right) \\ &= n(n-1) \int_0^1 \int_0^y (x+y)(y-x)^{n-3} dx dy \\ &\quad + \frac{\theta_1 - \theta_2}{\theta_2} n(n-1) \int_0^1 \int_0^y (y-x)^{n-3} dx dy \\ &= \frac{n}{n-2} + \frac{\theta_1 - \theta_2}{\theta_2} \frac{n}{n-2} \\ &= \frac{n}{n-2} \frac{\theta_1}{\theta_2}. \end{aligned}$$

Hence the UMVUE of θ_1/θ_2 is $\frac{n-2}{n}(X_{(n)} + X_{(1)})/(X_{(n)} - X_{(1)})$. ▮

Exercise 6 (#3.6). Let $(X_1, ..., X_n)$ be a random sample from the exponential distribution on (a, ∞) with scale parameter θ, where $\theta > 0$ and

$a \in \mathcal{R}$.
(i) Find the UMVUE of a when θ is known.
(ii) Find the UMVUE of θ when a is known.
(iii) Find the UMVUE's of θ and a.
(iv) Assume that θ is known. Find the UMVUE of $P(X_1 \geq t)$ and the UMVUE of $\frac{d}{dt}P(X_1 \geq t)$ for a fixed $t > a$.
(v) Find the UMVUE of $P(X_1 \geq t)$ for a fixed $t > a$.

Solution: (i) When θ is known, the smallest order statistic $X_{(1)}$ is complete and sufficient for a. Since $EX_{(1)} = a + \theta/n$, $X_{(1)} - \theta/n$ is the UMVUE of a.

(ii) When a is known, $T = \sum_{i=1}^n X_i$ is complete and sufficient for θ. Since $ET = n(a + \theta)$, $T/n - a$ is the UMVUE of θ.

(iii) Note that $(X_{(1)}, T - nX_{(1)})$ is complete and sufficient for (a, θ) and $2(T - nX_{(1)})/\theta$ has the chi-square distribution $\chi^2_{2(n-1)}$. Then $E(T - nX_{(1)}) = (n-1)\theta$ and the UMVUE of θ is $(T - nX_{(1)})/(n-1)$. Since $EX_{(1)} = a + \theta/n$, the UMVUE of a is $X_{(1)} - (T - nX_{(1)})/[n(n-1)]$.

(iv) Since $X_{(1)}$ is complete and sufficient for a, the UMVUE of

$$P(X_1 \geq t) = \begin{cases} e^{(a-t)/\theta} & t > a \\ 1 & t \leq a \end{cases}$$

is $g(X_{(1)})$ satisfying

$$P(X_1 \geq t) = E[g(X_{(1)})] = \frac{n}{\theta} \int_a^\infty g(x) e^{-n(x-a)/\theta} dx$$

for any a, which is the same as

$$\frac{ne^{t/\theta}}{\theta} \int_a^\infty g(x) e^{-nx/\theta} dx = e^{-(n-1)a/\theta}$$

for any $a < t$ and $g(a) = 1$ for $a \geq t$. Differentiating both sides of the above expression with respect to a, we obtain that

$$ne^{t/\theta} g(a) e^{-na/\theta} = (n-1) e^{-(n-1)a/\theta}.$$

Hence,

$$g(x) = \begin{cases} (1 - n^{-1}) e^{(x-t)/\theta} & x < t \\ 1 & x \geq t \end{cases}$$

and the UMVUE of $P(X_1 > t)$ is $g(X_{(1)})$. The UMVUE of $\frac{d}{dt}P(X_1 \geq t) = -\theta^{-1} e^{(a-t)/\theta}$ is then $-\theta^{-1} g(X_{(1)})$.

(v) The complete and sufficient statistic for (a, θ) is $U = (X_{(1)}, T - nX_{(1)})$. The UMVUE is $P(X_1 \geq t | U)$. Let $Y = T - nX_{(1)}$ and $A_j = \{X_{(1)} = X_j\}$.

Then $P(A_j) = n^{-1}$. If $t < X_{(1)}$, obviously $P(X_1 \geq t|U) = 1$. For $t \geq X_{(1)}$, consider $U = u = (x_{(1)}, y)$ and

$$\begin{aligned} P\big(X_1 \geq t|U = u\big) &= P\left(\frac{X_1 - X_{(1)}}{Y} \geq \frac{t - x_{(1)}}{y}\bigg|U = u\right) \\ &= P\left(\frac{X_1 - X_{(1)}}{Y} \geq \frac{t - x_{(1)}}{y}\right) \\ &= \sum_{j=1}^{n} P(A_j) P\left(\frac{X_1 - X_{(1)}}{Y} \geq \frac{t - x_{(1)}}{y}\bigg|A_j\right) \\ &= \frac{n-1}{n} P\left(\frac{X_1 - X_{(1)}}{Y} \geq \frac{t - x_{(1)}}{y}\bigg|A_n\right) \\ &= \frac{n-1}{n} P\left(\frac{X_1 - X_{(1)}}{\sum_{i=1}^{n-1}(X_i - X_{(1)})} \geq \frac{t - x_{(1)}}{y}\bigg|A_n\right) \\ &= \frac{n-1}{n}\left(1 - \frac{t - x_{(1)}}{y}\right)^{n-2}, \end{aligned}$$

where the second equality follows from the fact that U and $(X_1 - X_{(1)})/Y$ are independent (Basu's theorem), the fourth equality follows from the fact that the conditional probability given A_1 is 0 and the conditional probabilities given A_j, $j = 2, ..., n$, are all the same, the fifth equality follows from the fact that $Y = \sum_{i=1}^{n-1}(X_i - X_{(1)})$ on the event A_n, and the last equality follows from the fact that conditional on A_n, $X_i - X_{(1)}$, $i = 1, ..., n-1$, are independent and identically distributed as the exponential distribution on $(0, \infty)$ with scale parameter θ and $(X_1 - X_{(1)})/\sum_{i=1}^{n-1}(X_i - X_{(1)})$ has the beta distribution with density $(n-2)(1-x)^{n-3}I_{(0,1)}(x)$. Therefore, the UMVUE is equal to 1 when $t < X_{(1)}$ and

$$\left(1 - \frac{1}{n}\right)\left[1 - \frac{t - X_{(1)}}{\sum_{i=1}^{n}(X_i - X_{(1)})}\right]^{n-2}$$

when $X_{(1)} \leq t$. ∎

Exercise 7 (#3.7). Let $(X_1, ..., X_n)$ be a random sample from the Pareto distribution with Lebesgue density $\theta a^\theta x^{-(\theta+1)} I_{(a,\infty)}(x)$, where $\theta > 0$ and $a > 0$.
(i) Find the UMVUE of θ when a is known.
(ii) Find the UMVUE of a when θ is known.
(iii) Find the UMVUE's of a and θ.
Solution: (i) The joint Lebesgue density of $X_1, ..., X_n$ is

$$f(x_1, ..., x_n) = \theta^n a^{n\theta} \exp\left\{-(\theta + 1)\sum_{i=1}^{n} \log x_i\right\} I_{(a,\infty)}(x_{(1)}),$$

where $x_{(1)} = \min_{1\leq i\leq n} x_i$. When a is known, $T = \sum_{i=1}^n \log X_i$ is complete and sufficient for θ and $T - n\log a$ has the gamma distribution with shape parameter n and scale parameter θ^{-1}. Hence, $ET^{-1} = \theta/(n-1)$ and, thus, $(n-1)/T$ is the UMVUE of θ.
(ii) When θ is known, $X_{(1)}$ is complete and sufficient for a. Since $X_{(1)}$ has the Lebesgue density $n\theta a^{n\theta} x^{-(n\theta+1)} I_{(a,\infty)}(x)$, $EX_{(1)} = n\theta a/(n\theta - 1)$. Therefore, $(1 - n\theta)X_{(1)}/(n\theta)$ is the UMVUE of a.
(iii) When both a and θ are unknown, $(Y, X_{(1)})$ is complete and sufficient for (a, θ), where $Y = \sum_i (\log X_i - \log X_{(1)})$. Also, Y has the gamma distribution with shape parameter $n-1$ and scale parameter θ^{-1} and $X_{(1)}$ and Y are independent. Since $EY^{-1} = \theta/(n-2)$, $(n-2)/Y$ is the UMVUE of θ. Since

$$
\begin{aligned}
E\left\{\left[1 - \frac{Y}{n(n-1)}\right] X_{(1)}\right\} &= \left[1 - \frac{EY}{n(n-1)}\right] EX_{(1)} \\
&= \left(1 - \frac{1}{n\theta}\right) \frac{n\theta a}{n\theta - 1} \\
&= a,
\end{aligned}
$$

$\left[1 - \frac{Y}{n(n-1)}\right] X_{(1)}$ is the UMVUE of a. ∎

Exercise 8 (#3.11). Let X be a random variable having the negative binomial distribution with

$$
P(X = x) = \binom{x-1}{r-1} p^r (1-p)^{x-r}, \quad x = r, r+1, ...,
$$

where $p \in (0, 1)$ and r is a known positive integer.
(i) Find the UMVUE of p^t, where t is a positive integer and $t < r$.
(ii) Find the UMVUE of $\mathrm{Var}(X)$.
(iii) Find the UMVUE of $\log p$.
Solution. (i) Since X is complete and sufficient for p, the UMVUE of p^t is $h(X)$ with a function h satisfying $E[h(X)] = p^t$ for any p, i.e.,

$$
\sum_{x=r}^{\infty} h(x) \binom{x-1}{r-1} p^r (1-p)^{x-r} = p^t
$$

for any p. Let $q = 1 - p$. Then

$$
\sum_{x=r}^{\infty} h(x) \binom{x-1}{r-1} q^x = \frac{q^r}{(1-q)^{r-t}}
$$

for any $q \in (0, 1)$. From the negative binomial identity

$$
\sum_{x=j}^{\infty} \binom{x-1}{j-1} q^x = \frac{q^j}{(1-q)^j}
$$

with any positive integer j, we obtain that

$$\sum_{x=r}^{\infty} h(x)\binom{x-1}{r-1}q^x = \sum_{x=r-t}^{\infty} \binom{x-1}{r-t-1}q^{x+t} = \sum_{x=r}^{\infty} \binom{x-t-1}{r-t-1}q^x$$

for any q. Comparing the coefficients of q^x, we obtain that

$$h(x) = \frac{\binom{x-t-1}{r-t-1}}{\binom{x-1}{r-1}}, \quad x = r, r+1, ...$$

(ii) Note that $\text{Var}(X) = r(1-p)/p^2 = rq/(1-q)^2$. The UMVUE of $\text{Var}(X)$ is $h(X)$ with $E[h(X)] = rq/(1-q)^2$ for any $q \in (0,1)$. That is,

$$\sum_{x=r}^{\infty} h(x)\binom{x-1}{r-1}q^x = \frac{q^r}{(1-q)^r}\text{Var}(X) = r\frac{q^{r+1}}{(1-q)^{r+2}}$$

for any q. Using the negative binomial identity, this means that

$$\sum_{x=r}^{\infty} h(x)\binom{x-1}{r-1}q^x = r\sum_{x=r+2}^{\infty} \binom{x-1}{r+1}q^{x-1} = r\sum_{x=r+1}^{\infty} \binom{x}{r+1}q^x$$

for any q, which yields

$$h(x) = \begin{cases} 0 & x = r \\ \frac{r\binom{x}{r+1}}{\binom{x-1}{r-1}} & x = r+1, r+2, ... \end{cases}$$

(iii) Let $h(X)$ be the UMVUE of $\log p = \log(1-q)$. Then, for any $q \in (0,1)$,

$$\begin{aligned}
\sum_{x=r}^{\infty} h(x)\binom{x-1}{r-1}q^x &= \frac{q^r}{(1-q)^r}\log(1-q) \\
&= -\sum_{x=r}^{\infty}\binom{x-1}{r-1}q^x \sum_{i=1}^{\infty}\frac{q^i}{i} \\
&= \sum_{x=r+1}^{\infty}\sum_{k=0}^{x-r-1}\binom{r+k-1}{k}\frac{q^x}{k+r-x}.
\end{aligned}$$

Hence $h(r) = 0$ and

$$h(x) = \frac{1}{\binom{x-1}{r-1}}\sum_{k=0}^{x-r-1}\binom{r+k-1}{k}\frac{1}{k+r-x}$$

for $x = r+1, r+2, ...$ ∎

Exercise 9 (#3.12). Let $(X_1, ..., X_n)$ be a random sample from the Poisson distribution truncated at 0, i.e., $P(X_i = x) = (e^\theta - 1)^{-1}\theta^x/x!$, $x = 1, 2, ...$, $\theta > 0$. Find the UMVUE of θ when $n = 1, 2$.

Solution. Assume $n = 1$. Then X is complete and sufficient for θ and the UMVUE of θ is $h(X)$ with $E[h(X)] = \theta$ for any θ. Since

$$E[h(X)] = \frac{1}{e^\theta - 1} \sum_{x=1}^{\infty} h(x) \frac{\theta^x}{x!},$$

we must have

$$\sum_{x=1}^{\infty} \frac{h(x)\theta^x}{x!} = \theta(e^\theta - 1) = \theta \sum_{x=1}^{\infty} \frac{\theta^x}{x!} = \sum_{x=2}^{\infty} \frac{\theta^x}{(x-1)!}$$

for any θ. Comparing the coefficient of θ^x leads to $h(1) = 0$ and $h(x) = x$ for $x = 2, 3, ...$

Assume $n = 2$. Then $T = X_1 + X_2$ is complete and sufficient for θ. The UMVUE of θ is $h(T)$ with $E[h(T)] = \theta$ for any θ. Then

$$\theta(e^\theta - 1)^2 = \sum_{i=1}^{\infty} \sum_{j=1}^{\infty} \frac{h(i+j)e^{i+j}}{i!j!} = \sum_{t=2}^{\infty} h(t)\theta^t \sum_{i=0}^{t-1} \frac{1}{i!(t-i)!}.$$

On the other hand,

$$\theta(e^\theta - 1)^2 = \theta \left(\sum_{i=1}^{\infty} \frac{\theta^i}{i!} \right)^2 = \sum_{i=1}^{\infty} \sum_{j=1}^{\infty} \frac{\theta^i \theta^{j+1}}{i!j!} = \sum_{t=3}^{\infty} \theta^t \sum_{i=0}^{t-2} \frac{1}{i!(t-1-i)!}.$$

Comparing the coefficient of θ^x leads to $h(2) = 0$ and

$$h(t) = \sum_{i=0}^{t-2} \frac{1}{i!(t-1-i)!} \Bigg/ \sum_{i=0}^{t-1} \frac{1}{i!(t-i)!}$$

for $t = 3, 4, ...$ ∎

Exercise 10 (#3.14). Let $X_1, ..., X_n$ be a random sample from the log-distribution with

$$P(X_1 = x) = -(1-p)^x/(x \log p), \quad x = 1, 2, ...,$$

$p \in (0, 1)$. Let k be a fixed positive integer.

(i) For $n = 1, 2, 3$, find the UMVUE of p^k.

(ii) For $n = 1, 2, 3$, find the UMVUE of $P(X = k)$.

Solution. (i) Let $\theta = 1-p$. Then $p^k = \sum_{r=0}^{k} \binom{k}{r} (-1)^r \theta^r$. Hence, it suffices to obtain the UMVUE for θ^r. Note that the distribution of X_1 is from a power series distribution with $\gamma(x) = x^{-1}$ and $c(\theta) = -\log(1-\theta)$ (see Example 3.5 in Shao, 2003). The statistic $T = \sum_{i=1}^{n} X_i$ is complete and sufficient for θ. By the result in Example 3.5 of Shao (2003), the UMVUE of θ^r is

$$\frac{\gamma_n(T-r)}{\gamma_n(T)} I_{\{r,r+1,...\}}(T),$$

where $\gamma_n(t)$ is the coefficient of θ^t in $\left(\sum_{y=1}^{\infty} \frac{\theta^y}{y}\right)^n$, i.e., $\gamma_n(t) = 0$ for $t < n$ and

$$\gamma_n(t) = \sum_{y_1+\cdots+y_n=t-n, y_i \ge 0} \frac{1}{(y_1+1)\cdots(y_n+1)}$$

for $t = n, n+1, ...$ When $n = 1, 2, 3$, $\gamma_n(t)$ has a simpler form. In fact, $\gamma_1(1) = 0$ and

$$\gamma_1(t) = t^{-1}, \quad t = 2, 3, ...;$$

$\gamma_2(1) = \gamma_2(2) = 0$ and

$$\gamma_2(t) = \sum_{l=0}^{t-2} \frac{1}{(l+1)(t-l-1)}, \quad t = 3, 4, ...;$$

$\gamma_3(1) = \gamma_3(2) = \gamma_3(3) = 0$ and

$$\gamma_3(t) = \sum_{l_1=0}^{t-3}\sum_{l_2=0}^{t-3} \frac{1}{(l_1+1)(l_2+1)(t-l_1-l_2-2)}, \quad t = 4, 5, ...$$

(ii) By Example 3.5 in Shao (2003), the UMVUE of $P(X_1 = k)$ is

$$\frac{\gamma_{n-1}(T-k)}{k\gamma_n(T)} I_{\{k,k+1,...\}}(T),$$

where $\gamma_n(t)$ is given in the solution of part (i). ∎

Exercise 11 (#3.19). Let $Y_1, ..., Y_n$ be a random sample from the uniform distribution on the interval $(0, \theta)$ with an unknown $\theta \in (1, \infty)$.
(i) Suppose that we only observe

$$X_i = \begin{cases} Y_i & \text{if } Y_i \ge 1 \\ 1 & \text{if } Y_i < 1, \end{cases} \qquad i = 1, ..., n.$$

Derive a UMVUE of θ.
(ii) Suppose that we only observe

$$X_i = \begin{cases} Y_i & \text{if } Y_i \le 1 \\ 1 & \text{if } Y_i > 1, \end{cases} \qquad i = 1, ..., n.$$

Derive a UMVUE of the probability $P(Y_1 > 1)$.
Solution. (i) Let m be the Lebesgue measure and δ be the point mass on $\{1\}$. The joint probability density of $X_1, ..., X_n$ with respect to $\delta+m$ is (see, e.g., Exercise 16 in Chapter 1) $\theta^{-n} I_{(0,\theta)}(X_{(n)})$, where $X_{(n)} = \max_{1\le i\le n} X_i$. Hence $X_{(n)}$ is complete and sufficient for θ and the UMVUE of θ is $h(X_{(n)})$ satisfying $E[h(X_{(n)})] = \theta$ for all $\theta > 1$. The probability density of $X_{(n)}$ with respect to $\delta + m$ is $\theta^{-n} I_{\{1\}}(x) + n\theta^{-n} x^{n-1} I_{(1,\theta)}(x)$. Hence

$$E[h(X_{(n)})] = \frac{h(1)}{\theta^n} + \frac{n}{\theta^n}\int_1^\theta h(x)x^{n-1}dx.$$

Then

$$\theta^{n+1} = h(1) + n\int_1^\theta h(x)x^{n-1}dx$$

for all $\theta > 1$. Letting $\theta \to 1$ we obtain that $h(1) = 1$. Differentiating both sides of the previous expression with respect to θ we obtain that

$$(n+1)\theta^n = nh(\theta)\theta^{n-1} \qquad \theta > 1.$$

Hence $h(x) = (n+1)x/n$ when $x > 1$.
(ii) The joint probability density of $X_1, ..., X_n$ with respect to $\delta + m$ is $\theta^{-r}(1-\theta^{-1})^{n-r}$, where r is the observed value of R = the number of X_i's that are less than 1. Hence, R is complete and sufficient for θ. Note that R has the binomial distribution with size n and probability θ^{-1} and $P(Y_1 > 1) = 1 - \theta^{-1}$. Hence, the UMVUE of $P(Y_1 > 1)$ is $1 - R/n$. ∎

Exercise 12 (#3.22). Let $(X_1, ..., X_n)$ be a random sample from $P \in \mathcal{P}$ containing all symmetric distributions with finite means and with Lebesgue densities on $\mathcal{R}$.
(i) When $n = 1$, show that X_1 is the UMVUE of μ.
(ii) When $n > 1$, show that there is no UMVUE of $\mu = EX_1$.
Solution. (i) Consider the sub-family $\mathcal{P}_1 = \{N(\mu, 1) : \mu \in \mathcal{R}\}$. Then X_1 is complete for $P \in \mathcal{P}_1$. Hence, $E[h(X_1)] = 0$ for any $P \in \mathcal{P}$ implies that $E[h(X_1)] = 0$ for any $P \in \mathcal{P}_1$ and, thus, $h = 0$ a.e. Lebesgue measure. This shows that 0 is the unique estimator of 0 when the family $\mathcal{P}$ is considered. Since $EX_1 = \mu$, X_1 is the unique unbiased estimator of μ and, hence, it is the UMVUE of μ.
(ii) Suppose that T is a UMVUE of μ. Let $\mathcal{P}_1 = \{N(\mu, 1) : \mu \in \mathcal{R}\}$. Since the sample mean $\bar{X}$ is UMVUE when $\mathcal{P}_1$ is considered, by using the same argument in the solution for Exercise 4(iv), we can show that $T = \bar{X}$ a.s. P for any $P \in \mathcal{P}_1$. Since the Lebesgue measure is dominated by any $P \in \mathcal{P}_1$, we conclude that $T = \bar{X}$ a.e. Lebesgue measure. Let $\mathcal{P}_2$ be the family given in Exercise 5. Then $(X_{(1)} + X_{(n)})/2$ is the UMVUE when $\mathcal{P}_2$ is considered, where $X_{(j)}$ is the jth order statistic. Then $\bar{X} = (X_{(1)} + X_{(n)})/2$ a.s. P for any $P \in \mathcal{P}_2$, which is impossible. Hence, there is no UMVUE of μ. ∎

Exercise 13 (#3.24). Suppose that T is a UMVUE of an unknown parameter θ. Show that T^k is a UMVUE of $E(T^k)$, where k is any positive integer for which $E(T^{2k}) < \infty$.
Solution. Let U be an unbiased estimator of 0. Since T is a UMVUE of θ, $E(TU) = 0$ for any P, which means that TU is an unbiased estimator of 0. Then $E(T^2U) = E[T(TU)] = 0$ if $ET^4 < \infty$. By Theorem 3.2 in Shao (2003), T^2 is a UMVUE of ET^2. Similarly, we can show that T^3 is a UMVUE of ET^3,..., T^k is a UMVUE of ET^k. ∎

Exercise 14 (#3.27). Let X be a random variable having the Lebesgue density $[(1-\theta)+\theta/(2\sqrt{x})]I_{(0,1)}(x)$, where $\theta \in [0,1]$. Show that there is no UMVUE of θ based on an observation X.
Solution. Consider estimators of the form $h(X) = a(X^{-1/2}+b)I_{(c,1)}(X)$ for some real numbers a and b, and $c \in (0,1)$. Note that

$$\int_0^1 h(x)dx = a\int_c^1 x^{-1/2}dx + ab\int_c^1 dx = 2a(1-\sqrt{c}) + ab(1-c).$$

If $b = -2/(1+\sqrt{c})$, then $\int_0^1 h(x)dx = 0$ for any a and c. Also,

$$\int_0^1 \frac{h(x)}{2\sqrt{x}}dx = \frac{a}{2}\int_c^1 x^{-1}dx + \frac{ab}{2}\int_c^1 x^{-1/2}dx = -\frac{a}{2}\log c + ab(1-\sqrt{c}).$$

If $a = [b(1-\sqrt{c}) - 2^{-1}\log c]^{-1}$, then $\int_0^1 \frac{h(x)}{2\sqrt{x}}dx = 1$ for any b and c. Let $g_c = h$ with $b = -2/(1+\sqrt{c})$ and $a = [b(1-\sqrt{c}) - 2^{-1}\log c]^{-1}$, $c \in (0,1)$. Then

$$E[g_c(X)] = (1-\theta)\int_0^1 g_c(x)dx + \theta\int_0^1 \frac{g_c(x)}{2\sqrt{x}}dx = \theta$$

for any θ, i.e., $g_c(X)$ is unbiased for θ for any $c \in (0,1)$. The variance of $g_c(X)$ when $\theta = 0$ is

$$\begin{aligned} E[g_c(X)]^2 &= a^2\int_c^1 (x^{-1}+b^2+2bx^{-1/2})dx \\ &= a^2[-\log c + b^2(1-c) + 4b(1-\sqrt{c})] \\ &= \frac{-\log c + b^2(1-c) + 4b(1-\sqrt{c})}{[b(1-\sqrt{c}) - 2^{-1}\log c]^2}, \end{aligned}$$

where $b = -2/(1+\sqrt{c})$. Letting $c \to 0$, we obtain that $b \to -2$ and, thus, $E[g_c(X)]^2 \to 0$. This means that no minimum variance estimator within the class of estimators $g_c(X)$. Hence, there is no UMVUE of θ. ∎

Exercise 15 (#3.28). Let X be a random sample with $P(X=-1) = 2p(1-p)$ and $P(X=k) = p^k(1-p)^{3-k}$, $k = 0,1,2,3$, where $p \in (0,1)$.
(i) Determine whether there is a UMVUE of p.

(ii) Determine whether there is a UMVUE of $p(1-p)$.

Solution. (i) Suppose that $f(X)$ is an unbiased estimator of p. Then

$$p = 2f(-1)p(1-p) + f(0)(1-p)^3 + f(1)p(1-p)^2 + f(2)p^2(1-p) + f(3)p^3$$

for any p. Letting $p \to 0$, we obtain that $f(0) = 0$. Letting $p \to 1$, we obtain that $f(3) = 1$. Then

$$\begin{aligned} 1 &= 2f(-1)(1-p) + f(1)(1-p)^2 + f(2)p(1-p) + p^2 \\ &= 2f(-1) + f(1) + [f(2) - 2f(-1) - 2f(1)]p + [f(1) - f(2) + 1]p^2. \end{aligned}$$

Thus, $2f(-1)+f(1) = 1$, $f(2)-2f(-1)-2f(1) = 0$, and $f(1)-f(2)+1 = 0$. These three equations are not independent; in fact the second equation is a consequence of the first and the last equations. Let $f(2) = c$. Then $f(1) = c-1$ and $f(-1) = 1 - c/2$. Let $g_c(2) = c$, $g_c(1) = c-1$, $g_c(-1) = 1-c/2$, $g_c(0) = 0$, and $g_c(3) = 1$. Then the class of unbiased estimators of p is $\{g_c(X) : c \in \mathcal{R}\}$. The variance of $g_c(X)$ is

$$E[g_c(X)]^2 - p^2 = 2(1-c/2)^2p(1-p) + (c-1)^2p(1-p)^2 + c^2p^2(1-p) + p^3 - p^2.$$

Denote the right hand side of the above equation by $h(c)$. Then

$$h'(c) = -(2-c)p(1-p) + 2(c-1)p(1-p)^2 + 2cp^2(1-p).$$

Setting $h'(c) = 0$ we obtain that

$$0 = c - 2 + 2(c-1)(1-p) + 2cp = c - 2 + 2c - 2(1-p).$$

Hence, the function $h(c)$ reaches its minimum at $c = (4-2p)/3$, which depends on p. Therefore, there is no UMVUE of p.

(ii) Suppose that $f(X)$ is an unbiased estimator of $p(1-p)$. Then

$$\begin{aligned} p(1-p) = &\ 2f(-1)p(1-p) + f(0)(1-p)^3 + f(1)p(1-p)^2 \\ &+ f(2)p^2(1-p) + f(3)p^3 \end{aligned}$$

for any p. Letting $p \to 0$ we obtain that $f(0) = 0$. Letting $p \to 1$ we obtain that $f(3) = 0$. Then

$$1 = 2f(-1) + f(1)(1-p) + f(2)p$$

for any p, which implies that $f(2) = f(1)$ and $2f(-1) + f(1) = 1$. Let $f(-1) = c$. Then $f(1) = f(2) = 1-2c$. Let $g_c(-1) = c$, $g_c(0) = g_c(3) = 0$, and $g_c(1) = g_c(2) = 1-2c$. Then the class of unbiased estimators of $p(1-p)$ is $\{g_c(X) : c \in \mathcal{R}\}$. The variance of $g_c(X)$ is

$$\begin{aligned} E[g_c(X)]^2 - p^2 &= 2c^2p(1-p) + (1-2c)^2p(1-p)^2 \\ &\quad + (1-2c)^2p^2(1-p) - p^2 \\ &= 2c^2p(1-p) + (1-2c)^2p(1-p) - p^2 \\ &= [2c^2 + (1-2c)^2]p(1-p) - p^2, \end{aligned}$$

which reaches its minimum at $c = 1/3$ for any p. Thus, the UMVUE of $p(1-p)$ is $g_{1/3}(X)$. ∎

Exercise 16 (#3.29(a)). Let $(X_1, ..., X_n)$ be a random sample from the exponential distribution with density $\theta^{-1}e^{-(x-a)/\theta}I_{(a,\infty)}(x)$, where $a \leq 0$ and θ is known. Obtain a UMVUE of a.
Note. The minimum order statistic, $X_{(1)}$, is sufficient for a but not complete because $a \leq 0$.
Solution. Let $U(X_{(1)})$ be an unbiased estimator of 0. Then $E[U(X_{(1)})] = 0$ implies

$$\int_a^0 U(x)e^{-x/\theta}dx + \int_0^\infty U(x)e^{-x/\theta}dx = 0$$

for all $a \leq 0$. Hence, $U(x) = 0$ a.e. for $x \leq 0$ and $\int_0^\infty U(x)e^{-x/\theta}dx = 0$. Consider

$$h(X_{(1)}) = (bX_{(1)} + c)I_{(-\infty,0]}(X_{(1)})$$

with constants b and c. Then $E[h(X_{(1)})U(X_{(1)})] = 0$ for any a. By Theorem 3.2 in Shao (2003), $h(X_{(1)})$ is a UMVUE of its expectation

$$\begin{aligned} E[h(X_{(1)})] &= \frac{ne^{na/\theta}}{\theta}\int_a^0 (bx+c)e^{-nx/\theta}dx \\ &= c\left(1 - e^{na/\theta}\right) + ab + \frac{b\theta}{n}\left(1 - e^{na/\theta}\right), \end{aligned}$$

which equals a when $b = 1$ and $c = -\theta/n$. Therefore, the UMVUE of a is

$$h(X_{(1)}) = (X_{(1)} - \theta/n)I_{(-\infty,0]}(X_{(1)}). \quad ∎$$

Exercise 17 (#3.29(b)). Let $(X_1, ..., X_n)$ be a random sample from the distribution on $\mathcal{R}$ with Lebesgue density $\theta a^\theta x^{-(\theta+1)}I_{(a,\infty)}(x)$, where $a \in (0, 1]$ and θ is known. Obtain a UMVUE of a.
Solution. The minimum order statistic $X_{(1)}$ is sufficient for a and has Lebesgue density $n\theta a^{n\theta}x^{-(n\theta+1)}I_{(a,\infty)}(x)$. Let $U(X_{(1)})$ be an unbiased estimator of 0. Then $E[U(X_{(1)})] = 0$ implies

$$\int_a^1 U(x)x^{-(n\theta+1)}dx + \int_1^\infty U(x)x^{-(n\theta+1)}dx = 0$$

for all $a \in (0, 1]$. Hence, $U(x) = 0$ a.e. for $x \in (0, 1]$ and $\int_1^\infty U(x)x^{-(n\theta+1)}dx = 0$. Let

$$h(X_{(1)}) = cI_{(1,\infty)}(X_{(1)}) + bX_{(1)}I_{(0,1]}(X_{(1)})$$

with some constants b and c. Then

$$E[h(X_{(1)})U(X_{(1)})] = c\int_1^\infty U(x)x^{-(n\theta+1)}dx = 0.$$

By Theorem 3.2 in Shao (2003), $h(X_{(1)})$ is a UMVUE of its expectation

$$\begin{aligned}E[h(X_{(1)})] &= bn\theta a^{n\theta}\int_a^1 x^{-n\theta}dx + cn\theta a^{n\theta}\int_1^\infty x^{-(n\theta+1)}dx \\ &= \left(c - \frac{bn\theta}{n\theta-1}\right)a^{n\theta} + \frac{abn\theta}{n\theta-1},\end{aligned}$$

which equals a when $b = 1 - \frac{1}{n\theta}$ and $c = 1$. Hence, the UMVUE of a is

$$h(X_{(1)}) = I_{(1,\infty)}(X_{(1)}) + \left(1 - \frac{1}{n\theta}\right)X_{(1)}I_{(0,1]}(X_{(1)}). \ ∎$$

Exercise 18 (#3.30). Let $(X_1, ..., X_n)$ be a random sample from the population in a family $\mathcal{P}$ as described in Exercise 18 of Chapter 2. Find a UMVUE of θ.

Solution. Note that $\mathcal{P} = \mathcal{P}_1 \cup \mathcal{P}_2$, where $\mathcal{P}_1$ is the family of Poisson distributions with the mean parameter $\theta \in (0,1)$ and $\mathcal{P}_2$ is the family of binomial distributions with size 1 and probability θ. The sample mean $\bar{X}$ is the UMVUE of θ when either $\mathcal{P}_1$ or $\mathcal{P}_2$ is considered as the family of distributions. Hence $\bar{X}$ is the UMVUE of θ when $\mathcal{P}$ is considered as the family of distributions. ∎

Exercise 19 (#3.33). Find a function of θ for which the amount of information is independent of θ, when P_θ is
(i) the Poisson distribution with unknown mean $\theta > 0$;
(b) the binomial distribution with known size r and unknown probability $\theta \in (0,1)$;
(c) the gamma distribution with known shape parameter α and unknown scale parameter $\theta > 0$.

Solution. (i) The Fisher information about θ is $I(\theta) = \frac{1}{\theta}$. Let $\eta = \eta(\theta)$. If the Fisher information about η is

$$\tilde{I}(\eta) = \left(\frac{d\theta}{d\eta}\right)^2 I(\theta) = \left(\frac{d\theta}{d\eta}\right)^2 \frac{1}{\theta} = c$$

not depending on θ, then $\frac{d\eta}{d\theta} = 1/\sqrt{c\theta}$. Hence, $\eta(\theta) = 2\sqrt{\theta}/\sqrt{c}$.
(ii) The Fisher information about θ is $I(\theta) = \frac{r}{\theta(1-\theta)}$. Let $\eta = \eta(\theta)$. If the Fisher information about η is

$$\tilde{I}(\eta) = \left(\frac{d\theta}{d\eta}\right)^2 I(\theta) = \left(\frac{d\theta}{d\eta}\right)^2 \frac{r}{\theta(1-\theta)} = c$$

not depending on θ, then $\frac{d\eta}{d\theta} = \sqrt{r}/\sqrt{c\theta(1-\theta)}$. Choose $c = 4r$. Then $\eta(\theta) = \arcsin(\sqrt{\theta})$.

(iii) The Fisher information about θ is $I(\theta) = \frac{\alpha}{\theta^2}$. Let $\eta = \eta(\theta)$. If the Fisher information about η is

$$\tilde{I}(\eta) = \left(\frac{d\theta}{d\eta}\right)^2 I(\theta) = \left(\frac{d\theta}{d\eta}\right)^2 \frac{\alpha}{\theta^2} = \alpha,$$

then $\frac{d\eta}{d\theta} = \theta^{-1}$ and, hence, $\eta(\theta) = \log\theta$. ∎

Exercise 20 (#3.34). Let $(X_1, ..., X_n)$ be a random sample from a distribution on $\mathcal{R}$ with the Lebesgue density $\frac{1}{\sigma} f\left(\frac{x-\mu}{\sigma}\right)$, where $f(x) > 0$ is a known Lebesgue density and $f'(x)$ exists for all $x \in \mathcal{R}$, $\mu \in \mathcal{R}$, and $\sigma > 0$. Let $\theta = (\mu, \sigma)$. Show that the Fisher information about θ contained in $X_1, ..., X_n$ is

$$I(\theta) = \frac{n}{\sigma^2} \begin{pmatrix} \int \frac{[f'(x)]^2}{f(x)} dx & \int \frac{f'(x)[xf'(x)+f(x)]}{f(x)} dx \\ \int \frac{f'(x)[xf'(x)+f(x)]}{f(x)} dx & \int \frac{[xf'(x)+f(x)]^2}{f(x)} dx \end{pmatrix},$$

assuming that all integrals are finite.

Solution. Let $g(\mu, \sigma, x) = \log \frac{1}{\sigma} f\left(\frac{x-\mu}{\sigma}\right)$. Then

$$\frac{\partial}{\partial \mu} g(\mu, \sigma, x) = -\frac{f'\left(\frac{x-\mu}{\sigma}\right)}{\sigma f\left(\frac{x-\mu}{\sigma}\right)}$$

and

$$\frac{\partial}{\partial \sigma} g(\mu, \sigma, x) = -\frac{(x-\mu) f'\left(\frac{x-\mu}{\sigma}\right)}{\sigma f\left(\frac{x-\mu}{\sigma}\right)} - \frac{1}{\sigma}.$$

Then

$$\begin{aligned} E\left[\frac{\partial}{\partial \mu} g(\mu, \sigma, X_1)\right]^2 &= \frac{1}{\sigma^2} \int \left[\frac{f'\left(\frac{x-\mu}{\sigma}\right)}{f\left(\frac{x-\mu}{\sigma}\right)}\right]^2 \frac{1}{\sigma} f\left(\frac{x-\mu}{\sigma}\right) dx \\ &= \frac{1}{\sigma^2} \int \frac{\left[f'\left(\frac{x-\mu}{\sigma}\right)\right]^2}{f\left(\frac{x-\mu}{\sigma}\right)} d\left(\frac{x}{\sigma}\right) \\ &= \frac{1}{\sigma^2} \int \frac{[f'(x)]^2}{f(x)} dx, \end{aligned}$$

$$\begin{aligned} E\left[\frac{\partial}{\partial \sigma} g(\mu, \sigma, X_1)\right]^2 &= \frac{1}{\sigma^2} \int \left[\frac{x-\mu}{\sigma} \frac{f'\left(\frac{x-\mu}{\sigma}\right)}{f\left(\frac{x-\mu}{\sigma}\right)} + 1\right]^2 \frac{1}{\sigma} f\left(\frac{x-\mu}{\sigma}\right) dx \\ &= \frac{1}{\sigma^2} \int \left[x \frac{f'(x)}{f(x)} + 1\right]^2 f(x) dx \\ &= \frac{1}{\sigma^2} \int \frac{[xf'(x) + f(x)]^2}{f(x)} dx, \end{aligned}$$

and

$$
\begin{aligned}
& E\left[\frac{\partial}{\partial\mu} g(\mu,\sigma,X_1)\frac{\partial}{\partial\sigma} g(\mu,\sigma,X_1)\right] \\
&= \frac{1}{\sigma^2}\int \frac{f'\left(\frac{x-\mu}{\sigma}\right)}{f\left(\frac{x-\mu}{\sigma}\right)}\left[\frac{x-\mu}{\sigma}\frac{f'\left(\frac{x-\mu}{\sigma}\right)}{f\left(\frac{x-\mu}{\sigma}\right)}+1\right]\frac{1}{\sigma} f\left(\frac{x-\mu}{\sigma}\right) dx \\
&= \int \frac{f'(x)[xf'(x)+f(x)]}{f(x)} dx.
\end{aligned}
$$

The result follows since

$$
I(\theta) = nE\left[\frac{\partial}{\partial\theta}\log \tfrac{1}{\sigma} f\left(\tfrac{X_1-\mu}{\sigma}\right)\right]\left[\frac{\partial}{\partial\theta}\log \tfrac{1}{\sigma} f\left(\tfrac{X_1-\mu}{\sigma}\right)\right]^\tau . \quad ∎
$$

Exercise 21 (#3.36). Let X be a sample having a probability density $f_\theta(x)$ with respect to ν, where θ is a k-vector of unknown parameters. Let $T(X)$ be a statistic having a probability density $g_\theta(t)$ with respect to λ. Suppose that $\frac{\partial}{\partial\theta} f_\theta(x)$ and $\frac{\partial}{\partial\theta} g_\theta(t)$ exist for any x and t and that, on any set $\{\|\theta\| \leq c\}$, there are functions $u_c(x)$ and $v_c(t)$ such that $|\frac{\partial}{\partial\theta} f_\theta(x)| \leq u_c(x)$, $|\frac{\partial}{\partial\theta} g_\theta(t)| \leq v_c(t)$, $\int u_c(x)d\nu < \infty$, and $\int v_c(t)d\lambda < \infty$. Show that
(i) $I_X(\theta) - I_T(\theta)$ is nonnegative definite, where $I_X(\theta)$ is the Fisher information about θ contained in X and $I_T(\theta)$ is the Fisher information about θ contained in T;
(ii) $I_X(\theta) = I_T(\theta)$ if T is sufficient for θ.
Solution. (i) For any event $T^{-1}(B)$,

$$
\begin{aligned}
\int_{T^{-1}(B)} \frac{\partial}{\partial\theta}\log f_\theta(X)dP &= \int_{T^{-1}(B)} \frac{\partial}{\partial\theta} f_\theta(x)d\nu \\
&= \frac{\partial}{\partial\theta}\int_{T^{-1}(B)} f_\theta(x)d\nu \\
&= \frac{\partial}{\partial\theta} P\big(T^{-1}(B)\big) \\
&= \frac{\partial}{\partial\theta}\int_B g_\theta(t)d\lambda \\
&= \int_B \frac{\partial}{\partial\theta} g_\theta(t)d\lambda \\
&= \int_B \left[\frac{\partial}{\partial\theta}\log g_\theta(t)\right] g_\theta(t)d\lambda \\
&= \int_{T^{-1}(B)} \frac{\partial}{\partial\theta}\log g_\theta(T)dP,
\end{aligned}
$$

where the exchange of differentiation and integration is justified by the dominated convergence theorem under the given conditions. This shows that

$$E\left[\frac{\partial}{\partial\theta}\log f_\theta(X)\middle|T\right] = \frac{\partial}{\partial\theta}\log g_\theta(T) \quad \text{a.s.}$$

Then

$$\begin{aligned}
&E\left[\frac{\partial}{\partial\theta}\log f_\theta(X)\right]\left[\frac{\partial}{\partial\theta}\log g_\theta(T)\right]^\tau \\
&= E\left\{E\left[\frac{\partial}{\partial\theta}\log f_\theta(X)\middle|T\right]\left[\frac{\partial}{\partial\theta}\log g_\theta(T)\right]^\tau\right\} \\
&= E\left[\frac{\partial}{\partial\theta}\log g_\theta(T)\right]\left[\frac{\partial}{\partial\theta}\log g_\theta(T)\right]^\tau \\
&= I_T(\theta).
\end{aligned}$$

Then the nonnegative definite matrix

$$E\left[\frac{\partial}{\partial\theta}\log f_\theta(X) - \frac{\partial}{\partial\theta}\log g_\theta(T)\right]\left[\frac{\partial}{\partial\theta}\log f_\theta(X) - \frac{\partial}{\partial\theta}\log g_\theta(T)\right]^\tau$$

is equal to $I_X(\theta) + I_T(\theta) - 2I_T(\theta) = I_X(\theta) - I_T(\theta)$. Hence $I_X(\theta) - I_T(\theta)$ is nonnegative definite.

(ii) If T is sufficient, then by the factorization theorem, $f_\theta(x) = \tilde{g}_\theta(t)h(x)$. Since $\frac{\partial}{\partial\theta}\log f_\theta(x) = \frac{\partial}{\partial\theta}\log\tilde{g}_\theta(t)$, the result in part (i) of the solution implies that

$$\frac{\partial}{\partial\theta}\log\tilde{g}_\theta(T) = \frac{\partial}{\partial\theta}\log g_\theta(T) \quad \text{a.s.}$$

Therefore, $I_X(\theta) = I_T(\theta)$. ∎

Exercise 22 (#3.37). Let $(X_1, ..., X_n)$ be a random sample from the uniform distribution on the interval $(0, \theta)$ with $\theta > 0$.

(i) Show that $\frac{d}{d\theta}\int x f_\theta(x)dx \neq \int x \frac{d}{d\theta} f_\theta(x)dx$, where f_θ is the density of $X_{(n)}$, the largest order statistic.

(ii) Show that the Fisher information inequality does not hold for the UMVUE of θ.

Solution. (i) Note that $f_\theta(x) = n\theta^{-n}x^{n-1}I_{(0,\theta)}(x)$. Then

$$\int x\frac{d}{d\theta}f_\theta(x)dx = -\frac{n^2}{\theta^{n+1}}\int_0^\theta x^n dx = -\frac{n^2}{n+1}.$$

On the other hand,

$$\frac{d}{d\theta}\int x f_\theta(x)dx = \frac{d}{d\theta}\left(\frac{n}{\theta^n}\int_0^\theta x^n dx\right) = \frac{d}{d\theta}\left(\frac{n\theta}{n+1}\right) = \frac{n}{n+1}.$$

(ii) The UMVUE of θ is $(n+1)X_{(n)}/n$ with variance $\theta^2/[n(n+2)]$. On the other hand, the Fisher information is $I(\theta) = n\theta^{-2}$. Hence $[I(\theta)]^{-1} = \theta^2/n > \theta^2/[n(n+2)]$. ∎

Exercise 23 (#3.39). Let X be an observation with Lebesgue density $(2\theta)^{-1}e^{-|x|/\theta}$ with unknown $\theta > 0$. Find the UMVUE's of the parameters θ, θ^r $(r > 1)$, and $(1+\theta)^{-1}$ and, in each case, determine whether the variance of the UMVUE attains the Cramér-Rao lower bound.

Solution. For θ, Cramér-Rao lower bound is θ^2 and $|X|$ is the UMVUE of θ with $\text{Var}(|X|) = \theta^2$, which attains the lower bound.

For θ^r, Cramér-Rao lower bound is $r^2\theta^{2r}$. Since $E[|X|^r/\Gamma(r+1)] = \theta^r$, $|X|^r/\Gamma(r+1)$ is the UMVUE of θ^r with

$$\text{Var}\left(\frac{|X|^r}{\Gamma(r+1)}\right) = \theta^{2r}\left[\frac{\Gamma(2r+1)}{\Gamma(r+1)\Gamma(r+1)} - 1\right] > r^2\theta^{2r}$$

when $r > 1$.

For $(1+\theta)^{-1}$. Cramér-Rao lower bound is $\theta^2/(1+\theta)^4$. Since $E\left(e^{-|X|}\right) = (1+\theta)^{-1}$, $e^{-|X|}$ is the UMVUE of $(1+\theta)^{-1}$ with

$$\text{Var}\left(e^{-|X|}\right) = \frac{1}{1+2\theta} - \frac{1}{(1+\theta)^2} > \frac{\theta^2}{(1+\theta)^4}. \quad \blacksquare$$

Exercise 24 (#3.42). Let $(X_1, ..., X_n)$ be a random sample from $N(\mu, \sigma^2)$ with an unknown $\mu \in \mathcal{R}$ and a known $\sigma^2 > 0$. Find the UMVUE of $e^{t\mu}$ with a fixed $t \neq 0$ and show that the variance of the UMVUE is larger than the Cramér-Rao lower bound but the ratio of the variance of the UMVUE over the Cramér-Rao lower bound converges to 1 as $n \to \infty$.

Solution. The sample mean $\bar{X}$ is complete and sufficient for μ. Since

$$E\left(e^{t\bar{X}}\right) = e^{\mu t + \sigma^2 t^2/(2n)},$$

the UMVUE of $e^{t\mu}$ is $T(X) = e^{-\sigma^2 t^2/(2n) + t\bar{X}}$.

The Fisher information $I(\mu) = n/\sigma^2$. Then the Cramér-Rao lower bound is $\left(\frac{d}{d\mu}e^{t\mu}\right)^2 / I(\mu) = \sigma^2 t^2 e^{2t\mu}/n$. On the other hand,

$$\text{Var}(T) = e^{-\sigma^2 t^2/n} E e^{2t\bar{X}} - e^{2t\mu} = \left(e^{\sigma^2 t^2/n} - 1\right) e^{2t\mu} > \frac{\sigma^2 t^2 e^{2t\mu}}{n},$$

the Cramér-Rao lower bound. The ratio of the variance of the UMVUE over the Cramér-Rao lower bound is $(e^{\sigma^2 t^2/n} - 1)/(\sigma^2 t^2/n)$, which converges to 1 as $n \to \infty$, since $\lim_{x\to 0}(e^x - 1)/x = 1$. $\blacksquare$

Exercise 25 (#3.46, #3.47). Let $X_1, X_2, ...$ be independent and identically distributed random variables, m be a positive integer, and $h(x_1,..., x_m)$ be a function on $\mathcal{R}^m$ such that $E[h(X_1, ..., X_m)]^2 < \infty$ and h is symmetric in its m arguments. A U-statistic with kernel h (of order m) is defined as

$$U_n = \binom{n}{m}^{-1} \sum_{1 \le i_1 < \cdots < i_m \le n} h(X_{i_1}, ..., X_{i_m}),$$

where $\sum_{1\leq i_1<\cdots<i_m\leq n}$ denotes the summation over the $\binom{n}{m}$ combinations of m distinct elements $\{i_1,...,i_m\}$ from $\{1,...,n\}$. For $k=1,...,m$, define $h_k(x_1,...,x_k)=E[h(x_1,...,x_k,X_{k+1},...,X_m)]$ and $\zeta_k=\mathrm{Var}(h_k(X_1,...,X_k))$. Show that
(i) $\zeta_1\leq\zeta_2\leq\cdots\leq\zeta_m$;
(ii) $(n+1)\mathrm{Var}(U_{n+1})\leq n\mathrm{Var}(U_n)$ for any $n\geq m$;
(iii) if $\zeta_j=0$ for $j<k$ and $\zeta_k>0$, where $1\leq k\leq m$, then

$$\mathrm{Var}(U_n)=\frac{k!\binom{m}{k}^2\zeta_k}{n^k}+O\left(\frac{1}{n^{k+1}}\right);$$

(iv) $m^2\zeta_1\leq n\mathrm{Var}(U_n)\leq m\zeta_m$ for any $n\geq m$.

Solution. (i) For any $k=1,...,m-1$, let $W=h_{k+1}(X_1,...,X_k,X_{k+1})$ and $Y=(X_1,...,X_k)$. Then $\zeta_{k+1}=\mathrm{Var}(W)$ and $\zeta_k=\mathrm{Var}(E(W|Y))$, since

$$E(W|Y)=E[h_{k+1}(X_1,...,X_k,X_{k+1})|X_1,...,X_k)]=h_k(X_1,...,X_k).$$

The result follows from

$$\mathrm{Var}(W)=E\{E[(W-EW)^2|Y]\}\geq E\{[E(W|Y)-EW]^2\}=\mathrm{Var}(E(W|Y)),$$

where the inequality follows from Jensen's inequality for conditional expectations and the equality follows from $EW=E[E(W|Y)]$.

(ii) We use induction. The result is obvious when $m=1$, since U is an average of independent and identically distributed random variables when $m=1$. Assume that the result holds for any U-statistic with a kernel of order $m-1$. From Hoeffding's representation (e.g., Serfling, 1980, p. 178),

$$U_n-EU_n=W_n+S_n,$$

where W_n is a U-statistic with a kernel of order $m-1$, S_n is a U-statistic with variance $\binom{n}{m}^{-1}\eta_m$, η_m is a constant not depending on n, and $\mathrm{Var}(U_n)=\mathrm{Var}(W_n)+\mathrm{Var}(S_n)$. By the induction assumption, $(n+1)\mathrm{Var}(W_{n+1})\leq n\mathrm{Var}(W_n)$. Then, for any $n\geq m$,

$$\begin{aligned}
n\mathrm{Var}(U_n)&=n\mathrm{Var}(W_n)+n\mathrm{Var}(S_n)\\
&=n\mathrm{Var}(W_n)+n\binom{n}{m}^{-1}\eta_m\\
&=n\mathrm{Var}(W_n)+\frac{m!\eta_m}{(n-1)(n-2)\cdots(n-m+1)}\\
&\geq(n+1)\mathrm{Var}(W_{n+1})+\frac{m!\eta_m}{n(n-1)\cdots(n-m+2)}\\
&=(n+1)\mathrm{Var}(W_{n+1})+(n+1)\binom{n+1}{m}^{-1}\eta_m\\
&=(n+1)\mathrm{Var}(W_{n+1})+(n+1)\mathrm{Var}(S_{n+1})\\
&=(n+1)\mathrm{Var}(U_{n+1}).
\end{aligned}$$

(iii) From Hoeffding's theorem (e.g., Theorem 3.4 in Shao, 2003),

$$\mathrm{Var}(U_n) = \sum_{l=1}^{m} \frac{\binom{m}{l}\binom{n-m}{m-l}}{\binom{n}{m}} \zeta_l.$$

For any $l = 1, ..., m$,

$$\begin{aligned}
\frac{\binom{m}{l}\binom{n-m}{m-l}}{\binom{n}{m}} &= l!\binom{m}{l}^2 \frac{(n-m)(n-m-1)\cdots[n-m-(m-l-1)]}{n(n-1)\cdots[n-(m-1)]} \\
&= l!\binom{m}{l}^2 \left[\frac{1}{n^l} + O\left(\frac{1}{n^{l+1}}\right)\right] \\
&= O\left(\frac{1}{n^l}\right).
\end{aligned}$$

If $\zeta_j = 0$ for $j < k$ and $\zeta_k > 0$, where $1 \leq k \leq m$, then

$$\begin{aligned}
\mathrm{Var}(U_n) &= \sum_{l=k}^{m} \frac{\binom{m}{l}\binom{n-m}{m-l}}{\binom{n}{m}} \zeta_l \\
&= \frac{\binom{m}{k}\binom{n-m}{m-k}}{\binom{n}{m}} \zeta_k + \sum_{l=k+1}^{m} \frac{\binom{m}{l}\binom{n-m}{m-l}}{\binom{n}{m}} \zeta_l \\
&= \frac{k!\binom{m}{k}^2 \zeta_k}{n^k} + O\left(\frac{1}{n^{k+1}}\right) + \sum_{l=k+1}^{m} O\left(\frac{1}{n^l}\right) \\
&= \frac{k!\binom{m}{k}^2 \zeta_k}{n^k} + O\left(\frac{1}{n^{k+1}}\right).
\end{aligned}$$

(iv) From the result in (ii), $n\mathrm{Var}(U_n)$ is nonincreasing in n. Hence $n\mathrm{Var}(U_n) \leq m\mathrm{Var}(U_m) = m\zeta_m$ for any $n \geq m$. Also, $\lim_n[n\mathrm{Var}(U_n)] \leq n\mathrm{Var}(U_n)$ for any $n \geq m$. If $\zeta_1 > 0$, from the result in (iii), $\lim_n[n\mathrm{Var}(U_n)] = m^2\zeta_1$. Hence, $m^2\zeta_1 \leq n\mathrm{Var}(U_n)$ for any $n \geq m$, which obviously also holds if $\zeta_1 = 0$. ∎

Exercise 26 (#3.53). Let $h(x_1, x_2, x_3) = I_{(-\infty,0)}(x_1 + x_2 + x_3)$. Find h_k and ζ_k, $k = 1, 2, 3$, for the U-statistic with kernel h based on independent random variables $X_1, X_2, ...$ with a common cumulative distribution function F.

Solution. Let $G * H$ denote the convolution of the two cumulative distribution functions G and H. Then

$$h_1(x_1) = E[I_{(-\infty,0)}(x_1 + X_2 + X_3)] = F * F(-x_1),$$

$$h_2(x_1, x_2) = E[I_{(-\infty,0)}(x_1 + x_2 + X_3)] = F(-x_1 - x_2),$$

$$h_3(x_1, x_2, x_3) = I_{(-\infty,0)}(x_1 + x_2 + x_3),$$
$$\zeta_1 = \text{Var}(F * F(-X_1)),$$
$$\zeta_2 = \text{Var}(F(-X_1 - X_2)),$$

and

$$\zeta_3 = F * F * F(0)[1 - F * F * F(0)]. \ \blacksquare$$

Exercise 27 (#3.54). Let $X_1, ..., X_n$ be a random sample of random variables having finite EX_1^2 and EX_1^{-2}. Let $\mu = EX_1$ and $\bar{\mu} = EX_1^{-1}$. Find a U-statistic that is an unbiased estimator of $\mu\bar{\mu}$ and derive its variance and asymptotic distribution.
Solution. Consider $h(x_1, x_2) = (\frac{x_1}{x_2} + \frac{x_2}{x_1})/2$. Then the U-statistic

$$U_n = \frac{1}{n(n-1)} \sum_{1 \le i < j \le n} \left(\frac{X_i}{X_j} + \frac{X_j}{X_i} \right)$$

is unbiased for $E[h(X_1, X_2)] = \mu\bar{\mu}$. Define $h_1(x) = (x\bar{\mu} + x^{-1}\mu)/2$. Then

$$\zeta_1 = \text{Var}(h(X_1)) = \frac{\bar{\mu}^2 V(X_1) + \mu^2 \text{Var}(X_1^{-1}) + 2\mu\bar{\mu}(1 - \mu\bar{\mu})}{4}.$$

By Theorem 3.5 in Shao (2003),

$$\sqrt{n}(U_n - \mu\bar{\mu}) \to_d N(0, 4\zeta_1).$$

Using the formula for the variance of U-statistics given in the solution of the previous exercise, we obtain the variance of U_n as $[4(n-2)\zeta_1 + 2\zeta_2]/[n(n-1)]$, where $\zeta_2 = \text{Var}(h(X_1, X_2))$. ∎

Exercise 28 (#3.58). Suppose that

$$X_{ij} = \alpha_i + \theta t_{ij} + \varepsilon_{ij}, \quad i = 1, ..., a, \ j = 1, ..., b,$$

where α_i and θ are unknown parameters, t_{ij} are known constants, and ε_{ij} are independent and identically distributed random variables with mean 0. Find explicit forms for the least squares estimators (LSE's) of θ, α_i, $i = 1, ..., a$.
Solution. Write the model in the form of $X = Z\beta + \varepsilon$, where

$$X = (X_{11}, ..., X_{1b}, ..., X_{a1}, ..., X_{ab}),$$

$$\beta = (\alpha_1, ..., \alpha_a, \theta),$$

and

$$\varepsilon = (\varepsilon_{11}, ..., \varepsilon_{1b}, ..., \varepsilon_{a1}, ..., \varepsilon_{ab}).$$

Then the design matrix Z is

$$Z = \begin{pmatrix} J_b & 0 & 0 & t_1 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & J_b & t_a \end{pmatrix},$$

where $t_i = (t_{i1}, ..., t_{ib})$ and J_b is the b-vector of 1's. Solving the normal equation $(Z^t auZ)\hat{\beta} = Z^\tau X$, we obtain the LSE's

$$\hat{\theta} = \frac{\sum_{i=1}^a \sum_{j=1}^b t_{ij} X_{ij} - b \sum_i \bar{t}_{i\cdot} \bar{X}_{i\cdot}}{\sum_{j=1}^b (t_{ij} - \bar{t}_{i\cdot})^2},$$

where $\bar{t}_{i\cdot} = \frac{1}{b}\sum_{j=1}^b t_{ij}$, $\bar{X}_{i\cdot} = \frac{1}{b}\sum_{j=1}^b X_{ij}$, and

$$\hat{\alpha}_i = \bar{X}_{i\cdot} - \hat{\theta}\bar{t}_{i\cdot}, \quad i = 1, ..., a. \blacksquare$$

Exercise 29 (#3.59). Consider the polynomial model

$$X_i = \beta_0 + \beta_1 t_i + \beta_2 t_i^2 + \beta_3 t_i^3 + \varepsilon_i, \quad i = 1, ..., n,$$

where ε_i's are independent and identically distributed random variables with mean 0. Suppose that $n = 12$, $t_i = -1$, $i = 1, ..., 4$, $t_i = 0$, $i = 5, ..., 8$, and $t_i = 1$, $i = 9, ..., 12$. Show whether the following parameters are estimable (i.e., they can be unbiasedly estimated): $\beta_0 + \beta_2$, β_1, $\beta_0 - \beta_1$, $\beta_1 + \beta_3$, and $\beta_0 + \beta_1 + \beta_2 + \beta_3$.

Solution. Let $X = (X_1, ..., X_{12})$, $\varepsilon = (\varepsilon_1, ..., \varepsilon_{12})$, and $\beta = (\beta_0, \beta_1, \beta_2, \beta_3)$. Then $X = Z\beta + \varepsilon$ with

$$Z^\tau = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ -1 & -1 & -1 & -1 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 \\ -1 & -1 & -1 & -1 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 \end{pmatrix}$$

and

$$Z^\tau Z = \begin{pmatrix} 12 & 0 & 8 & 0 \\ 0 & 8 & 0 & 8 \\ 8 & 0 & 8 & 0 \\ 0 & 8 & 0 & 8 \end{pmatrix}.$$

From the theory of linear models (e.g., Theorem 3.6 in Shao, 2003), a parameter $l^\tau \beta$ with a known vector l is estimable if and only if $l \in \mathcal{R}(Z^\tau Z)$. Note that $\beta_0 + \beta_2 = l^\tau \beta$ with $l = (1, 0, 1, 0)$, which is the third row of $Z^\tau Z$ divided by 8. Hence $\beta_0 + \beta_2$ is estimable. Similarly, $\beta_1 + \beta_3 = l^\tau \beta$

with $l = (0,1,0,1)$, which is the second row of $Z^\tau Z$ divided by 8 and, hence, $\beta_1 + \beta_3$ is estimable. Then $\beta_0 + \beta_1 + \beta_2 + \beta_3$ is estimable, since any linear combination of estimable functions is estimable. We now show that $\beta_0 - \beta_1 = l^\tau \beta$ with $l = (1,-1,0,0)$ is not estimable. If $\beta_0 - \beta_1$ is estimable, then there is $c = (c_1, ..., c_4)$ such that $l = Z^\tau Z c$, i.e.,

$$\begin{aligned} 12c_1 + 8c_3 &= 1 \\ 8c_2 + 8c_4 &= -1 \\ 8c_1 + 8c_3 &= 0 \\ 8c_2 + 8c_4 &= 0, \end{aligned}$$

where the second and the last equations have no solution. Similarly, the parameter β_1 is not estimable, since $8c_2 + 8c_4 = 1$ and $8c_2 + 8c_4 = 0$ can not hold at the same time. ∎

Exercise 30 (#3.60). Consider the one-way ANOVA model

$$X_{ij} = \mu + \alpha_i + \varepsilon_{ij}, \quad j = 1, ..., n_i, i = 1, ..., m,$$

where μ and α_i are unknown parameters and ε_{ij} are independent and identically distributed random variables with mean 0. Let

$$X = (X_{11}, ..., X_{1n_1}, ..., X_{m1}, ..., X_{mn_m}),$$

$$\varepsilon = (\varepsilon_{11}, ..., \varepsilon_{1n_1}, ..., \varepsilon_{m1}, ..., \varepsilon_{mn_m}),$$

and $\beta = (\mu, \alpha_1, ..., \alpha_m)$. Find the matrix Z in the linear model $X = Z\beta + \varepsilon$, the matrix $Z^\tau Z$, and the form of l for estimable $l^\tau \beta$.

Solution. Let $n = n_1 + \cdots + n_m$ and J_a be the a-vector of 1's. Then

$$Z = \begin{pmatrix} J_{n_1} & J_{n_1} & 0 & \cdots & 0 \\ J_{n_2} & 0 & J_{n_2} & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ J_{n_m} & 0 & 0 & \cdots & J_{n_m} \end{pmatrix}$$

and

$$Z^\tau Z = \begin{pmatrix} n & n_1 & n_2 & \cdots & n_m \\ n_1 & n_1 & 0 & \cdots & 0 \\ n_2 & 0 & n_2 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ n_m & 0 & 0 & \cdots & n_m \end{pmatrix}.$$

Note that $l^\tau \beta$ is estimable if and only if $l \in \mathcal{R}(Z^\tau Z)$, the linear space generated by the rows of $Z^\tau Z$. We now show that $l^\tau \beta$ is estimable if and only if $l_0 = l_1 + \cdots + l_m$ for $l = (l_0, l_1, ..., l_m) \in \mathcal{R}^{m+1}$.

If $l \in \mathcal{R}(Z^\tau Z)$, then there is a $c = (c_0, c_1, ..., c_m) \in \mathcal{R}^{m+1}$ such that $l = Z^\tau Zc$, i.e.,

$$
\begin{aligned}
&nc_0 + n_1c_1 + \cdots + n_mc_m = l_0 \\
&n_1c_0 + n_1c_1 = l_1 \\
&\cdots\cdots\cdots\cdots \\
&n_mc_0 + n_mc_m = l_m
\end{aligned}
$$

holds. Then $l_0 = l_1 + \cdots + l_m$. On the other hand, if $l_0 = l_1 + \cdots + l_m$, then the previous $m+1$ equations with $c_0, c_1, ..., c_m$ considered as variables have infinitely many solutions. Hence $l \in \mathcal{R}(Z^\tau Z)$. ∎

Exercise 31 (#3.61). Consider the two-way balanced ANOVA model

$$X_{ijk} = \mu + \alpha_i + \beta_j + \gamma_{ij} + \varepsilon_{ijk}, \quad i = 1, ..., a, j = 1, ..., b, k = 1, ..., c,$$

where a, b, and c are some positive integers, ε_{ijk}'s are independent and identically distributed random variables with mean 0, and μ, α_i's, β_j's, and γ_{ij}'s are unknown parameters. Let X be the vector of X_{ijk}'s, ε be the vector of ε_{ijk}'s, and $\beta = (\mu, \alpha_1, ..., \alpha_a, \beta_1, ..., \beta_b, \gamma_{11}, ..., \gamma_{1b}, ..., \gamma_{a1}, ..., \gamma_{ab})$.
(i) Obtain the design matrix Z in the model $X = Z\beta + \varepsilon$ and show that the rank of Z is ab.
(ii) Find the form of estimable $l^\tau\beta$, $l \in \mathcal{R}^{1+a+b+ab}$.
(iii) Obtain an LSE of β.
Solution. (i) Let J_t be the t-vector of 1's, I_t be the identity matrix of order t, A be the $ab \times b$ block diagonal matrix whose jth diagonal block is J_a, $j = 1, ..., b$,

$$B = (I_b \; I_b \cdots I_b),$$

and

$$\Lambda = (J_{ab} \; A \; B^\tau \; I_{ab}),$$

which is an $ab \times (1+a+b+ab)$ matrix. Then Z is the $(1+a+b+ab) \times abc$ matrix whose transpose is

$$Z^\tau = (\Lambda^\tau \; \Lambda^\tau \cdots \Lambda^\tau)$$

and

$$Z^\tau Z = c\Lambda^\tau\Lambda = c \begin{pmatrix} \Lambda_0^\tau\Lambda_0 & \Lambda_0^\tau \\ \Lambda_0 & I_{ab} \end{pmatrix},$$

where $\Lambda_0 = (J_{ab} \; A \; B^\tau)$. Clearly, the last ab rows of $Z^\tau Z$ are linearly independent. Hence the rank of Z, which is the same as the rank of $Z^\tau Z$, is no smaller than ab. On the other hand, the rank of Λ is no larger than ab and, hence, the rank of $Z^\tau Z$ is no larger than ab. Thus, the rank of Z is ab.
(ii) A function $l^\tau\beta$ with $l \in \mathcal{R}^{1+a+b+ab}$ is estimable if and only if l is a linear combination of the rows of $Z^\tau Z$. From the discussion in part (i)

of the solution, we know that $l^\tau\beta$ is estimable if and only if l is a linear combination of the rows in the matrix $(\Lambda_0^\tau\ I_{ab})$.
(iii) Any solution of $Z^\tau Z\hat\beta = Z^\tau X$ is an LSE of β. A direct calculation shows that an LSE of β is $(\hat\mu, \hat\alpha_1, ..., \hat\alpha_a, \hat\beta_1, ..., \hat\beta_b, \hat\gamma_{11}, ..., \hat\gamma_{1b}, ..., \hat\gamma_{a1}, ..., \hat\gamma_{ab})$, where $\hat\mu = \bar X_{\cdots}$, $\hat\alpha_i = \bar X_{i\cdot\cdot} - \bar X_{\cdots}$, $\hat\beta_j = \bar X_{\cdot j\cdot} - \bar X_{\cdots}$, $\hat\gamma_{ij} = \bar X_{ij\cdot} - \bar X_{i\cdot\cdot} - \bar X_{\cdot j\cdot} + \bar X_{\cdots}$, and a dot is used to denote averaging over the indicated subscript. ∎

Exercise 32 (#3.63). Assume that X is a random n-vector from the multivariate normal distribution $N_n(Z\beta, \sigma^2 I_n)$, where Z is an $n\times p$ known matrix of rank $r \le p < n$, β is a p-vector of unknown parameters, I_n is the identity matrix of order n, and $\sigma^2 > 0$ is unknown. Find the UMVUE's of $(l^\tau\beta)^2$, $l^\tau\beta/\sigma$, and $(l^\tau\beta/\sigma)^2$ for an estimable $l^\tau\beta$.
Solution. Let $\hat\beta$ be the LSE of β and $\hat\sigma^2 = \|X - Z\hat\beta\|^2/(n-r)$. Note that $(Z^\tau X, \hat\sigma^2)$ is complete and sufficient for (β, σ^2), $l^\tau\hat\beta$ has the normal distribution $N(l^\tau\beta, \sigma^2 l^\tau(Z^\tau Z)^- l)$, and $(n-r)\hat\sigma^2/\sigma^2$ has the chi-square distribution χ^2_{n-r}, where A^- is a generalized inverse of A. Since $E(l^\tau\hat\beta)^2 = [E(l^\tau\hat\beta)]^2 + \mathrm{Var}(l^\tau\hat\beta) = (l^\tau\beta)^2 + \sigma^2 l^\tau(Z^\tau Z)^- l$, the UMVUE of $(l^\tau\beta)^2$ is $(l^\tau\hat\beta)^2 - \hat\sigma^2 l^\tau(Z^\tau Z)^- l$. Since $\kappa_{n-r,-1}\hat\sigma^{-1}$ is the UMVUE of σ^{-1}, where $\kappa_{n-r,-1}$ is given in Exercise 4, and $l^\tau\hat\beta$ is independent of $\hat\sigma^2$, $\kappa_{n-r,-1}l^\tau\hat\beta\hat\sigma^{-1}$ is the UMVUE of $l^\tau\beta/\sigma$. A similar argument yields the UMVUE of $(l^\tau\beta/\sigma)^2$ as $(\kappa_{n-r,-2}l^\tau\hat\beta)^2\hat\sigma^{-2} - l^\tau(Z^\tau Z)^- l$. ∎

Exercise 33 (#3.65). Consider the one-way random effects model

$$X_{ij} = \mu + A_i + e_{ij}, \quad j = 1, ..., n, i = 1, ..., m,$$

where $\mu \in \mathcal{R}$ is an unknown parameter, A_i's are independent and identically distributed as $N(0, \sigma_a^2)$, e_{ij}'s are independent and identically distributed as $N(0, \sigma^2)$, and A_i's and e_{ij}'s are independent. Based on observed X_{ij}'s, show that the family of populations is an exponential family with sufficient and complete statistics $\bar X_{\cdot\cdot}$, $S_A = n\sum_{i=1}^m (\bar X_{i\cdot} - \bar X_{\cdot\cdot})^2$, and $S_E = \sum_{i=1}^m \sum_{j=1}^n (X_{ij} - \bar X_{i\cdot})^2$, where $\bar X_{\cdot\cdot} = (nm)^{-1}\sum_{i=1}^m\sum_{j=1}^n X_{ij}$ and $\bar X_{i\cdot} = n^{-1}\sum_{j=1}^n X_{ij}$. Find the UMVUE's of μ, σ_a^2, and σ^2.
Solution. Let $X_i = (X_{i1}, ..., X_{in})$, $i = 1, ..., m$. Then $X_1, ..., X_m$ are independent and identically distributed as the multivariate normal distribution $N_n(\mu J_n, \Sigma)$, where J_n is the n-vector of 1's and $\Sigma = \sigma_a^2 J_n J_n^\tau + \sigma^2 I_n$. The joint Lebesgue density of X_{ij}'s is

$$(2\pi)^{-\frac{mn}{2}}|\Sigma|^{-\frac{m}{2}}\exp\left\{-\frac{1}{2}\sum_{i=1}^n (X_i - \mu J_n)^\tau\Sigma^{-1}(X_i - \mu J_n)\right\}.$$

Note that

$$\Sigma^{-1} = (\sigma_a^2 J_n J_n^\tau + \sigma^2 I_n)^{-1} = \frac{1}{\sigma^2}I_n - \frac{\sigma_a^2}{\sigma^2(\sigma^2 + n\sigma_a^2)}J_n J_n^\tau.$$

Hence, the sum in the exponent of the joint density is equal to

$$\begin{aligned}
&\sum_{i=1}^{m}(X_i - \mu J_n)^\tau \Sigma^{-1}(X_i - \mu J_n)\\
&= \frac{1}{\sigma^2}\sum_{i=1}^{m}\sum_{j=1}^{n}(X_{ij}-\mu)^2 - \frac{n^2\sigma_a^2}{\sigma^2(\sigma^2+n\sigma_a^2)}\sum_{i=1}^{m}(\bar{X}_{i\cdot}-\mu)^2\\
&= \frac{1}{\sigma^2}\sum_{i=1}^{m}\sum_{j=1}^{n}(X_{ij}-\bar{X}_{i\cdot})^2 + \frac{n}{\sigma^2+n\sigma_a^2}\sum_{i=1}^{m}(\bar{X}_{i\cdot}-\mu)^2\\
&= \frac{S_E}{\sigma^2} + \frac{S_A}{\sigma^2+n\sigma_a^2} + \frac{nm}{\sigma^2+n\sigma_a^2}\sum_{i=1}^{m}(\bar{X}_{\cdot\cdot}-\mu)^2.
\end{aligned}$$

Therefore, the joint density of X_{ij}'s is from an exponential family with $(\bar{X}_{\cdot\cdot}, S_A, S_E)$ as the sufficient and complete statistics for $(\mu, \sigma_a^2, \sigma^2)$. The UMVUE of μ is $\bar{X}_{\cdot\cdot}$, since $E\bar{X}_{\cdot\cdot} = \mu$. Since $E(S_E) = m(n-1)\sigma^2$, the UMVUE of σ^2 is $S_E/[m(n-1)]$. Since $\bar{X}_{i\cdot}$, $i = 1, ..., m$ are independently from $N(\mu, \sigma_a^2 + \sigma^2/n)$, $E(S_A) = (m-1)(\sigma^2 + n\sigma_a^2)$ and, thus, the UMVUE of σ_a^2 is $S_A/[n(m-1)] - S_E/[mn(n-1)]$. ∎

Exercise 34 (#3.66). Consider the linear model $X = Z\beta + \varepsilon$, where Z is a known $n \times p$ matrix, β is a p-vector of unknown parameters, and ε is a random n-vector whose components are independent and identically distributed with mean 0 and Lebesgue density $\sigma^{-1}f(x/\sigma)$, where f is a known Lebesgue density and $\sigma > 0$ is unknown. Find the Fisher information about (β, σ) contained in X.

Solution. Let Z_i be the ith row of Z, $i = 1, ..., n$. Consider a fixed i and let $\theta = (Z_i^\tau\beta, \sigma^2)$. The Lebesgue density of X_i, the ith component of X, is $\sigma^{-1}f((x-\theta)/\sigma)$. From Exercise 20, the Fisher information about (θ, σ) contained in X_i is

$$I(\theta) = \frac{1}{\sigma^2}\begin{pmatrix} \int \frac{[f'(x)]^2}{f(x)}dx & \int \frac{f'(x)[xf'(x)+f(x)]}{f(x)}dx \\ \int \frac{f'(x)[xf'(x)+f(x)]}{f(x)}dx & \int \frac{[xf'(x)+f(x)]^2}{f(x)}dx \end{pmatrix}.$$

Let a_{ij} be the (i,j)th element of the matrix $\sigma^2 I(\theta)$. Since X_i's are independent, $\frac{\partial\theta}{\partial\beta} = Z_i^\tau$ and $\frac{\partial\theta}{\partial\sigma} = 1$, the Fisher information about $\eta = (\beta, \sigma)$ contained in X is

$$\begin{aligned}
\sum_{i=1}^{n}\frac{\partial\theta}{\partial\eta}I(\theta)\frac{\partial\theta^\tau}{\partial\eta} &= \frac{1}{\sigma^2}\sum_{i=1}^{n}\begin{pmatrix} Z_i & 0 \\ 0 & 1\end{pmatrix}\begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22}\end{pmatrix}\begin{pmatrix} Z_i^\tau & 0 \\ 0 & 1\end{pmatrix}\\
&= \frac{1}{\sigma^2}\sum_{i=1}^{n}\begin{pmatrix} a_{11}Z_iZ_i^\tau & a_{12}Z_i \\ a_{21}Z_i^\tau & a_{22}\end{pmatrix}. \quad ∎
\end{aligned}$$

Exercise 35 (#3.67). Consider the linear model $X = Z\beta + \varepsilon$, where Z is a known $n \times p$ matrix, β is a p-vector of unknown parameters, and ε is a random n-vector whose components are independent and identically distributed with mean 0 and variance σ^2. Let $c \in \mathcal{R}^p$. Show that if the equation $c = Z^\tau y$ has a solution, then there is a unique solution $y_0 \in \mathcal{R}(Z^\tau)$ such that $\mathrm{Var}(y_0^\tau X) \le \mathrm{Var}(y^\tau X)$ for any other solution of $c = Z^\tau y$.
Solution. Since $c = Z^\tau y$ has a solution, $c \in \mathcal{R}(Z) = \mathcal{R}(Z^\tau Z)$. Then, there is $\lambda \in \mathcal{R}^p$ such that $c = (Z^\tau Z)\lambda = Z^\tau y_0$ with $y_0 = Z\lambda \in \mathcal{R}(Z)$. This shows that $c = Z^\tau y$ has a solution in $\mathcal{R}(Z^\tau)$. Suppose that there is another $y_1 \in \mathcal{R}(Z^\tau)$ such that $c = Z^\tau y_1$. Then $y_0^\tau Z\beta = c^\tau\beta = y_1^\tau Z\beta$ for all $\beta \in \mathcal{R}^p$. Since $\mathcal{R}(Z^\tau) = \{Z\beta : \beta \in \mathcal{R}^p\}$, $y_0 = y_1$, i.e., the solution of $c = Z^\tau y$ in $\mathcal{R}(Z^\tau)$ is unique. For any $y \in \mathcal{R}^n$ satisfying $c = Z^\tau y$,

$$\begin{aligned}
\mathrm{Var}(y^\tau X) &= \mathrm{Var}(y^\tau X - y_0^\tau X + y_0^\tau X) \\
&= \mathrm{Var}(y^\tau X - y_0^\tau X) + \mathrm{Var}(y_0^\tau X) + 2\mathrm{Cov}((y - y_0)^\tau X, y_0^\tau X) \\
&= \mathrm{Var}(y^\tau X - y_0^\tau X) + \mathrm{Var}(y_0^\tau X) + 2E[(y - y_0)^\tau X X^\tau y_0] \\
&= \mathrm{Var}(y^\tau X - y_0^\tau X) + \mathrm{Var}(y_0^\tau X) + 2\sigma^2 (y - y_0)^\tau y_0 \\
&= \mathrm{Var}(y^\tau X - y_0^\tau X) + \mathrm{Var}(y_0^\tau X) + 2(y^\tau - y_0^\tau) Z\lambda \\
&= \mathrm{Var}(y^\tau X - y_0^\tau X) + \mathrm{Var}(y_0^\tau X) + 2(c^\tau - c^\tau)\lambda \\
&= \mathrm{Var}(y^\tau X - y_0^\tau X) + \mathrm{Var}(y_0^\tau X) \\
&\ge \mathrm{Var}(y_0^\tau X). \quad \blacksquare
\end{aligned}$$

Exercise 36 (#3.69). Consider the linear model $X = Z\beta + \varepsilon$, where Z is a known $n \times p$ matrix, β is a p-vector of unknown parameters, and ε is a random n-vector whose components are independent and identically distributed with mean 0 and variance σ^2. Let X_i be the ith component of X, Z_i be the ith row of Z, h_{ij} be the (i,j)th element of $Z(Z^\tau Z)^- Z^\tau$, $h_i = h_{ii}$, $\hat\beta$ be an LSE of β, and $\hat X_i = Z_i^\tau \hat\beta$. Show that
(i) $\mathrm{Var}(\hat X_i) = \sigma^2 h_i$;
(ii) $\mathrm{Var}(X_i - \hat X_i) = \sigma^2(1 - h_i)$;
(iii) $\mathrm{Cov}(\hat X_i, \hat X_j) = \sigma^2 h_{ij}$;
(iv) $\mathrm{Cov}(X_i - \hat X_i, X_j - \hat X_j) = -\sigma^2 h_{ij}$, $i \ne j$;
(v) $\mathrm{Cov}(\hat X_i, X_j - \hat X_j) = 0$.
Solution. (i) Since $Z_i \in \mathcal{R}(Z)$, $Z_i^\tau \beta$ is estimable and

$$\mathrm{Var}(Z_i^\tau \hat\beta) = \sigma^2 Z_i^\tau (Z^\tau Z)^- Z_i = \sigma^2 h_i.$$

(ii) Note that

$$\hat X_i = Z_i^\tau \hat\beta = Z_i^\tau (Z^\tau Z)^- Z^\tau X = \sum_{j=1}^n h_{ij} X_j.$$

Hence,

$$X_i - \hat{X}_i = (1 - h_i)X_i - \sum_{j \neq i} h_{ij} X_j.$$

Since X_i's are independent and $\mathrm{Var}(X_i) = \sigma^2$, we obtain that

$$\begin{aligned}\mathrm{Var}(X_i - \hat{X}_i) &= (1 - h_i)^2\sigma^2 + \sigma^2 \sum_{j \neq i} h_{ij}^2 \\ &= (1 - h_i)^2\sigma^2 + (h_i - h_i^2)\sigma^2 \\ &= (1 - h_i)\sigma^2,\end{aligned}$$

where the second equality follows from the fact that $\sum_{j=1}^n h_{ij}^2 = h_{ii} = h_i$, a property of the projection matrix $Z(Z^\tau Z)^- Z^\tau$.
(iii) Using the formula for $\hat{X}_i$ in part (ii) of the solution and the independence of X_i's,

$$\mathrm{Cov}(\hat{X}_i, \hat{X}_j) = \mathrm{Cov}\left(\sum_{k=1}^n h_{ik}X_k, \sum_{l=1}^n h_{jl}X_l\right) = \sigma^2 \sum_{k=1}^m h_{ik}h_{jk} = \sigma^2 h_{ij},$$

where the last equality follows from the fact that $Z(Z^\tau Z)^- Z^\tau$ is a projection matrix.
(iv) For $i \neq j$,

$$\mathrm{Cov}(X_i, \hat{X}_j) = \mathrm{Cov}\left(X_i, \sum_{k=1}^n h_{jk}X_k\right) = \sigma^2 h_{ij}$$

and, thus,

$$\begin{aligned}\mathrm{Cov}(X_i - \hat{X}_i, X_j - \hat{X}_j) &= -\mathrm{Cov}(X_i, \hat{X}_j) - \mathrm{Cov}(X_j, \hat{X}_i) + \mathrm{Cov}(\hat{X}_i, \hat{X}_j) \\ &= -\sigma^2 h_{ij} - \sigma^2 h_{ji} + \sigma^2 h_{ij} \\ &= -\sigma^2 h_{ij}.\end{aligned}$$

(v) From part (iii) and part (iv) of the solution,

$$\mathrm{Cov}(\hat{X}_i, X_j - \hat{X}_j) = \mathrm{Cov}(\hat{X}_i, X_j) - \mathrm{Cov}(\hat{X}_i, \hat{X}_j) = \sigma^2 h_{ij} - \sigma^2 h_{ij} = 0. \blacksquare$$

Exercise 37 (#3.70). Consider the linear model $X = Z\beta + \varepsilon$, where Z is a known $n \times p$ matrix, β is a p-vector of unknown parameters, and ε is a random n-vector whose components are independent and identically distributed with mean 0 and variance σ^2. Let $Z = (Z_1, Z_2)$ and $\beta = (\beta_1, \beta_2)$, where Z_j is $n \times p_j$ and β_j is a p_j-vector, $j = 1, 2$. Assume that $(Z_1^\tau Z_1)^{-1}$ and $[Z_2^\tau Z_2 - Z_2^\tau Z_1(Z_1^\tau Z_1)^{-1} Z_1^\tau Z_2]^{-1}$ exist.
(i) Derive the LSE of β in terms of Z_1, Z_2, and X.

(ii) Let $\hat{\beta} = (\hat{\beta}_1, \hat{\beta}_2)$ be the LSE in (i). Calculate the covariance between $\hat{\beta}_1$ and $\hat{\beta}_2$.
(iii) Suppose that it is known that $\beta_2 = 0$. Let $\tilde{\beta}_1$ be the LSE of β_1 under the reduced model $X = Z_1\beta_1 + \varepsilon$. Show that, for any $l \in \mathcal{R}^{p_1}$, $l^\tau \tilde{\beta}_1$ is better than $l^\tau \hat{\beta}_1$ in terms of their variances.
Solution. (i) Note that

$$Z^\tau Z = \begin{pmatrix} Z_1^\tau Z_1 & Z_1^\tau Z_2 \\ Z_2^\tau Z_1 & Z_2^\tau Z_2 \end{pmatrix}.$$

From matrix algebra,

$$(Z^\tau Z)^{-1} = \begin{pmatrix} A & B \\ B^\tau & C \end{pmatrix},$$

where

$$C = [Z_2^\tau Z_2 - Z_2^\tau Z_1 (Z_1^\tau Z_1)^{-1} Z_1^\tau Z_2]^{-1},$$
$$B = -(Z_1^\tau Z_1)^{-1} C$$

and

$$A = (Z_1^\tau Z_1)^{-1} + (Z_1^\tau Z_1)^{-1} Z_1^\tau Z_2 C Z_2^\tau Z_1 (Z_1^\tau Z_1)^{-1}.$$

The LSE of β is

$$\hat{\beta} = (Z^\tau Z)^{-1} Z^\tau X = \begin{pmatrix} A & B \\ B^\tau & C \end{pmatrix} \begin{pmatrix} Z_1^\tau X \\ Z_2^\tau X \end{pmatrix} = \begin{pmatrix} A Z_1^\tau X + B Z_2^\tau X \\ B^\tau Z_1^\tau X + C Z_2^\tau X \end{pmatrix}.$$

(ii) Since $\text{Var}(\hat{\beta}) = \sigma^2 (Z^\tau Z)^{-1}$, $\text{Cov}(\hat{\beta}_1, \hat{\beta}_2) = \sigma^2 B$.
(iii) Note that $\text{Var}(l^\tau \tilde{\beta}_1) = \sigma^2 l^\tau (Z_1^\tau Z_1)^{-1} l$. From part (i) of the solution,

$$\text{Var}(l^\tau \hat{\beta}_1) = \sigma^2 l^\tau A l \geq \sigma^2 l^\tau (Z_1^\tau Z_1)^{-1} l. \blacksquare$$

Exercise 38 (#3.71, #3.72). Consider the linear model $X = Z\beta + \varepsilon$, where Z is a known $n \times p$ matrix, β is a p-vector of unknown parameters, and ε is a random n-vector with $E(\varepsilon) = 0$ and finite $\text{Var}(\varepsilon) = \Sigma$. Show the following statements are equivalent:
(a) The LSE $l^\tau \hat{\beta}$ is the best linear unbiased estimator (BLUE) of $l^\tau \beta$.
(e) $\text{Var}(\varepsilon) = Z\Lambda_1 Z^\tau + U\Lambda_2 U^\tau$ for some matrices Λ_1 and Λ_2, where U is a matrix such that $Z^\tau U = 0$ and $\mathcal{R}(U^\tau) + \mathcal{R}(Z^\tau) = \mathcal{R}^n$.
(f) $\text{Var}(\varepsilon) Z = ZB$ for some matrix B.
(g) $\mathcal{R}(Z^\tau)$ is generated by r eigenvectors of $\text{Var}(\varepsilon)$, where r is the rank of Z.
Solution. (i) From the proof in Shao (2003, p. 191), (a) is equivalent to (c) $Z^\tau \text{Var}(\varepsilon) U = 0$ and (c) implies (e). Hence, to show that (a) and (e) are

equivalent, it suffices to show that (e) implies (c). Since $Z(Z^\tau Z)^- Z^\tau Z = Z$, (e) implies that

$$Z^\tau \mathrm{Var}(\varepsilon)U = Z^\tau Z Z(Z^\tau Z)^- Z^\tau \mathrm{Var}(\varepsilon)U = Z^\tau Z\mathrm{Var}(\varepsilon)Z(Z^\tau Z)^- Z^\tau U = 0.$$

(ii) We now show that (f) and (c) are equivalent. If (f) holds, then $\mathrm{Var}(\varepsilon)Z = ZB$ for some matrix B and

$$Z^\tau \mathrm{Var}(\varepsilon)U = B^\tau Z^\tau U = 0.$$

If (c) holds, then (e) holds. Then

$$\mathrm{Var}(\varepsilon)Z = \mathrm{Var}(\varepsilon)Z(Z^\tau Z)^- Z^\tau Z = Z(Z^\tau Z)^- Z^\tau \mathrm{Var}(\varepsilon)Z$$

and (f) holds with $B = (Z^\tau Z)^- Z^\tau \mathrm{Var}(\varepsilon)Z$.
(iii) Assume that (g) holds. Then $\mathcal{R}(Z^\tau) = \mathcal{R}(\xi_1, ..., \xi_r)$, the linear space generated by r linearly independent eigenvectors $\xi_1, ..., \xi_r$ of $\mathrm{Var}(\varepsilon)$. Let $\xi_{r+1}, ..., \xi_n$ be the other $n - r$ linearly independent eigenvectors of $\mathrm{Var}(\varepsilon)$ that are orthogonal to $\xi_1, ..., \xi_r$. Then $\mathcal{R}(U^\tau) = \mathcal{R}(\xi_{r+1}, ..., \xi_n)$. For $j \le r$, $\mathrm{Var}(\varepsilon)\xi_j = a_j\xi_j$ for some constant a_j. For $k \ge r+1$, $\xi_j^\tau \mathrm{Var}(\varepsilon)\xi_k = a\xi_j^\tau \xi_k = 0$. Hence, $Z^\tau \mathrm{Var}(\varepsilon)U = 0$, i.e., (c) holds.

Now, assume (c) holds. Let $\xi_1, ..., \xi_n$ be n orthogonal eigenvectors of $\mathrm{Var}(\varepsilon)$ and M be the matrix with ξ_i as the ith column. Decompose M as $M = M_Z + M_U$, where columns of M_Z are in $\mathcal{R}(Z^\tau)$ and columns of M_U are in $\mathcal{R}(U^\tau)$. Then

$$\mathrm{Var}(\varepsilon)M_Z + \mathrm{Var}(\varepsilon)M_U = M_Z D + M_U D,$$

where D is a diagonal matrix. Multiplying the transposes of both sides of the above equation by M_U from the right, we obtain that, by (c),

$$M_U^\tau \mathrm{Var}(\epsilon)M_U = DM_U^\tau M_U$$

which is the same as

$$\mathrm{Var}(\varepsilon)M_U = M_U D,$$

and, hence,

$$\mathrm{Var}(\varepsilon)M_Z = M_Z D.$$

This means that column vectors of M_Z are eigenvectors of $\mathrm{Var}(\varepsilon)$. Then (g) follows from $\mathcal{R}(Z) = \mathcal{R}(M_Z)$. ∎

Exercise 39 (#3.74). Suppose that

$$X = \mu J_n + H\xi + e,$$

where $\mu \in \mathcal{R}$ is an unknown parameter, J_n is the n-vector of 1's, H is an $n \times p$ known matrix of full rank, ξ is a random p-vector with $E(\xi) = 0$ and

$\text{Var}(\xi) = \sigma_\xi^2 I_p$, e is a random n-vector with $E(e) = 0$ and $\text{Var}(e) = \sigma^2 I_n$, and ξ and e are independent. Show that the LSE of μ is the BLUE if and only if the row totals of HH^τ are the same.

Solution. From the result in the previous exercise, it suffices to show that the LSE of μ is the BLUE if and only if J_n is an eigenvector of $\text{Var}(H\xi + e) = \sigma_\xi^2 HH^\tau + \sigma^2 I_n$. Since

$$(\sigma_\xi^2 HH^\tau + \sigma^2 I_n)J_n = \sigma_\xi^2 \eta + \sigma^2 J_n,$$

where η is the vector of row totals of HH^τ, J_n is an eigenvector of the matrix $\text{Var}(H\xi + e)$ if and only if $\eta = cJ_n$ for some constant. ∎

Exercise 40 (#3.75). Consider a linear model

$$X_{ij} = \mu + \alpha_i + \beta_j + \varepsilon_{ij}, \quad i = 1, ..., a, j = 1, ..., b,$$

where μ, α_i's, and β_j's are unknown parameters, $E(\varepsilon_{ij}) = 0$, $\text{Var}(\varepsilon_{ij}) = \sigma^2$, $\text{Cov}(\varepsilon_{ij}, \varepsilon_{i'j'}) = 0$ if $i \neq i'$, and $\text{Cov}(\varepsilon_{ij}, \varepsilon_{ij'}) = \sigma^2\rho$ if $j \neq j'$. Show that the LSE of $l^\tau\beta$ is the BLUE for any $l \in \mathcal{R}(Z)$.

Solution. Write the model in the form of $X = Z\beta + \varepsilon$. Then $\text{Var}(\varepsilon)$ is a block diagonal matrix whose jth diagonal block is $\sigma^2(1-\rho)I_a + \sigma^2\rho J_a J_a^\tau$, $j = 1, ..., b$, where I_a is the identity matrix of order a and J_a is the a-vector of 1's. Let A and B be as defined in Exercise 31. Then $Z = (J_{ab}\ A\ B^\tau)$. Let Λ be the $(1+a+b) \times (1+a+b)$ matrix whose first element is $\sigma^2\rho$ and all the other elements are 0. Then, $Z\Lambda Z^\tau$ is a block diagonal matrix whose jth diagonal block is $\sigma^2\rho J_a J_a^\tau$, $j = 1, ..., b$. Thus,

$$\text{Var}(\varepsilon) = \sigma^2(1-\rho)I_{ab} + Z\Lambda Z^\tau.$$

This shows that (c) in Exercise 38 holds. Hence, the LSE of $l^\tau\beta$ is the BLUE for any $l \in \mathcal{R}(Z)$. ∎

Exercise 41 (#3.76). Consider the linear model $X = Z\beta + \varepsilon$, where Z is a known $n \times p$ matrix, β is a p-vector of unknown parameters, and ε is a random n-vector with $E(\varepsilon) = 0$ and $\text{Var}(\varepsilon) =$ a block diagonal matrix whose ith block diagonal V_i is $n_i \times n_i$ and has a single eigenvalue λ_i with eigenvector J_{n_i} (the n_i-vector of 1's) and a repeated eigenvalue ρ_i with multiplicity $n_i - 1$, $i = 1, ..., k$, $\sum_{i=1}^k n_i = n$. Let U be the $n \times k$ matrix whose ith column is U_i, where $U_1 = (J_{n_1}^\tau, 0, ..., 0)$, $U_2 = (0, J_{n_2}^\tau, ..., 0)$,..., $U_k = (0, 0, ..., J_{n_k}^\tau)$, and let $\hat{\beta}$ be the LSE of β.

(i) If $\mathcal{R}(Z^\tau) \subset \mathcal{R}(U^\tau)$ and $\lambda_i \equiv \lambda$, show that $l^\tau\hat{\beta}$ is the BLUE of $l^\tau\beta$ for any $l \in \mathcal{R}(Z)$.

(ii) If $Z^\tau U_i = 0$ for all i and $\rho_i \equiv \rho$, show that $l^\tau\hat{\beta}$ is the BLUE of $l^\tau\beta$ for any $l \in \mathcal{R}(Z)$.

Solution. (i) Condition $\mathcal{R}(Z^\tau) \subset \mathcal{R}(U^\tau)$ implies that there exists a matrix B such that $Z = UB$. Then

$$\text{Var}(\varepsilon)Z = \text{Var}(\varepsilon)UB = \lambda UB = \lambda Z$$

and, thus,

$$Z(Z^\tau Z)^- Z^\tau \text{Var}(\varepsilon) = \lambda Z(Z^\tau Z)^- Z^\tau,$$

which is symmetric. Hence the result follows from the result in Exercise 38.

(ii) Let Λ_ρ be the $(n-k) \times (n-k)$ matrix whose columns are the $n-k$ eigenvectors corresponding to the eigenvalue ρ. Then $Z^\tau U_i = 0$ for all i implies that $\mathcal{R}(Z^\tau) \subset \mathcal{R}(\Lambda_\rho^\tau)$ and there exists a matrix C such that $Z = \Lambda_\rho C$. Since

$$\text{Var}(\varepsilon)Z = \text{Var}(\varepsilon)\Lambda_\rho C = \rho\Lambda_\rho C = \rho Z,$$

we obtain that

$$Z(Z^\tau Z)^- Z^\tau \text{Var}(\varepsilon) = \rho Z(Z^\tau Z)^- Z^\tau,$$

which is symmetric. Hence the result follows from the result in Exercise 38. ∎

Exercise 42 (#3.80). Consider the linear model $X = Z\beta + \varepsilon$, where Z is a known $n \times p$ matrix, β is a p-vector of unknown parameters, and $\varepsilon = (\varepsilon_1, ..., \varepsilon_n)$ with independent and identically distributed $\varepsilon_1, ..., \varepsilon_n$ having $E(\varepsilon_i) = 0$ and $\text{Var}(\varepsilon_i) = \sigma^2$. Let Z_i be the ith row of Z, $\hat{X}_i = Z_i^\tau \hat{\beta}$, $\hat{\beta}$ be the LSE of β, and $h_i = Z_i^\tau (Z^\tau Z)^- Z_i$.

(i) Show that for any $\epsilon > 0$,

$$P(|\hat{X}_i - E\hat{X}_i| \geq \epsilon) \geq \min\{P(\varepsilon_i \geq \epsilon/h_i), P(\varepsilon_i \leq -\epsilon/h_i)\}.$$

(ii) Show that $\hat{X}_i - E\hat{X}_i \to_p 0$ if and only if $\lim_n h_i = 0$.

Solution. (i) For independent random variables U and Y and $\epsilon > 0$,

$$\begin{aligned} P(|U + Y| \geq \epsilon) &\geq P(U \geq \epsilon)P(Y \geq 0) + P(U \leq -\epsilon)P(Y < 0) \\ &\geq \min\{P(U \geq \epsilon), P(U \leq -\epsilon)\}. \end{aligned}$$

Using the result in the solution of Exercise 36,

$$\hat{X}_i - E\hat{X}_i = \sum_{j=1}^n h_{ij}(X_j - EX_j) = \sum_{j=1}^n h_{ij}\varepsilon_j = h_i\varepsilon_i + \sum_{j \neq i} h_{ij}\varepsilon_j.$$

Then the result follows by taking $U = h_i\varepsilon_i$ and $Y = \sum_{j \neq i} h_{ij}\varepsilon_j$.

(ii) If $\hat{X}_i - E\hat{X}_i \to_p 0$, then it follows from the result in (i) that

$$\lim_n \min\{P(\varepsilon_i \geq \epsilon/h_i), P(\varepsilon_i \leq -\epsilon/h_i)\} = 0,$$

which holds only if $\lim_n h_i = 0$. Suppose now that $\lim_n h_i = 0$. From Exercise 36, $\lim_n \text{Var}(\hat{X}_i) = \lim_n \sigma^2 h_i = 0$. Therefore, $\hat{X}_i - E\hat{X}_i \to_p 0$. ∎

Exercise 43 (#3.81). Let Z be an $n \times p$ matrix, Z_i be the ith row of Z, $h_i = Z_i^\tau (Z^\tau Z)^- Z_i$, and λ_n be the largest eigenvalue of $(Z^\tau Z)^-$. Show that if $\lim_n \lambda_n = 0$ and $\lim_n Z_n^\tau (Z^\tau Z)^- Z_n = 0$, then $\lim_n \max_{1\leq i\leq n} h_i = 0$.
Solution. Since $Z^\tau Z$ depends on n, we denote $(Z^\tau Z)^-$ by A_n. Let i_n be the integer such that $h_{i_n} = \max_{1\leq i\leq n} h_i$. If $\lim_n i_n = \infty$, then

$$\lim_n h_{i_n} = \lim_n Z_{i_n}^\tau A_n Z_{i_n} \leq \lim_n Z_{i_n}^\tau A_{i_n} Z_{i_n} = 0,$$

where the inequality follows from $i_n \leq n$ and, thus, $A_{i_n} - A_n$ is nonnegative definite. If $i_n \leq c$ for all n, then

$$\lim_n h_{i_n} = \lim_n Z_{i_n}^\tau A_n Z_{i_n} \leq \lim_n \lambda_n \max_{1\leq i\leq c} \|Z_i\|^2 = 0.$$

Therefore, for any subsequence $\{j_n\} \subset \{i_n\}$ with $\lim_n j_n = a \in (0, \infty]$, $\lim_n h_{j_n} = 0$. This shows that $\lim_n h_{i_n} = 0$. ∎

Exercise 44 (#3.84). Consider the one-way random effects model

$$X_{ij} = \mu + A_i + e_{ij}, \quad j = 1, ..., n_i, i = 1, ..., m,$$

where $\mu \in \mathcal{R}$ is an unknown parameter, A_i's are independent and identically distributed with mean 0 and variance σ_a^2, e_{ij}'s are independent with mean 0, and A_i's and e_{ij}'s are independent. Assume that $\{n_i\}$ is bounded and $E|e_{ij}|^{2+\delta} < \infty$ for some $\delta > 0$. Show that the LSE $\hat{\mu}$ of μ is asymptotically normal and derive an explicit form of $\text{Var}(\hat{\mu})$.
Solution. The LSE of μ is $\hat{\mu} = \bar{X}_{\cdot\cdot}$, the average of X_{ij}'s. The model under consideration can be written as $X = Z\mu + \varepsilon$ with $Z = J_n$, $Z^\tau Z = n$, and

$$\lim_n \max_{1\leq i\leq n} Z_i^\tau (Z^\tau Z)^- Z_i = \lim_n \frac{1}{n} = 0.$$

Since we also have $E|e_{ij}|^{2+\delta} < \infty$ and $\{n_i\}$ is bounded, by Theorem 3.12(i) in Shao (2003),

$$\frac{\hat{\mu} - \mu}{\sqrt{\text{Var}(\hat{\mu})}} \to_d N(0, 1),$$

where $\text{Var}(\hat{\mu}) = \text{Var}(\bar{X}_{\cdot\cdot}) = n^{-2} \sum_{i=1}^m (n_i^2 \sigma_a^2 + n_i \sigma^2)$. ∎

Exercise 45 (#3.85). Suppose that

$$X_i = \rho t_i + \varepsilon_i, \quad i = 1, ..., n,$$

where $\rho \in \mathcal{R}$ is an unknown parameter, t_i's are known and in (a, b), a and b are known positive constants, and ε_i's are independent random variables

satisfying $E(\varepsilon_i) = 0$, $E|\varepsilon_i|^{2+\delta} < \infty$ for some $\delta > 0$, and $\text{Var}(\varepsilon_i) = \sigma^2 t_i$ with an unknown $\sigma^2 > 0$.
(i) Obtain the LSE of ρ.
(ii) Obtain the BLUE of ρ.
(iii) Show that both the LSE and BLUE are asymptotically normal and obtain the asymptotic relative efficiency of the BLUE with respect to the LSE.

Solution. (i) The LSE of ρ is

$$\hat{\rho} = \frac{\sum_{i=1}^n t_i X_i}{\sum_{i=1}^n t_i^2}.$$

(iii) Let $X = (X_1, ..., X_n)$ and $c = (c_1, ..., c_n)$. Consider minimizing

$$E(c^\tau X - \rho)^2 = \sum_{i=1}^n t_i c_i^2$$

under the constraint $\sum_{i=1}^n c_i t_i = 1$ (to ensure unbiasedness), which yields $c_i = (\sum_{i=1}^n t_i)^{-1}$. Hence, the BLUE of ρ is

$$\tilde{\rho} = \frac{\sum_{i=1}^n X_i}{\sum_{i=1}^n t_i}.$$

(iii) The asymptotic normality of the LSE and BLUE follows directly from Lindeberg's central limit theorem. Since

$$\text{Var}(\hat{\rho}) = \frac{\sigma^2 \sum_{i=1}^n t_i^3}{(\sum_{i=1}^n t_i^2)^2}$$

and

$$\text{Var}(\tilde{\rho}) = \frac{\sigma^2}{\sum_{i=1}^n t_i},$$

the asymptotic relative efficiency of the BLUE with respect to the LSE is

$$\frac{\left(\sum_{i=1}^n t_i^2\right)^2}{(\sum_{i=1}^n t_i^3)(\sum_{i=1}^n t_i)}. \quad \blacksquare$$

Exercise 46 (#3.87). Suppose that $X = (X_1, ..., X_n)$ is a simple random sample without replacement from a finite population $\mathcal{P} = \{y_1, ..., y_N\}$ with all $y_i \in \mathcal{R}$.
(i) Show that a necessary condition for $h(y_1, ..., y_N)$ to be estimable is that h is symmetric in its N arguments.
(ii) Find the UMVUE of $P(X_i \le X_j)$, $i \neq j$.
(iii) Find the UMVUE of $\text{Cov}(X_i, X_j)$, $i \neq j$.

Solution. (i) If $h(y_1, ..., y_N)$ is estimable, then there exists a function $u(x_1, ..., x_n)$ that is symmetric in its arguments and satisfies

$$h(y_1, ..., y_N) = E[u(X_1, ..., X_n)] = \frac{1}{\binom{N}{n}} \sum_{1 \le i_1 < \cdots < i_n \le N} u(y_{i_1}, ..., y_{i_n}).$$

Hence, h is symmetric in its arguments.
(ii) From Watson-Royall's theorem (e.g., Theorem 3.13 in Shao, 2003), the order statistics are complete and sufficient. Hence, for any estimable parameter, its UMVUE is the unbiased estimator $g(X_1, ..., X_n)$ that is symmetric in its arguments. Thus, the UMVUE of $P(X_i \le X_j)$, $i \neq j$, is

$$\frac{1}{\binom{n}{2}} \sum_{1 \le i < j \le n} \frac{I_{(-\infty, X_i]}(X_j) + I_{(-\infty, X_j]}(X_i)}{2}.$$

(iii) From the argument in part (ii) of the solution, the UMVUE of $E(X_i X_j)$ when $i \neq j$ is

$$U_1 = \frac{1}{\binom{n}{2}} \sum_{1 \le i < j \le n} X_i X_j.$$

Let $\bar{X}$ be the sample mean. Since

$$E(\bar{X}^2) = \frac{1}{nN} \sum_{i=1}^N y_i^2 + \frac{2(n-1)}{nN(N-1)} \sum_{1 \le i < j \le N} y_i y_j$$

and

$$E\left(\frac{1}{n} \sum_{i=1}^n X_i^2\right) = \frac{1}{N} \sum_{i=1}^N y_i^2,$$

the UMVUE of $2\sum_{1 \le i < j \le N} y_i y_j$ is

$$U_2 = \frac{nN(N-1)}{n-1} \left(\bar{X}^2 - \frac{1}{n^2} \sum_{i=1}^n X_i^2\right).$$

From

$$\begin{aligned} \text{Cov}(X_i, X_j) &= E(X_i X_j) - \left(\frac{1}{N} \sum_{i=1}^N y_i\right)^2 \\ &= E(X_i X_j) - \frac{1}{N^2} \sum_{i=1}^N y_i^2 - \frac{2}{N^2} \sum_{1 \le i < j \le n} y_i y_j, \end{aligned}$$

the UMVUE of $\text{Cov}(X_i, X_j)$, $i \neq j$, is

$$U_1 - \frac{1}{nN} \sum_{i=1}^n X_i^2 - \frac{U_2}{N^2}. \quad \blacksquare$$

Exercise 47 (#3.100). Let $(X_1, ..., X_n)$ be a random sample from the normal distribution $N(\mu, \sigma^2)$, where $\mu \in \mathcal{R}$ and $\sigma^2 > 0$. Consider the estimation of $\vartheta = E[\Phi(a + bX_1)]$, where Φ is the cumulative distribution function of $N(0, 1)$ and a and b are known constants. Obtain an explicit form of a function $g(\mu, \sigma^2) = \vartheta$ and the asymptotic mean squared error of $\hat{\vartheta} = g(\bar{X}, S^2)$, where $\bar{X}$ and S^2 are the sample mean and variance.

Solution. Let Z be a random variable that has distribution $N(0, 1)$ and is independent of X_1. Define $Y = Z - bX_1$. Then Y has distribution $N(-b\mu, 1 + b^2\sigma^2)$ and

$$\begin{aligned} E[\Phi(a + bX_1)] &= E[P(Z \le a + bX_1)] \\ &= P(Z - bX_1 \le a) \\ &= P(Y \le a) \\ &= \Phi\left(\frac{a + b\mu}{\sqrt{1 + b^2\sigma^2}}\right). \end{aligned}$$

Hence

$$g(\mu, \sigma^2) = \Phi\left(\frac{a + b\mu}{\sqrt{1 + b^2\sigma^2}}\right).$$

From Example 2.8 in Shao (2003),

$$\sqrt{n}\begin{pmatrix} \bar{X} - \mu \\ S^2 - \sigma^2 \end{pmatrix} \to_d N_2\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma^2 & 0 \\ 0 & 2\sigma^4 \end{pmatrix}\right).$$

Then, by the δ-method,

$$\sqrt{n}(\hat{\vartheta} - \vartheta) = \sqrt{n}[g(\bar{X}, S^2) - \vartheta] \to_d N(0, \kappa),$$

where

$$\kappa = \left[\frac{b^2\sigma^2}{1 + b^2\sigma^2} + \frac{(a + b\mu)^2 b^4\sigma^2}{2(1 + b^2\sigma^2)}\right]\left[\Phi'\left(\frac{a + b\mu}{\sqrt{1 + b^2\sigma^2}}\right)\right]^2.$$

The asymptotic mean squared error of $\hat{\vartheta}$ is κ/n. ∎

Exercise 48 (#3.103). Let $(X_1, ..., X_n)$ be a random sample from P in a parametric family. Obtain moment estimators of parameters in the following cases.

(i) P is the gamma distribution with shape parameter $\alpha > 0$ and scale parameter $\gamma > 0$.

(ii) P has Lebesgue density $\theta^{-1}e^{-(x-a)/\theta}I_{(a,\infty)}(x)$, $a \in \mathcal{R}$, $\theta > 0$.

(iii) P has Lebesgue density $\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}x^{\alpha-1}(1-x)^{\beta-1}I_{(0,1)}(x)$, $\alpha > 0$, $\beta > 0$.

(iv) P is the log-normal distribution with parameter (μ, σ^2) (i.e., $\log X_1$

has distribution $N(\mu, \sigma^2)$, $\mu \in \mathcal{R}$, $\sigma > 0$.
(v) P is the negative binomial distribution with discrete probability density $\binom{x-1}{r-1} p^r (1-p)^{x-r}$, $x = r, r+1, ..., p \in (0,1)$, $r = 1, 2, ...$.
Solution. Let $\mu_k = E(X_1^k)$ and $\hat{\mu}_k = n^{-1} \sum_{i=1}^n X_i^k$.
(i) Note that $\mu_1 = \alpha\gamma$ and $\mu_2 - \mu_1^2 = \alpha\gamma^2$. Hence, the moment estimators are $\hat{\gamma} = (\hat{\mu}_2 - \hat{\mu}_1^2)/\hat{\mu}_1$ and $\hat{\alpha} = \hat{\mu}_1^2/(\hat{\mu}_2 - \hat{\mu}_1^2)$.
(ii) Note that $\mu_1 = a + \theta$ and $\mu_2 - \mu_1^2 = \theta^2$. Hence, the moment estimators are $\hat{\theta} = \sqrt{\hat{\mu}_2 - \hat{\mu}_1^2}$ and $\hat{a} = \hat{\mu}_1 - \hat{\theta}$.
(iii) Note that $\mu_1 = \alpha/(\alpha+\beta)$ and $\mu_2 = \alpha(\alpha+1)/[(\alpha+\beta)(\alpha+\beta+1)]$. Then $1 + \beta/\alpha = \mu_1^{-1}$, which leads to $\mu_2 = \mu_1(1+\alpha^{-1})/(\mu_1^{-1} + \alpha^{-1})$. Then the moment estimators are $\hat{\alpha} = \hat{\mu}_1(\hat{\mu}_1 - \hat{\mu}_2)/(\hat{\mu}_2 - \hat{\mu}_1^2)$ and $\hat{\beta} = (\hat{\mu}_1 - \hat{\mu}_2)(1 - \hat{\mu}_1)/(\hat{\mu}_2 - \hat{\mu}_1^2)$.
(iv) Note that $\mu_1 = e^{\mu + \sigma^2/2}$ and $\mu_2 = e^{2\mu + 2\sigma^2}$. Then $\mu_2/\mu_1^2 = e^{\sigma^2}$, i.e., $\sigma^2 = \log(\mu_2/\mu_1^2)$. Then $\mu = \log \mu_1 + \sigma^2/2$. Hence, the moment estimators are $\hat{\sigma}^2 = \log(\hat{\mu}_2/\hat{\mu}_1^2)$ and $\hat{\mu} = \log \hat{\mu}_1 - \frac{1}{2}\log(\hat{\mu}_2/\hat{\mu}_1^2)$.
(v) Note that $\mu_1 = r/p$ and $\mu_2 - \mu_1^2 = r(1-p)/p^2$. Then $r = p\mu_1$ and $(\mu_2 - \mu_1^2)p = \mu_1(1-p)$. Hence, the moment estimators are $\hat{p} = \hat{\mu}_1/(\hat{\mu}_2 - \hat{\mu}_1^2 + \hat{\mu}_1)$ and $\hat{r} = \hat{\mu}_1^2/(\hat{\mu}_2 - \hat{\mu}_1^2 + \hat{\mu}_1)$. ∎

Exercise 49 (#3.106). In Exercise 11(i), find a moment estimator of θ and derive its asymptotic distribution. In Exercise 11(ii), obtain a moment estimator of θ^{-1} and its asymptotic relative efficiency with respect to the UMVUE of θ^{-1}.
Solution. (i) From Exercise 11(i),

$$\mu_1 = EX_1 = P(Y_1 < 1) + \frac{1}{\theta}\int_1^\theta x dx = \frac{1}{\theta} + \frac{\theta^2 - 1}{2\theta} = \frac{\theta^2 + 1}{2\theta}.$$

Let $\bar{X}$ be the sample mean. Setting $\bar{X} = (\theta^2+1)/(2\theta)$, we obtain that $\theta^2 - 2\bar{X}\theta + 1 = 0$, which has solutions $\bar{X} \pm \sqrt{\bar{X}^2 - 1}$. Since $\bar{X} \geq 1$, $\bar{X} - \sqrt{\bar{X}^2 - 1} < 1$. Since $\theta \geq 1$, the moment estimator of θ is $\hat{\theta} = \bar{X} + \sqrt{\bar{X}^2 - 1}$.

From the central limit theorem,

$$\sqrt{n}(\bar{X} - \mu_1) \to_d N\left(0, \frac{\theta^3 + 2}{3\theta} - \frac{(\theta^2+2)^2}{4\theta^2}\right).$$

By the δ-method with $g(x) = x + \sqrt{x^2 - 1}$,

$$\sqrt{n}(\hat{\theta} - \theta) \to_d N\left(0, \left(1 + \frac{\theta}{\sqrt{\theta^2 - 1}}\right)^2 \left[\frac{\theta^3+2}{3\theta} - \frac{(\theta^2+2)^2}{4\theta^2}\right]\right).$$

(ii) From Exercise 11(ii),

$$\mu_1 = EX_1 = \frac{1}{\theta}\int_0^1 x dx + P(Y_1 > 1) = \frac{1}{2\theta} + 1 - \frac{1}{\theta} = 1 - \frac{1}{2\theta}.$$

Hence the moment estimator of θ^{-1} is $2(1-\bar{X})$. From the central limit theorem,

$$\sqrt{n}(\bar{X}-\mu_1) \to_d N\left(0, \frac{1}{3\theta} - \frac{1}{4\theta^2}\right).$$

By the δ-method with $g(x) = 2(1-x)$,

$$\sqrt{n}[2(1-\bar{X}) - \theta^{-1}] \to_d N\left(0, \frac{4}{3\theta} - \frac{1}{\theta^2}\right).$$

Let $R_i = 0$ if $X_i = 1$ and $R_i = 1$ if $X_i \neq 1$. From the solution of Exercise 11(ii), the UMVUE of θ^{-1} is $\bar{R} = n^{-1}\sum_{i=1}^n R_i$. By the central limit theorem,

$$\sqrt{n}(\bar{R}-\theta^{-1}) \to_d N\left(0, \frac{1}{\theta} - \frac{1}{\theta^2}\right).$$

Hence, the asymptotic relative efficiency of $2(1-\bar{X})$ with respect to $\bar{R}$ is equal to $(\theta-1)/(\frac{4}{3}\theta - 1)$. ▮

Exercise 50 (#3.107). Let $(X_1, ..., X_n)$ be a random sample from a population having the Lebesgue density $f_{\alpha,\beta}(x) = \alpha\beta^{-\alpha}x^{\alpha-1}I_{(0,\beta)}(x)$, where $\alpha > 0$ and $\beta > 0$ are unknown. Obtain a moment estimator of $\theta = (\alpha, \beta)$ and its asymptotic distribution.

Solution. Let $\mu_j = EX_1^j$. Note that

$$\mu_1 = \frac{\alpha}{\beta^\alpha}\int_0^\beta x^\alpha dx = \frac{\alpha\beta}{\alpha+1}$$

and

$$\mu_2 = \frac{\alpha}{\beta^\alpha}\int_0^\beta x^{\alpha+1} dx = \frac{\alpha\beta^2}{\alpha+2}.$$

Then $\beta = (1+\frac{1}{\alpha})\mu_1$ and

$$\left(1+\frac{1}{\alpha}\right)^2 \mu_1^2 = \left(1+\frac{2}{\alpha}\right)\mu_2,$$

which leads to

$$\frac{1}{\alpha} = \frac{\mu_2 - \mu_1^2 \pm \sqrt{\mu_2^2 - \mu_1\mu_2}}{\mu_1^2}.$$

Since $\alpha > 0$, we obtain the moment estimators

$$\hat{\alpha} = \frac{\hat{\mu}_1^2}{\hat{\mu}_2 - \hat{\mu}_1^2 + \sqrt{\hat{\mu}_2^2 - \hat{\mu}_1\hat{\mu}_2}}$$

and

$$\hat{\beta} = \frac{\hat{\mu}_2 + \sqrt{\hat{\mu}_2^2 - \hat{\mu}_1\hat{\mu}_2}}{\hat{\mu}_1},$$

where $\hat{\mu}_j = n^{-1}\sum_{i=1}^n X_i^j$. Let $\gamma = (\mu_1, \mu_2)$ and $\hat{\gamma} = (\hat{\mu}_1, \hat{\mu}_2)$. From the central limit theorem,

$$\sqrt{n}(\hat{\gamma} - \gamma) \to_d N(0, \Sigma),$$

where

$$\Sigma = \begin{pmatrix} \mu_2 - \mu_1^2 & \mu_3 - \mu_1\mu_2 \\ \mu_3 - \mu_1\mu_2 & \mu_4 - \mu_2^2 \end{pmatrix}.$$

Let $\alpha(x, y) = x^2/(y - x^2 + \sqrt{y^2 - xy})$ and $\beta(x, y) = (y + \sqrt{y^2 - xy})/x$. Then

$$\frac{\partial(\alpha, \beta)}{\partial(x, y)} = \begin{pmatrix} \frac{2x}{y-x^2+\sqrt{y^2-xy}} + \frac{x^2(4x+y/\sqrt{y^2-xy})}{2(y-x^2+\sqrt{y^2-xy})^2} & -\frac{x^2[1+(y-x/2)/\sqrt{y^2-xy}]}{(y-x^2+\sqrt{y^2-xy})^2} \\ -\frac{y}{2x\sqrt{y^2-xy}} - \frac{y+\sqrt{y^2-xy}}{x^2} & +\frac{1}{x} + \frac{2y-x}{2x\sqrt{y^2-xy}} \end{pmatrix}.$$

Let $\hat{\theta} = (\hat{\alpha}, \hat{\beta})$ and $\Lambda = \frac{\partial(\alpha,\beta)}{\partial(x,y)}|_{x=\mu_1, y=\mu_2}$. Then, by the δ-method,

$$\sqrt{n}(\hat{\theta} - \theta) \to_d N(0, \Lambda\Sigma\Lambda^\tau). \blacksquare$$

Exercise 51 (#3.108). Let $(X_1, ..., X_n)$ be a random sample from the following discrete distribution:

$$P(X_1 = 1) = \frac{2(1-\theta)}{2-\theta}, \quad P(X_1 = 2) = \frac{\theta}{2-\theta},$$

where $\theta \in (0, 1)$ is unknown. Obtain a moment estimator of θ and its asymptotic distribution.

Solution. Note that

$$EX_1 = \frac{2(1-\theta)}{2-\theta} + \frac{2\theta}{2-\theta} = \frac{2}{2-\theta}.$$

Hence, a moment estimator of θ is $\hat{\theta} = 2(1 - \bar{X}^{-1})$, where $\bar{X}$ is the sample mean. Note that

$$\text{Var}(X_1) = \frac{2(1-\theta)}{2-\theta} + \frac{4\theta}{2-\theta} - \frac{4}{(2-\theta)^2} = \frac{4\theta - 2\theta^2 - 4}{(2-\theta)^2}.$$

By the central limit theorem and δ-method,

$$\sqrt{n}(\hat{\theta} - \theta) \to_d N\left(0, \frac{(2-\theta)^2(2\theta - \theta^2 - 2)}{2}\right). \blacksquare$$

Exercise 52 (#3.110). Let $(X_1, ..., X_n)$ be a random sample from a population having the Lebesgue density

$$f_{\theta_1,\theta_2}(x) = \begin{cases} (\theta_1 + \theta_2)^{-1} e^{-x/\theta_1} & x > 0 \\ (\theta_1 + \theta_2)^{-1} e^{x/\theta_2} & x \leq 0, \end{cases}$$

where $\theta_1 > 0$ and $\theta_2 > 0$ are unknown. Obtain a moment estimator of (θ_1, θ_2) and its asymptotic distribution.

Solution. Let $\mu_j = EX_1^j$ and $\hat{\mu}_j = n^{-1} \sum_{i=1}^n X_i^j$. Note that

$$\mu_1 = \frac{1}{\theta_1 + \theta_2} \left(\int_{-\infty}^0 x e^{x/\theta_2} dx + \int_0^\infty x e^{-x/\theta_1} dx \right) = \theta_1 - \theta_2$$

and

$$\mu_2 = \frac{1}{\theta_1 + \theta_2} \left(\int_{-\infty}^0 x^2 e^{x/\theta_2} dx + \int_0^\infty x^2 e^{-x/\theta_1} dx \right) = 2(\theta_1^2 + \theta_2^2 - \theta_1\theta_2).$$

Then, $\mu_2 - \mu_1^2 = \theta_1^2 + \theta_2^2$. Since $\theta_1 = \mu_1 + \theta_2$, we obtain that

$$2\theta_2^2 + 2\mu_1\theta_2 + 2\mu_1^2 - \mu_2 = 0,$$

which has solutions

$$\frac{-\mu_1 \pm \sqrt{2\mu_2 - 3\mu_1^2}}{2}.$$

Since $\theta_2 > 0$, the moment estimators are

$$\hat{\theta}_2 = \frac{-\hat{\mu}_1 + \sqrt{2\hat{\mu}_2 - 3\hat{\mu}_1^2}}{2}$$

and

$$\hat{\theta}_1 = \frac{\hat{\mu}_1 + \sqrt{2\hat{\mu}_2 - 3\hat{\mu}_1^2}}{2}.$$

Let $g(x, y) = (\sqrt{2y - 3x} - x)/2$ and $h(x, y) = (\sqrt{2y - 3x} + x)/2$. Then

$$\frac{\partial(g, h)}{\partial(x, y)} = \begin{pmatrix} -\frac{1}{2} - \frac{3}{4\sqrt{2y-3x}} & \frac{1}{2\sqrt{2y-3x}} \\ \frac{1}{2} - \frac{3}{4\sqrt{2y-3x}} & \frac{1}{2\sqrt{2y-3x}} \end{pmatrix}.$$

Let $\gamma = (\mu_1, \mu_2)$, $\hat{\gamma} = (\hat{\mu}_1, \hat{\mu}_2)$, $\theta = (\theta_1, \theta_2)$, and $\hat{\theta} = (\hat{\theta}_1, \hat{\theta}_2)$. From the central limit theorem,

$$\sqrt{n}(\hat{\gamma} - \gamma) \to_d N(0, \Sigma),$$

where Σ is as defined in the solution of Exercise 50. By the δ method,

$$\sqrt{n}(\hat{\theta} - \theta) \to_d N(0, \Lambda\Sigma\Lambda^\tau),$$

where $\Lambda = \frac{\partial(g,h)}{\partial(x,y)}|_{x=\mu_1,y=\mu_2}$. ∎

Exercise 53 (#3.111). Let $(X_1,...,X_n)$ be a random sample from P with discrete probability density $f_{\theta,j}$, where $\theta \in (0,1)$, $j = 1,2$, $f_{\theta,1}$ is the Poisson distribution with mean θ, and $f_{\theta,2}$ is the binomial distribution with size 1 and probability θ. Let $h_k(\theta,j) = E_{\theta,j}(X_1^k)$, $k = 1,2$, where $E_{\theta,j}$ is the expectation is with respect to $f_{\theta,j}$. Show that

$$\lim_n P\left(\hat{\mu}_k = h_k(\theta,j) \text{ has a solution}\right) = 0$$

when X_i's are from the Poisson distribution, where $\hat{\mu}_k = n^{-1}\sum_{i=1}^n X_i^k$, $k = 1,2$.

Solution. Note that $h_1(\theta,1) = h_1(\theta,2) = \theta$. Hence $h_1(\theta,j) = \hat{\mu}_1$ has a solution $\theta = \hat{\mu}_1$. Assume that X_i's are from the Poisson distribution with mean θ. Then $\hat{\mu}_2 \to_p \theta + \theta^2$. Since $h_2(\theta,1) = \theta - \theta^2$,

$$\lim_n P\left(\hat{\mu}_2 = h_2(\theta,1)\right) = 0.$$

It remains to show that

$$\lim_n P\left(\hat{\mu}_2 = h_2(\theta,2)\right) = 0.$$

Since $h_2(\theta,2) = \theta + \theta^2$ and $\theta = \hat{\mu}_1$ is a solution to the equation $h_1(\theta,1) = h_1(\theta,2) = \theta$, it suffices to show that

$$\lim_n P\left(\hat{\mu}_2 = \hat{\mu}_1 + \hat{\mu}_1^2\right) = 0.$$

Let $\gamma = (\mu_1,\mu_2)$ and $\hat{\gamma} = (\hat{\mu}_1,\hat{\mu}_2)$. From the central limit theorem,

$$\sqrt{n}(\hat{\gamma} - \gamma) \to_d N(0,\Sigma),$$

where Σ is as defined in the solution of Exercise 50. Then, we only need to show that Σ is not singular. When X_1 has the Poisson distribution with mean θ, a direct calculation shows that $\mu_1 = \theta$, $\mu_2 = \theta+\theta^2$, $\mu_3 = \theta+3\theta^2+\theta^3$, and $\mu_4 = \theta + 7\theta^2 + 6\theta^3 + \theta^4$. Hence,

$$\Sigma = \begin{pmatrix} \theta & \theta + 2\theta^2 \\ \theta + 2\theta^2 & \theta + 6\theta^2 + 4\theta^3 \end{pmatrix}.$$

The determinant of Σ is equal to

$$\theta^2 + 6\theta^3 + 4\theta^4 - (\theta + 2\theta^2)^2 = 2\theta^3 > 0.$$

Hence Σ is not singular. ∎

Exercise 54 (#3.115). Let $X_1,...,X_n$ be a random sample from a population on $\mathcal{R}$ having a finite sixth moment. Consider the estimation of μ^3, where $\mu = EX_1$. Let $\bar{X}$ be the sample mean. When $\mu = 0$, find the asymptotic relative efficiency of the V-statistic $\bar{X}^3$ with respect to the U-statistic $U_n = \binom{n}{3}^{-1} \sum_{1\leq i<j<k\leq n} X_i X_j X_k$.

Solution. We adopt the notation in Exercise 25. Note that U_n is a U-statistic with $\zeta_1 = \zeta_2 = 0$, since $\mu = 0$. The order of the kernel of U_n is 3. Hence, by Exercise 25(iii),

$$\mathrm{Var}(U_n) = \frac{6\zeta_3}{n^3} + O\left(\frac{1}{n^4}\right),$$

where $\zeta_3 = \mathrm{Var}(X_1X_2X_3) = E(X_1^2X_2^2X_3^2) = \sigma^6$ and $\sigma^2 = EX_1^2 = \mathrm{Var}(X_1)$. The asymptotic mean squared error of U_n is then $6\sigma^6/n^3$.

From the central limit theorem and $\mu = 0$, $\sqrt{n}\bar{X} \to_d N(0, \sigma^2)$. Then $n^{3/2}\bar{X}^3/\sigma^3 \to_d Z^3$, where Z is a random variable having distribution $N(0, 1)$. Then the asymptotic mean square error of $\bar{X}^3$ is $\sigma^6 EZ^6/n^3$. Note that $EZ^6 = 15$. Hence, the asymptotic relative efficiency of $\bar{X}^3$ with respect to U_n is $6/15 = 2/5$. ∎

Chapter 4

Estimation in Parametric Models

Exercise 1 (#4.1). Show that the priors in the following cases are conjugate priors:
(i) $X = (X_1, ..., X_n)$ is a random sample from $N_k(\theta, I_k)$, $\theta \in \mathcal{R}^k$, and the prior is $N_k(\mu_0, \Sigma_0)$;
(ii) $X = (X_1, ..., X_n)$ is a random sample from the binomial distribution with probability θ and size k (a known positive integer), $\theta \in (0, 1)$, and the prior is the beta distribution with parameter (α, β);
(iii) $X = (X_1, ..., X_n)$ is a random sample from the uniform distribution on the interval $(0, \theta)$, $\theta > 0$, and the prior has Lebesgue density $ba^b\theta^{-(b+1)}I_{(a,\infty)}(\theta)$;
(iv) $X = (X_1, ..., X_n)$ is a random sample from the exponential distribution with Lebesgue density $\theta^{-1}e^{-x/\theta}I_{(0,\infty)}(x)$, $\theta > 0$, and the prior of θ^{-1} is the gamma distribution with shape parameter α and scale parameter γ.

Solution. (i) Let $T = \sum_{i=1}^n X_i$ and $A = nI_k + \Sigma_0^{-1}$. The product of the density of X and the prior density is

$$C_X \exp\left\{-\frac{\|T - n\theta\|^2}{2} - \frac{(\theta - \mu_0)^\tau \Sigma_0^{-1}(\theta - \mu_0)}{2}\right\}$$

$$= D_X \exp\left\{-\frac{\left[\theta - A^{-1}(\Sigma_0^{-1}\mu_0 + T)\right]^\tau A \left[\theta - A^{-1}(\Sigma_0^{-1}\mu_0 + T)\right]}{2}\right\},$$

where C_X and D_X are quantities depending on X but not θ. Thus, the posterior distribution of θ given X is $N_k\left(A^{-1}(\Sigma_0^{-1}\mu_0 + T), A^{-1}\right)$.
(ii) Let $T = \sum_{i=1}^n X_i$. The product of the density of X and the prior density is

$$C_X\theta^{T+\alpha-1}(1-\theta)^{nk-T+\beta-1},$$

where C_X does not depend on θ. Thus, the posterior distribution of θ given X is the beta distribution with parameter $(T+\alpha, nk-T+\beta)$.
(iii) Let $X_{(n)}$ be the largest order statistic. The product of the density of X and the prior density is

$$\theta^{-n} I_{(0,\theta)}(X_{(n)}) b a^b \theta^{-(b+1)} I_{(a,\infty)}(\theta) = \theta^{-(n+b+1)} I_{(\max\{X_{(n)},a\},\infty)}(\theta).$$

Thus, the posterior distribution of θ given X has the same form as the prior with a replaced by $\max\{X_{(n)}, a\}$ and b replaced by $b+n$.
(iv) Let $T = \sum_{i=1}^n X_i$. The product of the density of X and the prior density is

$$C_X \theta^{-(n+\alpha+1)} \exp\left\{-(T+\gamma^{-1})/\theta\right\},$$

where C_X does not depend on θ. Thus, the posterior distribution of θ^{-1} given X is the gamma distribution with shape parameter $n+\alpha$ and scale parameter $(T+\gamma^{-1})^{-1}$. ∎

Exercise 2 (#4.2). In Exercise 1, find the posterior mean and variance for each case.
Solution. (i) Since the posterior is a normal distribution,

$$E(\theta|X) = (\Sigma_0^{-1}\mu_0 + T)A^{-1}$$

and

$$\mathrm{Var}(\theta|X) = A^{-1},$$

where $T = \sum_{i=1}^n X_i$ and $A = nI_k + \Sigma_0^{-1}$.
(ii) Since the posterior is a beta distribution,

$$E(\theta|X) = \frac{T+\alpha}{nk+\alpha+\beta}$$

and

$$\mathrm{Var}(\theta|X) = \frac{(T+\alpha)(nk-T+\beta)}{(nk+\alpha+\beta)^2(nk+\alpha+\beta+1)},$$

where $T = \sum_{i=1}^n X_i$.
(iii) A direct calculation shows that

$$E(\theta|X) = \frac{\max\{X_{(n)}, a\}(b+n)}{(b+n-1)}$$

and

$$\mathrm{Var}(\theta|X) = \frac{\max\{X_{(n)}^2, a^2\}(b+n)}{(b+n-1)^2(b+n-2)}.$$

(iv) Let $T = \sum_{i=1}^n X_i$. Then

$$E(\theta|X) = \frac{T+\gamma^{-1}}{\alpha+n}$$

and

$$\text{Var}(\theta|X) = \frac{(T+\gamma^{-1})^2}{(n+\alpha-1)(n+\alpha-2)} - \frac{(T+\gamma^{-1})^2}{(n+\alpha)^2}. \ \blacksquare$$

Exercise 3 (#4.4). Let $X = (X_1, ..., X_n)$ be a random sample from the uniform distribution on the interval $(0, \theta)$, where $\theta > 0$ is unknown. Let the prior of θ be the log-normal distribution with parameter (μ_0, σ_0^2), where $\mu_0 \in \mathcal{R}$ and $\sigma_0 > 0$ are known constants.
(i) Find the posterior density of $\log\theta$.
(ii) Find the rth posterior moment of θ.
(iii) Find a value that maximizes the posterior density of θ.
Solution. (i) Let $X_{(n)}$ be the largest order statistic. The product of the density of X and the prior density is proportional to

$$\frac{1}{\theta^{n+1}} \exp\left\{-\frac{(\log\theta - \mu_0)^2}{2\sigma_0^2}\right\} I_{(X_{(n)},\infty)}(\theta).$$

Then the posterior density of $\vartheta = \log\theta$ given X is

$$\frac{1}{\sqrt{2\pi}\sigma_0 C_X} \exp\left\{-\frac{(\vartheta - \mu_0 + n\sigma_0^2)^2}{2\sigma_0^2}\right\} I_{(\log X_{(n)},\infty)}(\vartheta),$$

where

$$C_X = \Phi\left(\frac{\mu_0 - n\sigma_0^2 - \log X_{(n)}}{\sigma_0}\right)$$

and Φ is the cumulative distribution function of the standard normal distribution.
(ii) Note that $E(\theta^r|X) = E\left(e^{r\log\theta}|X\right)$ and $\log\theta$ given X has a truncated normal distribution as specified in part (i) of the solution. Therefore,

$$E(\theta^r|X) = C_X^{-1} e^{r[2\mu_0 - (2n-r)\sigma_0^2]/2} \Phi\left(\frac{\mu_0 - (n-r)\sigma_0^2 - \log X_{(n)}}{\sigma_0}\right).$$

(iii) From part (i) of the solution, the posterior density of θ given X is

$$\frac{1}{\sqrt{2\pi}\sigma_0 C_X \theta} \exp\left\{-\frac{(\log\theta - \mu_0 + n\sigma_0^2)^2}{2\sigma_0^2}\right\} I_{(X_{(n)},\infty)}(\theta).$$

Without the indicator function $I_{(X_{(n)},\infty)}(\theta)$, the above function has a unique maximum at $e^{\mu_0 - (n+1)\sigma_0^2}$. Therefore, the posterior of θ given X is maximized at $\max\{e^{\mu_0 - (n+1)\sigma_0^2}, X_{(n)}\}$. $\blacksquare$

Exercise 4 (#4.6). Let $\bar{X}$ be the sample mean of a random sample of size n from $N(\theta, \sigma^2)$ with a known $\sigma > 0$ and an unknown $\theta \in \mathcal{R}$. Let $\pi(\theta)$

be a prior density with respect to a σ-finite measure ν on $\mathcal{R}$.
(i) Show that the posterior mean of θ, given $\bar{X} = x$, is of the form

$$\delta(x) = x + \frac{\sigma^2}{n} \frac{d \log(p(x))}{dx},$$

where $p(x)$ is the marginal density of $\bar{X}$, unconditional on θ.
(ii) Express the posterior variance of θ (given $\bar{X} = x$) as a function of the first two derivatives of $\log p(x)$.
(iii) Find explicit expressions for $p(x)$ and $\delta(x)$ in (i) when the prior is $N(\mu_0, \sigma_0^2)$ with probability $1 - \epsilon$ and a point mass at μ_1 with probability ϵ, where μ_0, μ_1, and σ_0^2 are known constants.
Solution. (i) Note that $\bar{X}$ has distribution $N(\theta, \sigma^2/n)$. The product of the density of $\bar{X}$ and $\pi(\theta)$ is

$$\frac{\sqrt{n}}{\sqrt{2\pi}\sigma} e^{-n(x-\theta)^2/(2\sigma^2)} \pi(\theta).$$

Hence,

$$p(x) = \int \frac{\sqrt{n}}{\sqrt{2\pi}\sigma} e^{-n(x-\theta)^2/(2\sigma^2)} \pi(\theta) d\nu$$

and

$$p'(x) = \frac{n}{\sigma^2} \int \frac{\sqrt{n}}{\sqrt{2\pi}\sigma} (\theta - x) e^{-n(x-\theta)^2/(2\sigma^2)} \pi(\theta) d\nu.$$

Then, the posterior mean is

$$\begin{aligned}
\delta(x) &= \frac{1}{p(x)} \int \frac{\sqrt{n}}{\sqrt{2\pi}\sigma} \theta e^{-n(x-\theta)^2/(2\sigma^2)} \pi(\theta) d\nu \\
&= x + \frac{1}{p(x)} \int \frac{\sqrt{n}}{\sqrt{2\pi}\sigma} (\theta - x) e^{-n(x-\theta)^2/(2\sigma^2)} \pi(\theta) d\nu \\
&= x + \frac{\sigma^2}{n} \frac{p'(x)}{p(x)} \\
&= x + \frac{\sigma^2}{n} \frac{d \log(p(x))}{dx}.
\end{aligned}$$

(ii) From the result in part (i) of the solution,

$$p''(x) = \frac{n^2}{\sigma^4} \int \frac{\sqrt{n}}{\sqrt{2\pi}\sigma} (\theta - x)^2 e^{-n(x-\theta)^2/(2\sigma^2)} \pi(\theta) d\nu - \frac{n}{\sigma^2} p(x).$$

Hence,

$$E[(\theta - x)^2 | \bar{X} = x] = \frac{\sigma^4}{n^2} \frac{p''(x)}{p(x)} + \frac{\sigma^2}{n}$$

and, therefore,

$$\begin{aligned}\text{Var}(\theta|\bar{X}=x) &= E[(\theta-x)^2|\bar{X}=x] - \left[E(\theta-x|\bar{X}=x)\right]^2 \\ &= \frac{\sigma^4}{n^2}\frac{p''(x)}{p(x)} + \frac{\sigma^2}{n} - \left[\frac{\sigma^2}{n}\frac{p'(x)}{p(x)}\right]^2 \\ &= \frac{\sigma^4}{n^2}\frac{d^2\log p(x)}{dx^2} + \frac{\sigma^2}{n}.\end{aligned}$$

(iii) If the prior is $N(\mu_0,\sigma_0^2)$, then the joint distribution of θ and $\bar{X}$ is normal and, hence, the marginal distribution of $\bar{X}$ is normal. The mean of $\bar{X}$ conditional on θ is θ. Hence the marginal mean of $\bar{X}$ is μ_0. The variance of $\bar{X}$ conditional on θ is σ^2/n. Hence the marginal variance of $\bar{X}$ is $\sigma_0^2+\sigma^2/n$. Thus, $p(x)$ is the density of $N(\mu_0,\sigma_0^2+\sigma^2/n)$ if the prior is $N(\mu_0,\sigma_0^2)$. If the prior is a point mass at μ_1, then

$$p(x) = \frac{\sqrt{n}}{\sqrt{2\pi}\sigma}e^{-n(x-\mu_1)^2/(2\sigma^2)},$$

which is the density of $N(\mu_1,\sigma^2/n)$. Therefore, $p(x)$ is the density of the mixture distribution $(1-\epsilon)N(\mu_0,\sigma_0^2+\sigma^2/n)+\epsilon N(\mu_1,\sigma^2/n)$, i.e.,

$$p(x) = (1-\epsilon)\phi\left(\frac{x-\mu_0}{\sqrt{\sigma_0^2+\sigma^2/n}}\right) + \epsilon\phi\left(\frac{x-\mu_1}{\sqrt{\sigma^2/n}}\right),$$

where $\phi(x)=e^{-x^2/2}/\sqrt{2\pi}$. Then

$$p'(x) = \frac{(1-\epsilon)(\mu_0-x)}{\sigma_0^2+\sigma^2/n}\phi\left(\frac{x-\mu_0}{\sqrt{\sigma_0^2+\sigma^2/n}}\right) + \frac{\epsilon(\mu_1-x)}{\sigma^2/n}\phi\left(\frac{x-\mu_1}{\sqrt{\sigma^2/n}}\right)$$

and $\delta(x)$ can be obtained using the formula in (i). ∎

Exercise 5 (#4.8). Let $X=(X_1,...,X_n)$ be a random sample from P with discrete probability density $f_{\theta,j}$, where $\theta\in(0,1)$, $j=1,2$, $f_{\theta,1}$ is the Poisson distribution with mean θ, and $f_{\theta,2}$ is the binomial distribution with size 1 and probability θ. Consider the estimation of θ under the squared error loss. Suppose that the prior of θ is the uniform distribution on $(0,1)$, the prior of j is $P(j=1)=P(j=2)=\frac{1}{2}$, and the joint prior of (θ,j) is the product probability of the two marginal priors. Show that the Bayes action is

$$\delta(x) = \frac{H(x)B(t+1)+G(t+1)}{H(x)B(t)+G(t)},$$

where $x=(x_1,...,x_n)$ is the vector of observations, $t=x_1+\cdots+x_n$, $B(t)=\int_0^1\theta^t(1-\theta)^{n-t}d\theta$, $G(t)=\int_0^1\theta^te^{-n\theta}d\theta$, and $H(x)$ is a function of x

with range $\{0,1\}$.

Note. Under the squared error loss, the Bayes action in an estimation problem is the posterior mean.

Solution. The marginal density is

$$m(x) = \frac{C(x)}{2}\int_0^1 e^{-n\theta}\theta^t d\theta + \frac{D(x)}{2}\int_0^1 \theta^t(1-\theta)^{n-t}d\theta$$
$$= \frac{C(x)G(t)+D(x)B(t)}{2},$$

where $C(x)=(x_1!\cdots x_n!)^{-1}$ and $D(x)=1$ if all components of x are 0 or 1 and is 0 otherwise. Then the Bayes action is

$$\delta(x) = \frac{C(x)\int_0^1 e^{-n\theta}\theta^{t+1}d\theta + D(x)\int_0^1 \theta^{t+1}(1-\theta)^{n-t}d\theta}{2m(x)}$$
$$= \frac{H(x)B(t+1)+G(t+1)}{H(x)B(t)+G(t)},$$

where $H(x)=D(x)/C(x)$ takes value 0 or 1. ∎

Exercise 6 (#4.10). Let X be a sample from P_θ, $\theta\in\Theta\subset\mathcal{R}$. Consider the estimation of θ under the loss $L(|\theta-a|)$, where L is an increasing function on $[0,\infty)$. Let $\pi(\theta|x)$ be the posterior density (with respect to Lebesgue measure) of θ given $X=x$. Suppose that $\pi(\theta|x)$ is symmetric about $\delta(x)\in\Theta$ and that $\pi(\theta|x)$ is nondecreasing for $\theta\le\delta(x)$ and nonincreasing for $\theta\ge\delta(x)$. Show that $\delta(x)$ is a Bayes action, assuming that all integrals involved are finite.

Solution. Without loss of generality, assume that $\delta(x)=0$. Then $\pi(\theta|x)$ is symmetric about 0. Hence, the posterior expected loss for any action a is

$$\rho(a) = \int L(|\theta-a|)\pi(\theta|x)d\theta$$
$$= \int L(|-\theta-a|)\pi(\theta|x)d\theta$$
$$= \rho(-a).$$

For any $a\ge 0$ and θ, define

$$H(\theta,a) = [L(|\theta+a|)-L(|\theta-a|)][\pi(\theta+a|x)-\pi(\theta-a|x)].$$

If $\theta+a\ge 0$ and $\theta-a\ge 0$, then $L(|\theta+a|)\ge L(|\theta-a|)$ and $\pi(\theta+a|x)\le\pi(\theta-a|x)$; if $\theta+a\le 0$ and $\theta-a\le 0$, then $L(|\theta+a|)\le L(|\theta-a|)$ and $\pi(\theta+a|x)\ge\pi(\theta-a|x)$; if $\theta-a\le 0\le\theta+a$, then $\pi(\theta+a|x)\le\pi(\theta-a|x)$ and $L(|\theta+a|)\ge L(|\theta-a|)$ when $\theta\ge 0$ and $\pi(\theta+a|x)\ge\pi(\theta-a|x)$ and

$L(|\theta+a|) \leq L(|\theta-a|)$ when $\theta \leq 0$. This shows that $H(\theta, a) \leq 0$ for any θ and $a \geq 0$. Then, for any $a \geq 0$,

$$\begin{aligned}
0 &\geq \int H(\theta,a)d\theta \\
&= \int L(|\theta+a|)\pi(\theta+a|x)d\theta + \int L(|\theta-a|)\pi(\theta-a|x)d\theta \\
&\quad - \int L(|\theta+a|)\pi(\theta-a|x)d\theta - \int L(|\theta-a|)\pi(\theta+a|x)d\theta \\
&= 2\int L(|\theta|)\pi(\theta|x)d\theta - \int L(|\theta+2a|)\pi(\theta|x)d\theta \\
&\quad - \int L(|\theta-2a|)\pi(\theta|x)d\theta \\
&= 2\rho(0) - \rho(2a) - \rho(-2a) \\
&= 2\rho(0) - 2\rho(2a).
\end{aligned}$$

This means that $\rho(0) \leq \rho(2a) = \rho(-2a)$ for any $a \geq 0$, which proves that 0 is a Bayes action. ∎

Exercise 7 (#4.11). Let X be a sample of size 1 from the geometric distribution with mean p^{-1}, where $p \in (0,1]$. Consider the estimation of p with the loss function $L(p,a) = (p-a)^2/p$.
(i) Show that δ is a Bayes action with a prior Π if and only if $\delta(x) = 1 - \int(1-p)^x d\Pi(p)/\int(1-p)^{x-1}d\Pi(p)$, $x = 1, 2, ...$
(ii) Let δ_0 be a rule such that $\delta_0(1) = 1/2$ and $\delta_0(x) = 0$ for all $x > 1$. Show that δ_0 is a limit of Bayes actions.
(iii) Let δ_0 be a rule such that $\delta_0(x) = 0$ for all $x > 1$ and $\delta_0(1)$ is arbitrary. Show that δ_0 is a generalized Bayes action.
Note. In estimating $g(\theta)$ under the family of densities $\{f_\theta : \theta \in \Theta\}$ and the loss function $w(\theta)[g(\theta)-a]^2$, where $\Theta \subset \mathcal{R}$, $w(\theta) \geq 0$ and $\int_\Theta w(\theta)[g(\theta)]^2 d\Pi < \infty$, the Bayes action is

$$\delta(x) = \frac{\int_\Theta w(\theta)g(\theta)f_\theta(x)d\Pi}{\int_\Theta w(\theta)f_\theta(x)d\Pi}.$$

Solution. (i) The discrete probability density of X is $(1-p)^{x-1}p$ for $x = 1, 2, ...$ Hence, for estimating p with loss $(p-a)^2/p$, the Bayes action when $X = x$ is

$$\delta(x) = \frac{\int_0^1 p^{-1}p(1-p)^{x-1}pd\Pi}{\int_0^1 p^{-1}(1-p)^{x-1}pd\Pi} = 1 - \frac{\int_0^1 (1-p)^x d\Pi}{\int_0^1 (1-p)^{x-1}d\Pi}.$$

(ii) Consider the prior with Lebesgue density $\frac{\Gamma(2\alpha)}{[\Gamma(\alpha)]^2}p^{\alpha-1}(1-p)^{\alpha-1}I_{(0,1)}(p)$.

The Bayes action is

$$\begin{aligned}\delta(x) &= 1 - \frac{\int_0^1 (1-p)^{x+\alpha-1} p^{\alpha-1} dp}{\int_0^1 (1-p)^{x+\alpha-2} p^{\alpha-1} dp} \\ &= 1 - \frac{\frac{\Gamma(x+2\alpha-1)}{\Gamma(x+\alpha-1)\Gamma(\alpha)}}{\frac{\Gamma(x+2\alpha)}{\Gamma(x+\alpha)\Gamma(\alpha)}} \\ &= 1 - \frac{x+\alpha-1}{x+2\alpha-1} \\ &= \frac{\alpha}{x+2\alpha-1}.\end{aligned}$$

Since

$$\lim_{\alpha\to 0} \delta(x) = \frac{1}{2} I_{\{1\}}(x) = \delta_0(x),$$

$\delta_0(x)$ is a limit of Bayes actions.
(iii) Consider the improper prior density $\frac{d\Pi}{dp} = [p^2(1-p)]^{-1}$. Then the posterior risk for action a is

$$\int_0^1 (p-a)^2 (1-p)^{x-2} p^{-2} dp.$$

When $x = 1$, the above integral diverges to infinity and, therefore, any a is a Bayes action. When $x > 1$, the above integral converges if and only if $a = 0$. Hence δ_0 is a Bayes action. ∎

Exercise 8 (#4.13). Let X be a sample from P_θ having probability density $f_\theta(x) = h(x)\exp\{\theta^\tau x - \zeta(\theta)\}$ with respect to ν on $\mathcal{R}^p$, where $\theta \in \mathcal{R}^p$. Let the prior be the Lebesgue measure on $\mathcal{R}^p$. Show that the generalized Bayes action under the loss $L(\theta, a) = \|E(X) - a\|^2$ is $\delta(x) = x$ when $X = x$ with $\int f_\theta(x)d\theta < \infty$.
Solution. Let $m(x) = \int f_\theta(x)d\theta$ and $\mu(\theta) = E(X)$. Similar to the case of univariate θ, the generalized Bayes action under loss $\|\mu(\theta) - a\|^2$ is $\delta(x) = \int \mu(\theta) f_\theta(x)d\theta/m(x)$. Let $A_c = (-\infty, c_1]\times\cdots\times(-\infty, c_p]$ for $c = (c_1, ..., c_p) \in \mathcal{R}^p$. Note that $f_c(x) = \int_{A_c} \frac{\partial f_\theta(x)}{\partial\theta} d\theta$, $c \in \mathcal{R}^p$. Since $m(x) = \int f_c(x)dc < \infty$, $\lim_{c_i\to\infty, i=1,...,p} f_c(x) = 0$. Hence, $\int \frac{\partial f_\theta(x)}{\partial\theta} d\theta = 0$. Since

$$\frac{\partial f_\theta(x)}{\partial\theta} = \left[x - \frac{\partial\zeta(\theta)}{\partial\theta}\right] f_\theta(x) = [x - \mu(\theta)]\, f_\theta(x),$$

we obtain that $x\int f_\theta(x)d\theta = \int \mu(\theta) f_\theta(x)d\theta$. This proves that $\delta(x) = x$. ∎

Exercise 9 (#4.14). Let $(X_1, ..., X_n)$ be a random sample of random variables with the Lebesgue density $\sqrt{2/\pi} e^{-(x-\theta)^2/2} I_{(\theta,\infty)}(x)$, where $\theta \in \mathcal{R}$

is unknown. Find the generalized Bayes action for estimating θ under the squared error loss, when the (improper) prior of θ is the Lebesgue measure on $\mathcal{R}$.

Solution. Let $\bar{X}$ be the sample mean and $X_{(1)}$ be the smallest order statistic. Then the product of the density of $X_1, ..., X_n$ and the prior density is

$$\left(\frac{2}{\pi}\right)^{n/2} \exp\left\{-\frac{1}{2}\sum_{i=1}^{n}(X_i-\theta)^2\right\} I_{(\theta,\infty)}(X_{(1)})$$

and, thus, the generalized Bayes action is

$$\delta = \frac{\int_0^{X_{(1)}} \theta e^{-n(\bar{X}-\theta)^2/2} d\theta}{\int_0^{X_{(1)}} e^{-n(\bar{X}-\theta)^2/2} d\theta} = \bar{X} + \frac{\int_0^{X_{(1)}} (\theta-\bar{X}) e^{-n(\bar{X}-\theta)^2/2} d\theta}{\int_0^{X_{(1)}} e^{-n(\bar{X}-\theta)^2/2} d\theta}.$$

Let Φ be the cumulative distribution function of the standard normal distribution. Then

$$\int_0^{X_{(1)}} e^{-n(\bar{X}-\theta)^2/2} d\theta = \frac{\sqrt{2\pi}[\Phi(\sqrt{n}(X_{(1)}-\bar{X})) - \Phi(-\sqrt{n}\bar{X})]}{\sqrt{n}}$$

and

$$\int_0^{X_{(1)}} (\theta-\bar{X}) e^{-n(\bar{X}-\theta)^2/2} d\theta = \frac{\sqrt{2\pi}[\Phi'(-\sqrt{n}\bar{X}) - \Phi'(\sqrt{n}(X_{(1)}-\bar{X}))]}{n}.$$

Hence, the generalized Bayes action is

$$\delta = \bar{X} + \frac{\Phi'(-\sqrt{n}\bar{X}) - \Phi'(\sqrt{n}(X_{(1)}-\bar{X}))}{\sqrt{n}[\Phi(\sqrt{n}(X_{(1)}-\bar{X})) - \Phi(-\sqrt{n}\bar{X})]}. \quad \blacksquare$$

Exercise 10 (#4.15). Let $(X_1, ..., X_n)$ be a random sample from $N(\mu, \sigma^2)$ and $\pi(\mu, \sigma^2) = \sigma^{-2} I_{(0,\infty)}(\sigma^2)$ be an improper prior for (μ, σ^2) with respect to the Lebesgue measure on $\mathcal{R}^2$.

(i) Show that the posterior density of (μ, σ^2) given $x = (x_1, ..., x_n)$ is $\pi(\mu, \sigma^2|x) = \pi_1(\mu|\sigma^2, x)\pi_2(\sigma^2|x)$, where $\pi_1(\mu|\sigma^2, x)$ is the density of the normal distribution $N(\bar{x}, \sigma^2/n)$, $\bar{x}$ is the sample mean of x_i's, $\pi_2(\sigma^2|x)$ is the density of ω^{-1}, and ω has the gamma distribution with shape parameter $(n-1)/2$ and scale parameter $[\sum_{i=1}^n (x_i-\bar{x})^2/2]^{-1}$.

(ii) Show that the marginal posterior density of μ given x is $f\left(\frac{\mu-\bar{x}}{\tau}\right)$, where $\tau^2 = \sum_{i=1}^n (x_i-\bar{x})^2/[n(n-1)]$ and f is the density of the t-distribution t_{n-1}.

(iii) Obtain the generalized Bayes action for estimating μ/σ under the squared error loss.

Solution. (i) The posterior density $\pi(\mu, \sigma^2|x)$ is proportional to

$$\sigma^{-(n+2)} \exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\bar{x})^2\right\} \exp\left\{-\frac{(\mu-\bar{x})^2}{2\sigma^2/n}\right\} I_{(0,\infty)}(\sigma^2),$$

which is proportional to $\pi_1(\mu|\sigma^2,x)\pi_2(\sigma^2|x)$.
(ii) The marginal posterior density of μ is

$$\begin{aligned}\pi(\mu|x) &= \int_0^\infty \pi(\mu,\sigma^2|x)d\sigma^2 \\ &\propto \int_0^\infty \sigma^{-(n+2)}\exp\left\{-\frac{n(\bar{x}-\mu)^2}{2\sigma^2}-\frac{1}{2\sigma^2}\sum_{i=1}^n (x_i-\bar{x})^2\right\}d\sigma^2 \\ &\propto \left[1+\frac{1}{n-1}\left(\frac{\mu-\bar{x}}{\tau}\right)^2\right]^{-n/2}.\end{aligned}$$

Hence, $\pi(\mu|x)$ is $f(\frac{\mu-\bar{x}}{\tau})$ with f being the density of the t-distribution t_{n-1}.
(iii) The generalized Bayes action is

$$\begin{aligned}\delta &= \int \frac{\mu}{\sigma}\pi_1(\mu|\sigma^2,x)\pi_2(\sigma^2|x)d\mu d\sigma^2 \\ &= \bar{x}\int \sigma^{-1}\pi_2(\sigma^2|x)d\sigma^2 \\ &= \frac{\Gamma(n/2)\bar{x}}{\Gamma((n-1)/2)\sqrt{\sum_{i=1}^n (x_i-\bar{x})^2/2}}. \quad \blacksquare\end{aligned}$$

Exercise 11 (#4.19). In (ii)-(iv) of Exercise 1, assume that the parameters in priors are unknown. Using the method of moments, find empirical Bayes actions under the squared error loss.
Solution. Define $\hat{\mu}_1 = \bar{X}$ (the sample mean) and $\hat{\mu}_2 = n^{-1}\sum_{i=1}^n X_i^2$.
(i) In Exercise 1(ii),

$$EX_1 = E[E(X_1|p)] = E(kp) = \frac{k\alpha}{\alpha+\beta}$$

and

$$EX_1^2 = E[E(X_1^2|p)] = E[kp(1-p)+k^2p^2] = \frac{k\alpha}{\alpha+\beta} + \frac{(k^2-k)\alpha(\alpha+1)}{(\alpha+\beta)(\alpha+\beta+1)}.$$

Setting $\hat{\mu}_1 = EX_1$ and $\hat{\mu}_2 = EX_1^2$, we obtain that

$$\hat{\alpha} = \frac{\hat{\mu}_2-\hat{\mu}_1-\hat{\mu}_1(k-1)}{\hat{\mu}_1(k-1)+k(1-\hat{\mu}_2/\hat{\mu}_1)}$$

and

$$\hat{\beta} = \frac{k\hat{\alpha}}{\hat{\mu}_1} - \hat{\alpha}.$$

Then the empirical Bayes action is $(n\bar{X}+\hat{\alpha})/(kn+\hat{\alpha}+\hat{\beta})$.
(ii) In Exercise 1(iii),

$$EX_1 = E[E(X_1|\theta)] = E(\theta/2) = \frac{ab}{2(b-1)}$$

and

$$EX_1^2 = E[E(X_1^2|\theta)] = E[(\theta/2)^2 + \theta^2/12] = E(\theta^2/3) = \frac{a^2 b}{3(b-2)}.$$

Setting $\hat{\mu}_1 = EX_1$ and $\hat{\mu}_2 = EX_1^2$, we obtain that

$$\hat{b} = 1 + \sqrt{3\hat{\mu}_2/(3\hat{\mu}_2 - 4\hat{\mu}_1^2)}$$

and

$$\hat{a} = 2\hat{\mu}_1(\hat{b}-1)/\hat{b}.$$

Therefore, the empirical Bayes action is $(n+\hat{b})\max\{X_{(n)},\hat{a}\}/(n+\hat{b}-1)$, where $X_{(n)}$ is the largest order statistic.
(iii) In Exercise 1(iv),

$$EX_1 = E[E(X_1|\theta)] = E(\theta) = \frac{1}{\gamma(\alpha-1)}$$

and

$$EX_1^2 = E[E(X_1^2|\theta)] = E(2\theta^2) = \frac{2}{\gamma^2(\alpha-1)(\alpha-2)}.$$

Setting $\hat{\mu}_1 = EX_1$ and $\hat{\mu}_2 = EX_1^2$, we obtain that

$$\hat{\alpha} = \frac{2\hat{\mu}_2 - 2\hat{\mu}_1^2}{\hat{\mu}_2 - 2\hat{\mu}_1^2}$$

and

$$\hat{\gamma} = \frac{1}{(\hat{\alpha}-1)\hat{\mu}_1}.$$

The empirical Bayes action is $(\hat{\gamma}n\bar{X}+1)/[\hat{\gamma}(n+\hat{\alpha}-1)]$. ∎

Exercise 12 (#4.20). Let $X = (X_1, ..., X_n)$ be a random sample from $N(\mu, \sigma^2)$ with an unknown $\mu \in \mathcal{R}$ and a known $\sigma^2 > 0$. Consider the prior $\Pi_{\mu|\xi} = N(\mu_0, \sigma_0^2)$, $\xi = (\mu_0, \sigma_0^2)$, and the second-stage improper joint prior for ξ be the product of $N(a, v^2)$ and the Lebesgue measure on $(0,\infty)$, where a and v are known. Under the squared error loss, obtain a formula for the generalized Bayes action in terms of a one-dimensional integral.

Solution. Let $\bar{x}$ be the observed sample mean. From Exercise 2, the Bayes action when ξ is known is

$$\delta(x,\xi) = \frac{\sigma^2}{n\sigma_0^2+\sigma^2}\mu_0 + \frac{n\sigma_0^2}{n\sigma_0^2+\sigma^2}\bar{x}.$$

By formula (4.8) in Shao (2003), the Bayes action is

$$\int \delta(x,\xi) f(\xi|x) d\xi,$$

where $f(\xi|x)$ is the conditional density of ξ given $X = x$. The joint density of (X,μ,ξ) is

$$\left(\frac{1}{\sqrt{2\pi}\sigma}\right)^n \frac{1}{2\pi\sigma_0 v} \exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^n (x_i-\mu)^2 - \frac{(\mu-\mu_0)^2}{2\sigma_0^2} - \frac{(\mu_0-\mu)^2}{2v^2}\right\}.$$

Integrating out μ in the joint density of (X,μ,ξ) and using the identity

$$\int_{-\infty}^{\infty} \exp\left\{-\frac{at^2-2bt+c}{2}\right\} dt = \sqrt{\frac{2\pi}{a}} \exp\left\{\frac{b^2}{2a} - \frac{c}{2}\right\}$$

for any $a > 0$ and real b and c, we obtain the joint density of (X,ξ) as

$$\frac{d}{\sigma_0\sqrt{\frac{n}{\sigma^2}+\frac{1}{\sigma_0^2}}} \exp\left\{-\frac{y}{2\sigma^2} - \frac{(\mu_0-a)^2}{2v^2} - \frac{\mu_0^2}{2\sigma_0^2} + \frac{\left(\frac{n\bar{x}}{\sigma^2}+\frac{\mu_0}{\sigma_0^2}\right)^2}{2\left(\frac{n}{\sigma^2}+\frac{1}{\sigma_0^2}\right)}\right\},$$

where y is the observed value of $\sum_{i=1}^n X_i^2$ and $d = (2\pi)^{-(n+1)/2}\sigma^{-n}v^{-1}$. This implies that

$$E(\mu_0|\sigma_0^2,x) = \frac{a\sigma_0^2(n\sigma_0^2+\sigma^2) - n\sigma_0^2 v^2\bar{x}}{\sigma_0^2(n\sigma_0^2+\sigma^2)+n\sigma_0^2 v^2}.$$

Integrating out μ_0 in the joint density of (X,ξ) and using the previous identity again yields the joint density of (X,σ_0^2) as

$$f(x,\sigma_0^2) = \frac{de^{-y/(2\sigma^2)}}{\sigma_0\sqrt{\frac{n}{\sigma^2}+\frac{1}{\sigma_0^2}}\sqrt{\frac{1}{v^2}+\frac{1}{\sigma_0^2}-\frac{\frac{1}{\sigma_0^4}}{\frac{n}{\sigma^2}+\frac{1}{\sigma_0^2}}}}$$

$$\times \exp\left\{\left(\frac{a}{v^2} - \frac{\frac{n\bar{x}}{\sigma_0^2\sigma^2}}{\frac{n}{\sigma^2}+\frac{1}{\sigma_0^2}}\right)^2 \Bigg/ \left[2\left(\frac{1}{v^2}+\frac{1}{\sigma_0^2}-\frac{\frac{1}{\sigma_0^4}}{\frac{n}{\sigma^2}+\frac{1}{\sigma_0^2}}\right)\right] - \frac{\frac{n^2\bar{x}^2}{2\sigma^4}}{\frac{n}{\sigma^2}+\frac{1}{\sigma_0^2}}\right\}$$

Then the generalized Bayes action is

$$\int \delta(x,\xi) f(\xi|x) d\xi = \frac{\int_0^\infty \left[\frac{\sigma^2 E(\mu_0|\sigma_0^2,x)}{n\sigma_0^2+\sigma^2} + \frac{n\sigma_0^2}{n\sigma_0^2+\sigma^2}\bar{x} \right] f(\sigma_0^2,x) d\sigma_0^2}{\int_0^\infty f(\sigma_0^2,x) d\sigma_0^2}. \blacksquare$$

Exercise 13 (#4.21). Let $X = (X_1, ..., X_n)$ be a random sample from the uniform distribution on $(0, \theta)$, where $\theta > 0$ is unknown. Let $\pi(\theta) = ba^b\theta^{-(b+1)} I_{(a,\infty)}(\theta)$ be a prior density with respect to the Lebesgue measure, where $b > 1$ is known but $a > 0$ is an unknown hyperparameter. Consider the estimation of θ under the squared error loss.
(i) Show that the empirical Bayes method using the method of moments produces the empirical Bayes action $\delta(\hat{a})$, where $\delta(a) = \frac{b+n}{b+n-1} \max\{a, X_{(n)}\}$, $\hat{a} = \frac{2(b-1)}{bn} \sum_{i=1}^n X_i$, and $X_{(n)}$ is the largest order statistic.
(ii) Let $h(a) = a^{-1} I_{(0,\infty)}(a)$ be an improper Lebesgue prior density for a. Obtain explicitly the generalized Bayes action.
Solution. (i) Note that $EX_1 = E[E(X_1|\theta)] = E(\theta/2) = ab/[2(b-1)]$. Then $\hat{a} = \frac{2(b-1)}{bn} \sum_{i=1}^n X_i$ is the moment estimator of a. From Exercise 2, the empirical Bayes action is $\delta(\hat{a})$.
(ii) The joint density for (X, θ, a) is

$$ba^{b-1}\theta^{-(n+b+1)} I_{(X_{(n)},\infty)}(\theta) I_{(0,\theta)}(a).$$

Hence, the joint density for (X, θ) is

$$\int_0^\theta ba^{b-1}\theta^{-n-(b+1)} I_{(X_{(n)},\infty)}(\theta) da = \theta^{-(n+1)} I_{(X_{(n)},\infty)}(\theta)$$

and the generalized Bayes action is

$$\frac{\int_{X_{(n)}}^\infty \theta^{-n} d\theta}{\int_{X_{(n)}}^\infty \theta^{-(n+1)} d\theta} = \frac{nX_{(n)}}{n-1}. \blacksquare$$

Exercise 14 (#4.25). Let $(X_1, ..., X_n)$ be a random sample from the exponential distribution on $(0, \infty)$ with scale parameter 1. Suppose that we observe $T = X_1 + \cdots + X_\theta$, where θ is an unknown positive integer. Consider the estimation of θ under the loss function $L(\theta, a) = (\theta - a)^2/\theta$ and the geometric distribution with mean p^{-1} as the prior for θ, where $p \in (0, 1)$ is known.
(i) Show that the posterior expected loss is

$$E[L(\theta, a)|T = t] = 1 + \xi - 2a + (1 - e^{-\xi})a^2/\xi,$$

where $\xi = (1 - p)t$.
(ii) Find the Bayes estimator of θ and show that its posterior expected loss

is $1 - \xi \sum_{m=1}^{\infty} e^{-m\xi}$.
(iii) Find the marginal distribution of $(1-p)T$, unconditional on θ.
(iv) Obtain an explicit expression for the Bayes risk of the Bayes estimator in part (ii).

Solution. (i) For a given θ, T has the gamma distribution with shape parameter θ and scale parameter 1. Hence, the joint probability density of (T, θ) is

$$f(t,\theta) = \frac{1}{(\theta-1)!} t^{\theta-1} e^{-t} p(1-p)^{\theta-1}, \quad t > 0, \theta = 1, 2, ...$$

and

$$\sum_{\theta=1}^{\infty} f(t,\theta) = pe^{-t} \sum_{\theta=1}^{\infty} \frac{[(1-p)t]^{\theta-1}}{(\theta-1)!} = pe^{-pt}.$$

Then,

$$\begin{aligned} E[L(\theta,a)|T=t] &= p^{-1}e^{pt} \sum_{\theta=1}^{\infty} \frac{(\theta-a)^2}{\theta} f(t,\theta) \\ &= e^{-\xi} \sum_{\theta=1}^{\infty} \frac{\xi^{\theta-1}}{\theta!} \theta^2 - 2ae^{-\xi} \sum_{\theta=1}^{\infty} \frac{\xi^{\theta-1}}{\theta!} \theta \\ &\quad + a^2 e^{-\xi} \sum_{\theta=1}^{\infty} \frac{\xi^{\theta-1}}{\theta!} \\ &= 1 + \xi - 2a + (1-e^{-\xi})a^2/\xi. \end{aligned}$$

(ii) Since $E[L(\theta,a)|T=t]$ is a quadratic function of a, the Bayes estimator is $\delta(T) = (1-p)T/(1-e^{-(1-p)T})$. The posterior expected loss when $T = t$ is

$$E[L(\theta,\delta(t))|T=t] = 1 - \frac{\xi e^{-\xi}}{1-e^{-\xi}} = 1 - \xi \sum_{m=1}^{\infty} e^{-m\xi}.$$

(iii) As shown in part (i) of the solution, the marginal density of T is $\sum_{\theta=1}^{\infty} f(t,\theta) = pe^{-pt}$, which is the density of the exponential distribution on $(0,\infty)$ with scale parameter p^{-1}.
(iv) The Bayes risk of $\delta(T)$ is

$$\begin{aligned} E\{E[L(\theta,\delta(T))|T]\} &= 1 - E\left\{(1-p)T \sum_{m=1}^{\infty} e^{-m(1-p)T}\right\} \\ &= 1 - (1-p)p \sum_{m=1}^{\infty} \int_0^{\infty} te^{-m(1-p)t} e^{-pt} dt \\ &= 1 - (1-p)p \sum_{m=1}^{\infty} \frac{1}{[m(1-p)+p]^2}, \end{aligned}$$

where the first equality follows from the result in (ii) and the second equality follows from the result in (iii). ∎

Exercise 15 (#4.27). Let $(X_1, ..., X_n)$ be a random sample of binary random variables with $P(X_1 = 1) = p \in (0, 1)$.
(i) Show that the sample mean $\bar{X}$ is an admissible estimator of p under the loss function $(a - p)^2/[p(1 - p)]$.
(ii) Show that $\bar{X}$ is an admissible estimator of p under the squared error loss.
Note. A unique Bayes estimator under a given proper prior is admissible.
Solution. (i) Let $T = n\bar{X}$. Consider the uniform distribution on the interval $(0, 1)$ as the prior for p. Then the Bayes estimator under the loss function $(a - p)^2/[p(1 - p)]$ is

$$\frac{\int_0^1 p^T(1-p)^{n-T-1}dp}{\int_0^1 p^{T-1}(1-p)^{n-T-1}dp} = \frac{T}{n} = \bar{X}.$$

Since the Bayes estimator is unique, $\bar{X}$ is an admissible estimator under the given loss function.
(ii) From the result in (i), there does not exist an estimator U such that

$$\frac{E(U-p)^2}{p(1-p)} \leq \frac{E(\bar{X}-p)^2}{p(1-p)}$$

for any $p \in (0, 1)$ and with strict inequality holds for some p. Since $p \in (0, 1)$, this implies that there does not exist an estimator U such that

$$E(U - p)^2 \leq E(\bar{X} - p)^2$$

for any $p \in (0, 1)$ and with strict inequality holds for some p. Hence $\bar{X}$ is an admissible estimator of p under the squared error loss. ∎

Exercise 16 (#4.28). Let $X = (X_1, ..., X_n)$ be a random sample from $N(\mu, 1)$, $\mu \in \mathcal{R}$. Show that the sample mean $\bar{X}$ is an admissible estimator of μ under the loss function $L(\mu, a) = |\mu - a|$.
Solution. Consider a sequence of priors, $N(0, j)$, $j = 1, 2, ...$ From Exercise 1, the posterior mean under the jth prior is $\delta_j = a_j\bar{X}$, where $a_j = nj/(nj + 1)$. From Exercise 6, δ_j is a Bayes estimator of μ. Let r_T be the Bayes risk for an estimator T. Then, for any j, $r_{\bar{X}} \geq r_{\delta_j}$ and

$$r_{\delta_j} = E[E(|a_j\bar{X} - \mu||\mu)] = a_jE[E(|\bar{X} - a_j^{-1}\mu||\mu)] \geq a_jr_{\bar{X}},$$

where the last inequality follows from Exercise 11 in Chapter 1 and the fact that given μ, μ is a median of $\bar{X}$. Hence,

$$0 \geq r_{\delta_j} - r_{\bar{X}} \geq (a_j - 1)r_{\bar{X}} = -(nj + 1)^{-1}r_{\bar{X}},$$

which implies that $r_{\delta_j} - r_{\bar{X}}$ converges to 0 at rate j^{-1} as $j \to \infty$. On the other hand, for any finite interval (a, b), the prior probability of $\mu \in (a, b)$ is $\Phi(b/\sqrt{j}) - \Phi(a/\sqrt{j})$, which converges to 0 at rate $j^{-1/2}$, where Φ is the cumulative distribution of $N(0, 1)$. Thus, by Blyth's theorem (e.g., Theorem 4.3 in Shao, 2003), $\bar{X}$ is admissible. ∎

Exercise 17. Let X be an observation from the negative binomial distribution with a known size r and an unknown probability $p \in (0, 1)$. Show that $(X + 1)/(r + 1)$ is an admissible estimator of p^{-1} under the squared error loss.

Solution. It suffices to show that $\delta_0(X) = (X + 1)/(r + 1)$ is admissible under the loss function $p^2(a - p^{-1})^2$. The posterior distribution of p given X is the beta distribution with parameter $(r + \alpha, X - r + \beta)$. Under the loss function $p^2(a - p^{-1})^2$, the Bayes estimator is

$$\delta(X) = \frac{E(p|X)}{E(p^2|X)} = \frac{X + \alpha + \beta + 1}{r + \alpha + 1},$$

which has risk

$$R_\delta(p) = \frac{r(1 - p)}{(r + \alpha + 1)^2} + \frac{[(\alpha + \beta + 1)p - (\alpha + 1)]^2}{(r + \alpha + 1)^2}$$

and Bayes risk

$$r_\delta = \frac{r\beta}{(\alpha + \beta)(r + \alpha + 1)^2} + \frac{\beta^2}{(\alpha + \beta)^2(r + \alpha + 1)^2} + \frac{\alpha\beta(\alpha + \beta + 1)}{(\alpha + \beta)^2(r + \alpha + 1)^2}.$$

Also, $\delta_0(X)$ has risk

$$R_{\delta_0}(p) = \frac{r(1 - p) + (1 - p)^2}{(r + 1)^2}$$

and Bayes risk

$$r_{\delta_0} = \frac{r\beta}{(\alpha + \beta)(r + 1)^2} + \frac{\beta^2}{(\alpha + \beta)^2(r + 1)^2} + \frac{\alpha\beta}{(\alpha + \beta + 1)(\alpha + \beta)^2(r + 1)^2}.$$

Note that

$$\frac{\alpha + \beta}{\alpha\beta}(r_{\delta_0} - r_\delta) = A + B + C,$$

where

$$A = \frac{r}{\alpha}\left[\frac{1}{(r + 1)^2} - \frac{1}{(r + \alpha + 1)^2}\right] \to \frac{2r}{(r + 1)^3}$$

if $\alpha \to 0$ and $\beta \to 0$,

$$B = \frac{\beta}{\alpha(\alpha + \beta)}\left[\frac{1}{(r + 1)^2} - \frac{1}{(r + \alpha + 1)^2}\right] \to \frac{2}{(r + 1)^3}$$

if $\alpha \to 0$, $\beta \to 0$ and $\alpha/\beta \to 0$, and

$$\begin{aligned} C &= \frac{1}{\alpha+\beta}\left[\frac{1}{(r+1)^2(\alpha+\beta+1)} - \frac{\alpha+\beta+1}{(r+\alpha+1)^2}\right] \\ &= \frac{[\alpha-(\alpha+\beta)(r+1)][r+\alpha+1+(\alpha+\beta+1)(r+1)]}{(\alpha+\beta)(\alpha+\beta+1)(r+1)^2(r+\alpha+1)^2} \\ &\to -\frac{2}{(r+1)^2} \end{aligned}$$

if $\alpha \to 0$, $\beta \to 0$ and $\alpha/\beta \to 0$. Therefore,

$$\frac{\alpha+\beta}{\alpha\beta}(r_{\delta_0} - r_\delta) \to 0$$

if $\alpha \to 0$, $\beta \to 0$ and $\alpha/\beta \to 0$. For any $0 < a < b < 1$, the prior probability of $p \in (a, b)$ is

$$\pi_{\alpha,\beta} = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\int_a^b p^{\alpha-1}(1-p)^{\beta-1}dp.$$

Note that $a\Gamma(a) = \Gamma(a+1) \to 1$ as $a \to 0$. Hence,

$$\frac{\alpha+\beta}{\alpha\beta}\pi_{\alpha,\beta} \to \int_a^b p^{-1}(1-p)^{-1}dp$$

as $\alpha \to 0$ and $\beta \to 0$. From this and the proved result, $(r_{\delta_0} - r_\delta)/\pi_{\alpha,\beta} \to 0$ if $\alpha \to 0$, $\beta \to 0$ and $\alpha/\beta \to 0$. By Blyth's theorem, $\delta_0(X)$ is admissible. ∎

Exercise 18 (#4.30). Let $(X_1, ..., X_n)$ be a random sample of binary random variables with $P(X_1 = 1) = p \in (0, 1)$.
(i) Obtain the Bayes estimator of $p(1-p)$ when the prior is the beta distribution with known parameter (α, β), under the squared error loss.
(ii) Compare the Bayes estimator in (i) with the UMVUE of $p(1-p)$.
(iii) Discuss the bias, consistency, and admissibility of the Bayes estimator in (i).
(iv) Let $[p(1-p)]^{-1}I_{(0,1)}(p)$ be an improper Lebesgue prior density for p. Show that the posterior of p given X_i's is a probability density provided that the sample mean $\bar{X} \in (0, 1)$.
(v) Under the squared error loss, find the generalized Bayes estimator of $p(1-p)$ under the improper prior in (iv).
Solution. (i) Let $T = \sum_{i=1}^n X_i$. Since the posterior density given $T = t$ is proportional to

$$p^{t+\alpha-1}(1-p)^{n-t+\beta-1}I_{(0,1)}(p),$$

the Bayes estimator of $p(1-p)$ is

$$\delta(T) = \frac{\int_0^1 p^{T+\alpha}(1-p)^{n-T+\beta}dp}{\int_0^1 p^{T+\alpha-1}(1-p)^{n-T+\beta-1}dp} = \frac{(T+\alpha+1)(n-T+\beta)}{(n+\alpha+\beta+2)(n+\alpha+\beta+1)}.$$

(ii) By considering functions of the form $aT^2 + bT$ of the complete and sufficient statistic T, we obtain the UMVUE of $p(1-p)$ as

$$U(T) = \frac{T(n-T)}{n(n-1)}.$$

(iii) From part (ii) of the solution, $E[T(n-T)] = n(n-1)p(1-p)$. Then the bias of $\delta(T)$ is

$$\left[\frac{n(n-1)}{(n+\alpha+\beta+2)(n+\alpha+\beta+1)} - 1\right] p(1-p)$$
$$+\frac{(\alpha+1)(n+\beta)+pn(\beta-\alpha-1)}{(n+\alpha+\beta+2)(n+\alpha+\beta+1)},$$

which is of the order $O(n^{-1})$. Since $\lim_n (T/n) = p$ a.s. by the strong law of large numbers, $\lim_n \delta(T) = p(1-p)$ a.s. Hence the Bayes estimator $\delta(T)$ is consistent. Since $\delta(T)$ is a unique Bayes estimator, it is admissible.
(iv) The posterior density when $T = t$ is proportional to

$$p^{t-1}(1-p)^{n-t-1} I_{(0,1)}(p),$$

which is proper if and only if $0 < t < n$.
(v) The generalized Bayes estimator is

$$\frac{\int_0^1 p^T (1-p)^{n-T} dp}{\int_0^1 p^{T-1}(1-p)^{n-T-1} dp} = \frac{T(n-T)}{n(n+1)}. \blacksquare$$

Exercise 19 (#4.35(a)). Let $X = (X_1, ..., X_n)$ be a random sample from the uniform distribution on $(\theta, \theta+1)$, $\theta \in \mathcal{R}$. Consider the estimation of θ under the squared error loss. Let $\pi(\theta)$ be a continuous and positive Lebesgue density on $\mathcal{R}$. Derive the Bayes estimator under the prior π and show that it is a consistent estimator of θ.
Solution. Let $X_{(j)}$ be the jth order statistic. The joint density of X is

$$I_{(\theta,\theta+1)}(X_{(1)}) I_{(\theta,\theta+1)}(X_{(n)}) = I_{(X_{(n)}-1, X_{(1)})}(\theta).$$

Hence, the Bayes estimator is

$$\delta(X) = \frac{\int_{X_{(n)}-1}^{X_{(1)}} \theta\pi(\theta)d\theta}{\int_{X_{(n)}-1}^{X_{(1)}} \pi(\theta)d\theta}.$$

Note that $\lim_n X_{(1)} = \theta$ a.s. and $\lim_n X_{(n)} = \theta + 1$ a.s. Hence, almost surely, the interval $(X_{(n)} - 1, X_{(1)})$ shrinks to a single point θ as $n \to \infty$. Since π is continuous, this implies that $\lim_n \delta(X) = \theta$ a.s. $\blacksquare$

Exercise 20 (#4.36). Consider the linear model with observed vector X having distribution $N_n(Z\beta, \sigma^2 I_n)$, where Z is an $n \times p$ known matrix, $p < n$, $\beta \in \mathcal{R}^p$, and $\sigma^2 > 0$.
(i) Assume that σ^2 is known. Derive the posterior distribution of β when the prior distribution for β is $N_p(\beta_0, \sigma^2 V)$, where $\beta_0 \in \mathcal{R}^p$ is known and V is a known positive definite matrix, and find the Bayes estimator of $l^\tau \beta$ under the squared error loss, where $l \in \mathcal{R}^p$ is known.
(ii) Show that the Bayes estimator in (i) is admissible and consistent as $n \to \infty$, assuming that the minimum eigenvalue of $Z^\tau Z \to \infty$.
(iii) Repeat (i) and (ii) when σ^2 is unknown and σ^{-2} has the gamma distribution with shape parameter α and scale parameter γ, where α and γ are known.
(iv) In part (iii), obtain Bayes estimators of σ^2 and $l^\tau \beta/\sigma$ under the squared error loss and show that they are consistent under the condition in (ii).
Solution. (i) The product of the joint density of X and the prior is proportional to

$$\sigma^{-n} \exp\left\{-\frac{\|X - Z\beta\|^2}{2\sigma^2}\right\} \exp\left\{-\frac{(\beta - \beta_0)^\tau V^{-1}(\beta - \beta_0)}{2\sigma^2}\right\}.$$

Since

$$\exp\left\{-\frac{\|X - Z\beta\|^2}{2\sigma^2}\right\} = \exp\left\{-\frac{\text{SSR}}{2\sigma^2}\right\} \exp\left\{-\frac{(\hat\beta - \beta)^\tau Z^\tau Z(\hat\beta - \beta)}{2\sigma^2}\right\},$$

where $\hat\beta$ is the LSE of β and $\text{SSR} = \|X - Z\hat\beta\|^2$, the product of the joint density of X and the prior is proportional to

$$\exp\left\{-\frac{\beta^\tau(Z^\tau Z + V^{-1})\beta - 2\beta^\tau(V^{-1}\beta_0 + Z^\tau Z\hat\beta)}{2\sigma^2}\right\},$$

which is proportional to

$$\exp\left\{-\frac{(\beta - \beta^*)^\tau(Z^\tau Z + V^{-1})(\beta - \beta^*)}{2\sigma^2}\right\},$$

where

$$\beta^* = (Z^\tau Z + V^{-1})^{-1}(V^{-1}\beta_0 + Z^\tau Z\hat\beta).$$

This shows that the posterior of β is $N_p(\beta^*, \sigma^2(Z^\tau Z + V^{-1})^{-1})$. The Bayes estimator of $l^\tau\beta$ under the squared error loss is then $l^\tau\beta^*$.
(ii) Since the Bayes estimator $l^\tau\beta^*$ is unique, it is admissible. If the minimum eigenvalue of $Z^\tau Z \to \infty$ as $n \to \infty$, then $\hat\beta \to_p \beta$,

$$\lim_n l^\tau(Z^\tau Z + V^{-1})^{-1}V^{-1}\beta_0 = 0,$$

and

$$\lim_n (Z^\tau Z + V^{-1})^{-1} Z^\tau Z l = l.$$

Hence, $l^\tau \beta^* \to_p \beta$.
(iii) Let $\omega = \sigma^{-2}$. Then the product of the joint density of X and the prior is proportional to

$$\omega^{\alpha-1} e^{-\omega/\gamma} \omega^{n/2} \exp\left\{-\frac{\omega\|X - Z\beta\|^2}{2}\right\} \exp\left\{-\frac{\omega(\beta-\beta_0)^\tau V^{-1}(\beta-\beta_0)}{2}\right\},$$

which is proportional to (under the argument in part (i) of the solution),

$$\omega^{n/2+\alpha-1} e^{-\omega/\gamma} \exp\left\{-\frac{\omega \mathrm{SSR}}{2}\right\} \exp\left\{-\frac{(\beta-\beta^*)^\tau (Z^\tau Z + V^{-1})(\beta-\beta^*)}{2\sigma^2}\right\}.$$

Hence, the posterior of (β, ω) is $p(\beta|\omega)p(\omega)$ with $p(\beta|\omega)$ being the density of $N(\beta^*, \omega^{-1}(Z^\tau Z + V^{-1})^{-1})$ and $p(\omega)$ being the density of the gamma distribution with shape parameter $n/2 + \alpha$ and scale parameter $(\gamma^{-1} + \mathrm{SSR}/2)^{-1}$. The Bayes estimator of $l^\tau\beta$ is still $l^\tau\beta^*$ and the proof of its admissibility and consistency is the same as that in part (ii) of the solution.
(iv) From the result in part (iii) of the solution, the Bayes estimator of $\sigma^2 = \omega^{-1}$ is

$$\int_0^\infty \omega^{-1} p(\omega) d\omega = \frac{\gamma^{-1} + \mathrm{SSR}/2}{n/2 + \alpha - 1}.$$

It is consistent since $\mathrm{SSR}/n \to_p \sigma^2$. Using

$$E(\beta/\sigma) = E[E(\beta/\sigma|\sigma)] = E[\sigma^{-1}E(\beta|\sigma)]$$

and the fact that β^* does not depend on σ^2, we obtain the Bayes estimator of $l^\tau\beta/\sigma$ as

$$l^\tau\beta^* \int_0^\infty \omega^{1/2} p(\omega) d\omega = l^\tau\beta^* \frac{\Gamma(n/2 + \alpha + 1/2)}{\Gamma(n/2+\alpha)\sqrt{\gamma^{-1} + \mathrm{SSR}/2}}.$$

From the fact that

$$\lim_n \frac{\Gamma(n + \alpha + 1/2)}{\sqrt{n}\Gamma(n+\alpha)} = 1,$$

the consistency of the Bayes estimator follows from the consistency of $l^\tau\beta^*$ and SSR/n. ∎

Exercise 21 (#4.47). Let $(X_1, ..., X_n)$ be a random sample of random variables with the Lebesgue density $\sqrt{2/\pi} e^{-(x-\theta)^2/2} I_{(\theta,\infty)}(x)$, where $\theta \in \mathcal{R}$ is unknown. Find the MRIE (minimum risk invariant estimator) of θ under the squared error loss.
Note. See Sections 2.3.2 and 4.2 in Shao (2003) for definition and discussion of invariant estimators.

Solution. Let $f(x_1, ..., x_n)$ be the joint density of $(X_1, ..., X_n)$. When $\theta = 0$,

$$f(X_1 - t, ..., X_n - t) = \left(\frac{2}{\pi}\right)^{n/2} \exp\left\{-\frac{1}{2}\sum_{i=1}^{n}(X_i - t)^2\right\} I_{(t,\infty)}(X_{(1)})$$
$$= \left(\frac{2}{\pi}\right)^{n/2} e^{-(n-1)S^2/2} e^{-n(\bar{X}-t)^2/2} I_{(t,\infty)}(X_{(1)}),$$

where $\bar{X}$ is the sample mean, S^2 is the sample variance, and $X_{(1)}$ is the smallest order statistic. Under the squared error loss, the MRIE of θ is Pitman's estimator (e.g., Theorem 4.6 in Shao, 2003)

$$\frac{\int t f(X_1 - t, ..., X_n - t)dt}{\int f(X_1 - t, ..., X_n - t)dt} = \frac{\int_0^{X_{(1)}} t e^{-n(\bar{X}-t)^2/2} dt}{\int_0^{X_{(1)}} e^{-n(\bar{X}-t)^2/2} dt},$$

which is the same as the estimator δ given in Exercise 9. ∎

Exercise 22 (#4.48). Let $(X_1, ..., X_n)$ be a random sample from the exponential distribution on (μ, ∞) with a known scale parameter θ, where $\mu \in \mathcal{R}$ is unknown. Let $X_{(1)}$ be the smallest order statistic. Show that
(i) $X_{(1)} - \theta \log 2/n$ is an MRIE of μ under the absolute error loss $L(\mu - a) = |\mu - a|$;
(ii) $X_{(1)} - t$ is an MRIE under the loss function $L(\mu - a) = I_{(t,\infty)}(|\mu - a|)$.
Solution. Let $D = (X_1 - X_n, ..., X_{n-1} - X_n)$. Then the distribution of D does not depend on μ. Since $X_{(1)}$ is complete and sufficient for μ, by Basu's theorem, $X_{(1)}$ and D are independent. Since $X_{(1)}$ is location invariant, by Theorem 4.5(iii) in Shao (2003), $X_{(1)} - u_*$ is an MRIE of μ, where u_* minimizes $E_0[L(X_{(1)} - u)]$ over u and E_0 is the expectation taken under $\mu = 0$. Since $E_0[L(X_{(1)} - u)] = E_0|X_{(1)} - u|$, u_* is the median of the distribution of $X_{(1)}$ when $\mu = 0$ (Exercise 11 in Chapter 1). Since $X_{(1)}$ has Lebesgue density $n\theta^{-1}e^{-nx/\theta}I_{(0,\infty)}(x)$ when $\mu = 0$,

$$\frac{1}{2} = \frac{n}{\theta}\int_0^{u_*} e^{-nx/\theta} dx = 1 - e^{-nu_*/\theta}$$

and, hence, $u_* = \theta \log 2/n$.
(ii) Following the same argument in part (i) of the solution, we conclude that an MRIE of μ is $X_{(1)} - u_*$, where u_* minimizes $E_0[L(X_{(1)} - u)] = P_0(|X_{(1)} - u| > t)$ over u and E_0 and P_0 are the expectation and probability under $\mu = 0$. When $u \leq 0$, $P_0(|X_{(1)} - u| > t) \geq P_0(X_{(1)} > t)$. Hence, we only need to consider $u > 0$. A direct calculation shows that

$$E_0[L(X_{(1)} - u)] = P_0(X_{(1)} > u + t) + P_0(X_{(1)} < u - t)$$
$$= 1 - e^{-n\min\{u-t,0\}/\theta} + e^{-n(u+t)/\theta},$$

which is minimized at $u_* = t$. ∎

Exercise 23 (#4.52). Let $X = (X_1, ..., X_n)$ be a random sample from a population in a location family with unknown location parameter $\mu \in \mathcal{R}$ and T be a location invariant estimator of μ. Show that T is an MRIE under the squared error loss if and only if T is unbiased and $E[T(X)U(X)] = 0$ for any $U(X)$ satisfying $U(X_1 + c, ..., X_n + c) = U(X)$ for any c, $\text{Var}(U) < \infty$, and $E[U(X)] = 0$ for any μ.

Solution. Suppose that T is an MRIE of μ. Then T is unbiased. For any $U(X)$ satisfying $U(X_1 + c, ..., X_n + c) = U(X)$ for any c and $E[U(X)] = 0$ for any μ, $T + tU$ is location invariant and unbiased. Since T is an MRIE,

$$\text{Var}(T) \leq V(T + tU) = \text{Var}(T) + 2t\text{Cov}(T, U) + t^2\text{Var}(U),$$

which is the same as $0 \leq 2tE(TU) + t^2\text{Var}(U)$. This is impossible unless $E(TU) = 0$.

Suppose now that T is unbiased and $E[T(X)U(X)] = 0$ for any $U(X)$ satisfying $U(X_1 + c, ..., X_n + c) = U(X)$ for any c, $\text{Var}(U) < \infty$, and $E[U(X)] = 0$ for any μ. Let T_0 be Pitman's estimator (MRIE). Then $U = T - T_0$ satisfies $U(X_1 + c, ..., X_n + c) = U(X)$ for any c, $\text{Var}(U) < \infty$, and $E[U(X)] = 0$ for any μ. Then $E[T(T - T_0)] = 0$. Since T_0 is an MRIE, from the previous proof we know that $E[T_0(T - T_0)] = 0$. Then $E(T - T_0)^2 = E[T(T - T_0)] - E[T_0(T - T_0)] = 0$. Thus, $T = T_0$ a.s. and T is an MRIE. ∎

Exercise 24 (#4.56). Let $(X_1, ..., X_n)$ be a random sample from the uniform distribution on $(0, \sigma)$ and consider the estimation of $\sigma > 0$. Show that the MRIE of σ is $2^{(n+1)^{-1}} X_{(n)}$ when the loss is $L(\sigma, a) = |1 - a/\sigma|$, where $X_{(n)}$ is the largest order statistic.

Solution. By Basu's theorem, the scale invariant estimator $X_{(n)}$ is independent of $Z = (Z_1, ..., Z_n)$, where $Z_i = X_i/X_n$, $i = 1, ..., n-1$, and $Z_n = X_n/|X_n|$. By Theorem 4.8 in Shao (2003), the MRIE is $X_{(n)}/u_*$, where u_* minimizes $E_1|1 - X_{(n)}/u|$ over $u > 0$ and E_1 is the expectation under $\sigma = 1$. If $u \geq 1$, then $|1 - X_{(n)}/u| = 1 - X_{(n)}/u \geq 1 - X_{(n)} = |1 - X_{(n)}|$. Hence, we only need to consider $0 < u < 1$. Since $X_{(n)}$ has Lebesgue density $nx^{n-1}I_{(0,1)}(x)$ when $\sigma = 1$,

$$\begin{aligned}
E_1|(X_{(n)}/u) - 1| &= n\int_0^1 \left|\frac{x}{u} - 1\right| x^{n-1}dx \\
&= \frac{n}{u}\int_0^u (u - x)x^{n-1}dx + \frac{n}{u}\int_u^1 (x - u)x^{n-1}dx \\
&= u^n - \frac{n}{n+1}u^n + \frac{n}{n+1}\frac{1 - u^{n+1}}{u} - (1 - u^n) \\
&= \frac{2}{n+1}u^n + \frac{n}{n+1}\frac{1}{u} - 1,
\end{aligned}$$

which is minimized at $u_* = 2^{-(n+1)^{-1}}$. Thus, the MRIE is $2^{(n+1)^{-1}}X_{(n)}$. ∎

Exercise 25 (#4.59). Let $(X_1, ..., X_n)$ be a random sample from the Pareto distribution with Lebesgue density $\alpha\sigma^\alpha x^{-(\alpha+1)}I_{(\sigma,\infty)}(x)$, where $\sigma > 0$ is an unknown parameter and $\alpha > 2$ is known. Find the MRIE of σ under the loss function $L(\sigma, a) = (1 - a/\sigma)^2$.

Solution. By Basu's theorem, the scale invariant estimator $X_{(1)}$ is independent of $Z = (Z_1, ..., Z_n)$, where $X_{(1)}$ is the smallest order statistic, $Z_i = X_i/X_n$, $i = 1, ..., n-1$, and $Z_n = X_n/|X_n|$. By Theorem 4.8 in Shao (2003), the MRIE is $X_{(1)}/u_*$, where u_* minimizes $E_1(1 - X_{(1)}/u)^2$ over $u > 0$ and E_1 is the expectation under $\sigma = 1$. Since $X_{(1)}$ has Lebesgue density $n\alpha x^{-(n\alpha+1)}I_{(1,\infty)}(x)$ when $\sigma = 1$,

$$\begin{aligned} E_1(1 - X_{(1)}/u)^2 &= \frac{E_1(X_{(1)}^2) - 2uE_1(X_{(1)}) + u^2}{u^2} \\ &= \frac{n\alpha}{(n\alpha - 2)u^2} - \frac{2n\alpha}{(n\alpha - 1)u} + 1, \end{aligned}$$

which is minimized at $u_* = (n\alpha - 1)/(n\alpha - 2)$. Hence, the MRIE is equal to $(n\alpha - 2)X_{(1)}/(n\alpha - 1)$. ∎

Exercise 26 (#4.62). Let $(X_1, ..., X_n)$ be a random sample from the exponential distribution on (μ, ∞) with scale parameter σ, where $\mu \in \mathcal{R}$ and $\sigma > 0$ are unknown.

(i) Find the MRIE of σ under the loss $L(\sigma, a) = |1 - a/\sigma|^p$ with $p = 1$ or 2.

(ii) Under the loss function $L(\mu, \sigma, a) = (a - \mu)^2/\sigma^2$, find the MRIE of μ.

(iii) Compute the bias of the MRIE of μ in (ii).

Solution. Let $X_{(1)} = \min_{1\le i\le n} X_i$ and $T = \sum_{i=1}^n (X_i - X_{(1)})$. Then $(X_{(1)}, T)$ is complete and sufficient for (μ, σ); $X_{(1)}$ and T are independent; T is location-scale invariant and T/σ has the gamma distribution with shape parameter $n - 1$ and scale parameter 1; and $X_{(1)}$ is location-scale invariant and has Lebesgue density $n\sigma^{-1}e^{-n(x-\mu)/\sigma}I_{(\mu,\infty)}(x)$.

(i) Let $W = (W_1, ..., W_{n-1})$, where $W_i = (X_i - X_n)/(X_{n-1} - X_n)$, $i = 1, ..., n-2$, and $W_{n-1} = (X_{n-1} - X_n)/|X_{n-1} - X_n|$. By Basu's theorem, T is independent of W. Hence, according to formula (4.28) in Shao (2003), the MRIE of σ is T/u_*, where u_* minimizes $E_1|1 - T/u|^p$ over $u > 0$ and E_1 is the expectation taken under $\sigma = 1$.

When $p = 1$,

$$\begin{aligned} E_1|1 - T/u| &= \frac{1}{u}\left[\int_0^u (u - t)f_n(t)dt + \int_u^\infty (t - u)f_n(t)dt\right] \\ &= \int_0^u f_n(t)dt - \frac{1}{u}\int_0^u tf_n(t)dt \\ &\quad + \frac{1}{u}\int_u^\infty tf_n(t)dt - \int_u^\infty f_n(t)dt, \end{aligned}$$

where $f_n(t)$ denotes the Lebesgue density of the gamma distribution with shape parameter $n-1$ and scale parameter 1. The derivative of the above function with respect to u is

$$\psi(u) = \frac{1}{u^2}\left(\int_0^u tf_n(t)dt - \int_u^\infty tf_n(t)dt\right).$$

The solution to $\psi(u)=0$ is u_* satisfying

$$\int_0^{u_*} tf_n(t)dt = \int_{u_*}^\infty tf_n(t)dt.$$

Since $tf_n(t)$ is proportional to the Lebesgue density of the gamma distribution with shape parameter n and scale parameter 1, u_* is the median of the gamma distribution with shape parameter n and scale parameter 1.

When $p=2$,

$$E_1(1-T/u)^2 = \frac{E_1(T^2)-2uE_1(T)+u^2}{u^2} = \frac{n(n-1)}{u^2} - \frac{2(n-1)}{u} + 1,$$

which is minimized at $u_* = n$.

(ii) By Basu's theorem, $(X_{(1)},T)$ is independent of W defined in part (i) of the solution. By Theorem 4.9 in Shao (2003), an MRIE of μ is $X_{(1)} - u_*T$, where u_* minimizes $E_{0,1}(X_{(1)}-uT)^2$ over u and $E_{0,1}$ is the expectation taken under $\mu=0$ and $\sigma=1$. Note that

$$\begin{aligned} E_{0,1}(X_{(1)}-uT)^2 &= E_{0,1}(X_{(1)}^2) - 2uE_{0,1}(X_{(1)})E_{0,1}(T) + u^2E_{0,1}(T^2) \\ &= \frac{2}{n^2} - \frac{2(n-1)u}{n} + n(n-1)u^2, \end{aligned}$$

which is minimized at $u_* = n^{-2}$. Hence the MRIE of μ is $X_{(1)} - n^{-2}T$.

(iii) Note that

$$E(X_{(1)} - n^{-2}T) = \frac{\sigma}{n} + \mu - \frac{(n-1)\sigma}{n^2} = \mu + \frac{\sigma}{n^2}.$$

Hence, the bias of the MRIE in (ii) is σ/n^2. ∎

Exercise 27 (#4.67). Let $(X_1,...,X_n)$ be a random sample of binary random variables with $P(X_1=1)=p\in(0,1)$. Let T be a randomized estimator of p with probability $n/(n+1)$ being the sample mean $\bar{X}$ and probability $1/(n+1)$ being $\frac{1}{2}$. Under the squared error loss, show that T has a constant risk that is smaller than the maximum risk of $\bar{X}$.

Solution. The risk of T is

$$E\left[(p-\bar{X})^2\frac{n}{n+1} + \left(p-\tfrac{1}{2}\right)^2\frac{1}{n+1}\right] = \frac{1}{4(n+1)}.$$

The maximum risk of $\bar{X}$ is

$$\max_{0<p<1} \frac{p(1-p)}{n} = \frac{1}{4n} > \frac{1}{4(n+1)}. \quad \blacksquare$$

Exercise 28 (#4.68). Let X be a single sample from the geometric distribution with mean p^{-1}, where $p \in (0,1)$. Show that $I_{\{1\}}(X)$ is a minimax estimator of p under the loss function $L(p,a) = (a-p)^2/[p(1-p)]$.
Solution A. The risk function of any estimator $\delta(X)$ of p is

$$\begin{aligned} R_\delta(p) &= \sum_{x=1}^{\infty} [\delta(x) - p]^2 (1-p)^{x-2} \\ &= \frac{[\delta(1)-p]^2}{1-p} + \sum_{x=2}^{\infty} [\delta(x) - p]^2 (1-p)^{x-2}. \end{aligned}$$

If $\delta(1) \neq 1$, then $\lim_{p\to 1} R_\delta(p) = \infty$ and, hence, $\sup_{0<p<1} R_\delta(p) = \infty$. If $\delta(1) = 1$, then

$$\sup_{0<p<1} R_\delta(p) \geq \lim_{p\to 0} R_\delta(p) = 1 + \sum_{x=2}^{\infty} [\delta(x)]^2 \geq 1.$$

The risk of $I_{\{1\}}(X)$ is

$$1 - p + p^2 \sum_{x=2}^{\infty} (1-p)^{x-2} = 1.$$

Therefore, $I_{\{1\}}(X)$ is minimax.
Solution B. From Solution A, $I_{\{1\}}(X)$ has constant risk 1. Let Π_j be the beta distribution with parameter $(j^{-1}, 1)$, $j = 1, 2, ...$ Under prior Π_j, the Bayes estimator of p under loss $(a-p)^2/[p(1-p)]$ is

$$\delta_j(X) = \begin{cases} \frac{j^{-1}}{x-1+j^{-1}} & x \geq 2 \\ 1 & x = 1 \end{cases}$$

and its Bayes risk is

$$\begin{aligned} r_{\delta_j} &= \int_0^1 \frac{1-p}{jp} p^{j^{-1}} dp + \sum_{x=2}^{\infty} \int_0^1 \frac{[\delta_j(x) - p]^2 (1-p)^{x-2}}{jp} p^{j^{-1}} dp \\ &= \frac{j}{j+1} + \sum_{x=2}^{\infty} \frac{[\delta_j(x)]^2 \Gamma(x-1)\Gamma(j^{-1})}{j\Gamma(x+j^{-1}-1)} \\ &\quad - \sum_{x=2}^{\infty} \frac{2\delta_j(x)\Gamma(x-1)\Gamma(j^{-1}+1)}{j\Gamma(x+j^{-1})} + \sum_{x=2}^{\infty} \frac{\Gamma(x-1)\Gamma(j^{-1}+2)}{j\Gamma(x+j^{-1}+1)}. \end{aligned}$$

For any $x = 2, 3, ...$ and $j = 1, 2, ...,$

$$\frac{[\delta_j(x)]^2\Gamma(x-1)\Gamma(j^{-1})}{j\Gamma(x+j^{-1}-1)} \leq \frac{1}{j^3(x-1)^2}.$$

Hence,

$$\lim_{j\to\infty}\sum_{x=2}^{\infty}\frac{[\delta_j(x)]^2\Gamma(x-1)\Gamma(j^{-1})}{j\Gamma(x+j^{-1}-1)} = 0.$$

Similarly,

$$\lim_{j\to\infty}\sum_{x=2}^{\infty}\frac{2\delta_j(x)\Gamma(x-1)\Gamma(j^{-1}+1)}{j\Gamma(x+j^{-1})} = 0$$

and

$$\lim_{j\to\infty}\sum_{x=2}^{\infty}\frac{\Gamma(x-1)\Gamma(j^{-1}+2)}{j\Gamma(x+j^{-1}+1)} = 0.$$

Thus, $\lim_{j\to\infty} r_{\delta_j} = 1$. By Theorem 4.12 in Shao (2003), $I_{\{1\}}(X)$ is minimax. ▮

Exercise 29 (#4.72). Let $(X_1, ..., X_m)$ be a random sample from $N(\mu_x, \sigma_x^2)$ and $(Y_1, ..., Y_n)$ be a random sample from $N(\mu_y, \sigma_y^2)$. Assume that X_i's and Y_j's are independent. Consider the estimation of $\Delta = \mu_y - \mu_x$ under the squared error loss.
(i) Show that $\bar{Y} - \bar{X}$ is a minimax estimator of Δ when σ_x and σ_y are known, where $\bar{X}$ and $\bar{Y}$ are the sample means based on X_i's and Y_i's, respectively.
(ii) Show that $\bar{Y} - \bar{X}$ is a minimax estimator of Δ when $\sigma_x \in (0, c_x]$ and $\sigma_y \in (0, c_y]$, where c_x and c_y are constants.
Solution. (i) Let $\Pi_{x,j} = N(0, j)$ and $\Pi_{y,j} = N(0, j)$, $j = 1, 2, ...,$ and let $\Pi_{x,j} \times \Pi_{y,j}$ be the prior of (μ_x, μ_y). From Exercise 1, the Bayes estimators for μ_x and μ_y are $\frac{mj}{mj+\sigma_x^2}\bar{X}$ and $\frac{nj}{nj+\sigma_y^2}\bar{Y}$, respectively. Hence, the Bayes estimator of Δ is

$$\delta_j = \frac{nj}{nj+\sigma_y^2}\bar{Y} - \frac{mj}{mj+\sigma_x^2}\bar{X}$$

with Bayes risk

$$r_{\delta_j} = \frac{j\sigma_y^2}{nj+\sigma_y^2} + \frac{j\sigma_x^2}{mj+\sigma_x^2}.$$

Since

$$\lim_{j\to\infty} r_{\delta_j} = \frac{\sigma_y^2}{n} + \frac{\sigma_x^2}{m},$$

which does not depend on (μ_x, μ_y) and is equal to the risk of $\bar{Y} - \bar{X}$, by Theorem 4.12 in Shao (2003), $\bar{Y} - \bar{X}$ is minimax.
(ii) Let $\Theta = \{(\mu_x, \mu_y, \sigma_x^2, \sigma_y^2) : \mu_x \in \mathcal{R}, \mu_y \in \mathcal{R}, \sigma_x \in (0, c_x], \sigma_y \in (0, c_y]\}$

and $\Theta_0 = \{(\mu_x, \mu_y, \sigma_x^2, \sigma_y^2) : \mu_x \in \mathcal{R}, \mu_y \in \mathcal{R}, \sigma_x = c_x, \sigma_y = c_y\}$. From (i), $\bar{Y} - \bar{X}$ is minimax when Θ_0 is considered as the parameter space. Let $R_{\bar{Y}-\bar{X}}(\theta)$ be the risk function of $\bar{Y} - \bar{X}$. Since

$$\sup_{\theta\in\Theta} R_{\bar{Y}-\bar{X}}(\theta) = \frac{c_x^2}{m} + \frac{c_y^2}{n} = \sup_{\theta\in\Theta_0} R_{\bar{Y}-\bar{X}}(\theta),$$

we conclude that $\bar{Y} - \bar{X}$ is minimax. ∎

Exercise 30 (#4.73). Consider the linear model with observed vector X having distribution $N_n(Z\beta, \sigma^2 I_n)$, where Z is an $n \times p$ known matrix, $p < n$, $\beta \in \mathcal{R}^p$, and $\sigma^2 > 0$, and the estimation of $l^\tau\beta$ under the squared error loss, where $l \in \mathcal{R}(Z)$. Show that the LSE $l^\tau\hat{\beta}$ is minimax if $\sigma^2 \in (0, c]$ with a constant c.

Solution. Using the same argument in the solution for the previous exercise, we only need to show the minimaxity of $l^\tau\beta$ in the case where σ^2 is known.

Assume that σ^2 is known. The risk of $l^\tau\hat{\beta}$ is $\sigma^2 l^\tau(Z^\tau Z)^- l$, which does not depend on β. Consider a sequence of prior $N_p(0, j^{-1}I_p)$, $j = 1, 2, ...$ From Exercise 20, the Bayes estimator of $l^\tau\beta$ is

$$\delta_j = A_j Z^\tau Z\hat{\beta},$$

where $A_j = (Z^\tau Z + j^{-1}I_p)^{-1}$. The risk of δ_j is

$$\mathrm{Var}(\delta_j) + (E\delta_j - l^\tau\beta)^2 = \sigma^2 l^\tau A_j Z^\tau Z A_j l + \|(l^\tau A_j Z^\tau Z - l^\tau)\beta\|^2.$$

Hence, the Bayes risk of δ_j is

$$\begin{aligned} r_j &= \mathrm{Var}(\delta_j) + (l^\tau A_j Z^\tau Z - l^\tau)E(\beta\beta^\tau)(Z^\tau Z A_j l - l) \\ &= \sigma^2 l^\tau A_j Z^\tau Z A_j l + j\|Z^\tau Z A_j l - l\|^2. \end{aligned}$$

Since $l \in \mathcal{R}(Z) = \mathcal{R}(Z^\tau Z)$, there is $\zeta \in \mathcal{R}^p$ such that $l = Z^\tau Z\zeta$. Then

$$Z^\tau Z A_j l - l = (Z^\tau Z A_j Z^\tau Z - Z^\tau Z)\zeta.$$

Let Γ be an orthogonal matrix such that $\Gamma^\tau\Gamma = I_p$ and $\Gamma^\tau Z^\tau Z\Gamma = \Lambda$, a diagonal matrix whose kth diagonal element is λ_k. Then

$$\begin{aligned} B &= \Gamma^\tau(Z^\tau Z A_j Z^\tau Z - Z^\tau Z)\Gamma \\ &= \Gamma^\tau Z^\tau Z\Gamma\Gamma^\tau A_j\Gamma\Gamma^\tau Z^\tau Z\Gamma - \Lambda \\ &= \Lambda[\Gamma^\tau(Z^\tau Z + j^{-1}I_p)\Gamma]^{-1}\Lambda - \Lambda \\ &= \Lambda(\Lambda + j^{-1}I_p)^{-1}\Lambda - \Lambda, \end{aligned}$$

which is a diagonal matrix whose kth diagonal element is equal to $-j^{-1}\lambda_k/(\lambda_k + j^{-1})$. Then

$$j\|Z^\tau Z A_j l - l\|^2 = \zeta^\tau\Gamma B^2\Gamma^\tau\zeta \to 0$$

as $j \to \infty$. Similarly,

$$\sigma^2 l^\tau A_j Z^\tau Z A_j l \to \sigma^2 l^\tau (Z^\tau Z)^- l$$

as $j \to \infty$. This shows that $\lim_{j\to\infty} r_j =$ the risk of $l^\tau \hat{\beta}$. Hence, by Theorem 4.12 in Shao (2003), $l^\tau \hat{\beta}$ is minimax. ∎

Exercise 31 (#4.74). Let X be an observation having the hypergeometric distribution with discrete probability density

$$\frac{\binom{\theta}{x}\binom{N-\theta}{r-x}}{\binom{N}{r}}, \quad x = \max\{0, r-(N-\theta)\}, ..., \min\{r, \theta\},$$

where N and r are known and θ is an unknown integer between 1 and N. Consider the estimation of θ/N under the squared error loss.
(i) Show that the risk function of $T(X) = \alpha X/r + \beta$ is constant, where $\alpha = \{1 + \sqrt{(N-r)/[r(N-1)]}\}^{-1}$ and $\beta = (1-\alpha)/2$.
(ii) Show that T in (i) is the minimax estimator of θ/N and the Bayes estimator with the prior

$$\Pi(\{\theta\}) = \frac{\Gamma(2c)}{[\Gamma(c)]^2} \int_0^1 \binom{N}{\theta} t^{\theta+c-1}(1-t)^{N-\theta+c-1}dt, \quad \theta = 1, ..., N,$$

where $c = \beta/(\alpha/r - 1/N)$.
Solution. (i) From the property of the hypergeometric distribution, $E(X) = r\theta/N$ and $\mathrm{Var}(X) = r\theta(N-\theta)(N-r)/[N^2(N-1)]$. Hence, the risk of T is

$$\begin{aligned}
E\left(T - \frac{\theta}{N}\right)^2 &= \frac{\alpha^2}{r^2}\mathrm{Var}(X) + \left[\frac{\alpha}{r}E(X) + \beta - \frac{\theta}{N}\right]^2 \\
&= \frac{\alpha^2\theta(N-\theta)(N-r)}{rN^2(N-1)} + \left[\frac{(\alpha-1)\theta}{N} + \beta\right]^2 \\
&= \beta^2 + \left[\frac{\alpha^2(N-r)}{rN(N-1)} + \frac{2(\alpha-1)}{N}\right]\theta \\
&\quad + \left[\frac{(\alpha-1)^2}{N^2} - \frac{\alpha^2(N-r)}{rN^2(N-1)}\right]\theta^2.
\end{aligned}$$

Setting the coefficients in front of θ and θ^2 to 0, we conclude that T has a constant risk if $\alpha = \{1 + \sqrt{(N-r)/[r(N-1)]}\}^{-1}$ and $\beta = (1-\alpha)/2$.
(ii) The posterior of θ is proportional to

$$\frac{\binom{\theta}{x}\binom{N-\theta}{r-x}\binom{N}{\theta}}{\binom{N}{r}} \int_0^1 t^{\theta+c-1}(1-t)^{N-\theta+c-1}dt,$$

which is proportional to

$$\binom{N-r}{\theta-x}\int_0^1 t^{\theta+c-1}(1-t)^{N-\theta+c-1}dt,$$

$\theta = x, ..., N-r+x$. The posterior mean of θ is

$$\frac{\sum_{\theta=x}^{N-r+x}\theta\binom{N-r}{\theta-x}\int_0^1 t^{\theta+c-1}(1-t)^{N-\theta+c-1}dt}{\sum_{\theta=x}^{N-r+x}\binom{N-r}{\theta-x}\int_0^1 t^{\theta+c-1}(1-t)^{N-\theta+c-1}dt}.$$

From the property of the binomial distribution,

$$\sum_{\theta=x}^{N-r+x}\binom{N-r}{\theta-x}t^{\theta-x}(1-t)^{N-r-\theta+x} = 1$$

and

$$\sum_{\theta=x}^{N-r+x}(\theta-x)\binom{N-r}{\theta-x}t^{\theta-x}(1-t)^{N-r-\theta+x} = (N-r)t.$$

Hence, the posterior mean of θ is equal to

$$x + \frac{(N-r)\int_0^1 t^{x+c}(1-t)^{r-x+c-1}dt}{\int_0^1 t^{x+c-1}(1-t)^{r-x+c-1}dt} = x + \frac{(N-r)(x+c)}{r+2c}.$$

Then, the Bayes estimator of θ/N is

$$\left(1+\frac{N-r}{r+2c}\right)\frac{X}{N} + \frac{(N-r)c}{N(r+2c)}.$$

A direct calculation shows that when $c = \beta/(\alpha/r - 1/N)$ with β and α defined in (i), the Bayes estimator is equal to T. Since T has constant risk and is a unique Bayes estimator, T is minimax. ∎

Exercise 32 (#4.75). Let X be an observation from $N(\mu, 1)$ and let μ have the improper Lebesgue prior density $\pi(\mu) = e^{\mu}$. Under the squared error loss, show that the generalized Bayes estimator of μ is $X + 1$, which is neither minimax nor admissible.

Solution. The posterior density of μ is proportional to

$$\exp\left\{-\frac{(\mu-x)^2}{2}+\mu\right\} \propto \exp\left\{-\frac{[\mu-(x+1)]^2}{2}\right\}.$$

Thus, the posterior distribution of μ is $N(X+1, 1)$ and the generalized Bayes estimator is $E(\mu|X) = X+1$. Since the risk of $X+1$ is $E(X+1-\mu)^2 = 1+E(X-\mu)^2 > E(X-\mu)^2$, which is the risk of X, we conclude that $X+1$ is neither minimax nor admissible. ∎

Exercise 33 (#4.76). Let X be an observation from the Poisson distribution with unknown mean $\theta > 0$. Consider the estimation of θ under the squared error loss.
(i) Show that $\sup_\theta R_T(\theta) = \infty$ for any estimator $T = T(X)$, where $R_T(\theta)$ is the risk of T.
(ii) Let $\Im = \{aX + b : a \in \mathcal{R}, b \in \mathcal{R}\}$. Show that 0 is an admissible estimator of θ within $\Im$.
Solution. (i) When the gamma distribution with shape parameter α and scale parameter γ is used as the prior for θ, the Bayes estimator is $\delta(X) = \gamma(X + \alpha)/(\gamma + 1)$ with Bayes risk $r_\delta = \alpha\gamma^2/(\gamma + 1)$. Then, for any estimator T,

$$\sup_{\theta>0} R_T(\theta) \geq r_\delta = \frac{\alpha\gamma^2}{\gamma + 1} \to \infty$$

as $\gamma \to \infty$.
(ii) The risk of 0 is θ^2. The risk of $aX + b$ is

$$a^2\text{Var}(X) + [aE(X) + b - \theta]^2 = (a-1)^2\theta^2 + [2(a-1)b + a^2]\theta + b^2.$$

If 0 is inadmissible, then there are a and b such that

$$\theta^2 \geq (a-1)^2\theta^2 + [2(a-1)b + a^2]\theta + b^2$$

for all $\theta > 0$. Letting $\theta \to 0$, we obtain that $b = 0$. Then

$$\theta \geq (a-1)^2\theta + a^2$$

for all $\theta > 0$. Letting $\theta \to 0$ again, we conclude that $a = 0$. This shows that 0 is admissible within the class $\Im$. ∎

Exercise 34 (#4.78). Let $(X_1, ..., X_n)$ be a random sample from the uniform distribution on the interval $(\mu - \frac{1}{2}, \mu + \frac{1}{2})$ with an unknown $\mu \in \mathcal{R}$. Under the squared error loss, show that $(X_{(1)} + X_{(n)})/2$ is the unique minimax estimator of μ, where $X_{(j)}$ is the jth order statistic.
Solution. Let $f(x_1, ..., x_n)$ be the joint density of $X_1, ..., X_n$. Then

$$f(x_1 - \mu, ..., x_n - \mu) = \begin{cases} 1 & \mu - \frac{1}{2} \leq x_{(1)} \leq x_{(n)} \leq \mu + \frac{1}{2} \\ 0 & \text{otherwise.} \end{cases}$$

The Pitman estimator of μ is

$$\frac{\int_{-\infty}^{\infty} tf(X_1 - t, ..., X_n - t)dt}{\int_{-\infty}^{\infty} f(X_1 - t, ..., X_n - t)dt} = \frac{\int_{X_{(n)}-\frac{1}{2}}^{X_{(1)}+\frac{1}{2}} tdt}{\int_{X_{(n)}-\frac{1}{2}}^{X_{(1)}+\frac{1}{2}} dt} = \frac{X_{(1)} + X_{(n)}}{2}.$$

Hence, $(X_{(1)} + X_{(n)})/2$ is admissible. Since $(X_{(1)} + X_{(n)})/2$ has constant risk, it is the unique minimax estimator (otherwise $(X_{(1)} + X_{(n)})/2$ can not be admissible). ∎

Exercise 35 (#4.80). Let $(X_1, ..., X_n)$ be a random sample from the exponential distribution on $(0, \infty)$ with unknown mean $\theta > 0$ and $\bar{X}$ be the sample mean. Show that $(n\bar{X} + b)/(n+1)$ is an admissible estimator of θ under the squared error loss for any $b \geq 0$ and that $n\bar{X}/(n+1)$ is a minimax estimator of θ under the loss function $L(\theta, a) = (a - \theta)^2/\theta^2$.
Solution. The joint Lebesgue density of $X_1, ..., X_n$ is

$$\theta^{-n} e^{-n\bar{X}/\theta} I_{(0,\infty)}(X_{(1)}),$$

where $X_{(1)}$ is the smallest order statistic. Let $T(X) = \bar{X}$, $\vartheta = -\theta^{-1}$, and $c(\vartheta) = \vartheta^n$. Then the joint density is of the form $c(\vartheta)e^{\vartheta T}$ with respect a σ-finite measure and the range of ϑ is $(-\infty, 0)$. For any $\vartheta_0 \in (-\infty, 0)$,

$$\int_{-\infty}^{\vartheta_0} e^{-b\vartheta/n} \vartheta^{-1} d\vartheta = \int_{\theta_0}^{0} e^{-b\vartheta/n} \vartheta^{-1} d\vartheta = \infty.$$

By Karlin's theorem (e.g., Theorem 4.14 in Shao, 2003), we conclude that $(n\bar{X} + b)/(n+1)$ is admissible under the squared error loss. This implies that $(n\bar{X} + b)/(n+1)$ is also admissible under the loss function $L(\theta, a) = (a - \theta)^2/\theta^2$. Since the risk of $n\bar{X}/(n+1)$ is

$$\frac{1}{\theta^2} E\left(\frac{n\bar{X} + b}{n+1} - \theta\right)^2 = \frac{1}{n+1},$$

$n\bar{X}/(n+1)$ is an admissible estimator with constant risk. Hence, it is minimax. ∎

Exercise 36 (#4.82). Let X be a single observation. Consider the estimation of $E(X)$ under the squared error loss.
(i) Find all possible values of α and β such that $\alpha X + \beta$ are admissible when X has the Poisson distribution with unknown mean $\theta > 0$.
(ii) When X has the negative binomial distribution with a known size r and an unknown probability $p \in (0, 1)$, show that $\alpha X + \beta$ is admissible when $\alpha \leq \frac{r}{r+1}$ and $\beta > r(1 - \alpha)$.
Solution. (i) An application of the results in Exercises 35-36 of Chapter 2 shows that $\alpha X + \beta$ is an inadmissible estimator of EX when (a) $\alpha > 1$ or $\alpha < 0$ or (b) $\alpha = 1$ and $\beta \neq 0$. If $\alpha = 0$, then, by Exercise 36 of Chapter 2, $\alpha X + \beta$ is inadmissible when $\beta \leq 0$; by Exercise 34 of Chapter 2, $\alpha X + \beta$ is admissible when $\beta > 0$.

The discrete probability density X is $\theta^x e^{-\theta}/x! = e^{-e^{\vartheta}} e^{\vartheta x}/x!$, where $\vartheta = \log\theta \in (-\infty, \infty)$. Consider $\alpha \in (0, 1]$. Let $\alpha = (1 + \lambda)^{-1}$ and $\beta = \gamma\lambda/(1 + \lambda)$. Since

$$\int_{-\infty}^{0} \frac{e^{-\gamma\lambda\vartheta}}{e^{-\lambda e^{\vartheta}}} d\vartheta = \infty$$

if and only if $\lambda\gamma \geq 0$, and

$$\int_0^\infty \frac{e^{-\gamma\lambda\vartheta}}{e^{-\lambda e^\vartheta}} d\vartheta = \infty$$

if and only if $\lambda \geq 0$, we conclude that $\alpha X + \beta$ is admissible when $0 < \alpha < 1$ and $\beta \geq 0$; and $\alpha X + \beta$ is admissible when $\alpha = 1$ and $\beta = 0$. By Exercise 36 of Chapter 2, $\alpha X + \beta$ is inadmissible if $\beta < 0$.

The conclusion is that $\alpha X + \beta$ is admissible if and only if (α, β) is in the following set:

$$\{\alpha = 0, \beta > 0\} \cup \{\alpha = 1, \beta = 0\} \cup \{0 < \alpha < 1, \beta \geq 0\}.$$

(ii) The discrete probability density of X is $\binom{x-1}{r-1} \frac{p^r}{(1-p)^r} e^{x \log(1-p)}$. Let $\theta = \log(1-p) \in (-\infty, 0)$, $\alpha = (1+\lambda)^{-1}$, and $\beta = \gamma\lambda/(1+\lambda)$. Note that

$$\int_c^0 e^{-\lambda\gamma\theta} \left(\frac{e^\theta}{1-e^\theta}\right)^{\lambda r} d\theta = \infty$$

if and only if $\lambda r \geq 1$, i.e., $\alpha \leq \frac{r}{r+1}$;

$$\int_{-\infty}^c e^{-\lambda\gamma\theta} \left(\frac{e^\theta}{1-e^\theta}\right)^{\lambda r} d\theta = \infty$$

if and only if $\gamma > r$, i.e., $\beta > r\lambda/(1+\lambda) = r(1-\alpha)$. The result follows from Karlin's theorem. ∎

Exercise 37 (#4.83). Let X be an observation from the distribution with Lebesgue density $\frac{1}{2}c(\theta)e^{\theta x - |x|}$, $|\theta| < 1$.
(i) Show that $c(\theta) = 1 - \theta^2$.
(ii) Show that if $0 \leq \alpha \leq \frac{1}{2}$, then $\alpha X + \beta$ is admissible for estimating $E(X)$ under the squared error loss.
Solution. (i) Note that

$$\begin{aligned}
\frac{1}{c(\theta)} &= \frac{1}{2}\int_{-\infty}^\infty e^{\theta x - |x|} dx \\
&= \frac{1}{2}\left(\int_{-\infty}^0 e^{\theta x + x} dx + \int_0^\infty e^{\theta x - x} dx\right) \\
&= \frac{1}{2}\left(\int_0^\infty e^{-(1+\theta)} dx + \int_0^\infty e^{-(1-\theta)x} dx\right) \\
&= \frac{1}{2}\left(\frac{1}{1+\theta} + \frac{1}{1-\theta}\right) \\
&= \frac{1}{1-\theta^2}.
\end{aligned}$$

(ii) Consider first $\alpha > 0$. Let $\alpha = (1+\lambda)^{-1}$ and $\beta = \gamma\lambda/(1+\lambda)$. Note that

$$\int_{-1}^{0} \frac{e^{-\gamma\lambda\theta}}{(1-\theta^2)^\lambda} d\theta = \int_{0}^{1} \frac{e^{-\gamma\lambda\theta}}{(1-\theta^2)^\lambda} d\theta = \infty$$

if and only if $\lambda \geq 1$, i.e., $\alpha \leq \frac{1}{2}$. Hence, $\alpha X + \beta$ is an admissible estimator of $E(X)$ when $0 < \alpha \leq \frac{1}{2}$.

Consider $\alpha = 0$. Since

$$\begin{aligned} E(X) &= \frac{1-\theta^2}{2}\left(\int_{-\infty}^{0} xe^{\theta x + x} dx + \int_{0}^{\infty} xe^{\theta x - x} dx\right) \\ &= \frac{1-\theta^2}{2}\left(-\int_{0}^{\infty} xe^{-(1+\theta)} dx + \int_{0}^{\infty} xe^{-(1-\theta)x} dx\right) \\ &= \frac{1-\theta^2}{2}\left(\frac{1+\theta}{1-\theta} - \frac{1-\theta}{1+\theta}\right) \\ &= \frac{2\theta}{1-\theta^2}, \end{aligned}$$

which takes any value in $(-\infty, \infty)$, the constant estimator β is an admissible estimator of $E(X)$ (Exercise 34 in Chapter 2). ∎

Exercise 38 (#4.84). Let X be an observation with the discrete probability density $f_\theta(x) = [x!(1-e^{-\theta})]^{-1}\theta^x e^{-\theta} I_{\{1,2,...\}}(x)$, where $\theta > 0$ is unknown. Consider the estimation of $\theta/(1-e^{-\theta})$ under the squared error loss.

(i) Show that the estimator X is admissible.

(ii) Show that X is not minimax unless $\sup_\theta R_T(\theta) = \infty$ for the risk $R_T(\theta)$ of any estimator $T = T(X)$.

(iii) Find a loss function under which X is minimax and admissible.

Solution. (i) Let $\vartheta = \log\theta$. Then the range of ϑ is $(-\infty, \infty)$. The probability density of X is proportional to

$$\frac{\theta^x e^{-\theta}}{1-e^{-\theta}} = \frac{e^{-e^\vartheta}}{1-e^{-e^\vartheta}} e^{\vartheta x}.$$

Hence, by Corollary 4.3 in Shao (2003), X is admissible under the squared error loss for $E(X) = \theta/(1-e^{-\theta})$.

(ii) The risk of X is

$$\mathrm{Var}(X) = E(X^2) - [E(X)]^2 = \frac{\theta+\theta^2}{1-e^{-\theta}} - \frac{\theta^2}{(1-e^{-\theta})^2} = \frac{\theta - e^{-\theta}(\theta+\theta^2)}{(1-e^{-\theta})^2},$$

which diverges to ∞ as $\theta \to \infty$. Hence, X is not minimax unless $\sup_\theta R_T(\theta) = \infty$ for any estimator $T = T(X)$.

(iii) Consider the loss $[E(X) - a]^2/\mathrm{Var}(X)$. Since X is admissible under the loss $[E(X)-a]^2$, it is also admissible under the loss $[E(X)-a]^2/\mathrm{Var}(X)$. Since the risk of X under loss $[E(X)-a]^2/\mathrm{Var}(X)$ is 1, it is minimax. ∎

Exercise 39 (#4.91). Suppose that X is distributed as $N_p(\theta, I_p)$, where $\theta \in \mathcal{R}^p$. Consider the estimation of θ under the loss $(a-\theta)^\tau Q(a-\theta)$ with a known positive definite $p\times p$ matrix Q. Show that the risk of the estimator

$$\delta_{c,r}^Q = X - \frac{r(p-2)}{\|Q^{-1/2}(X-c)\|^2} Q^{-1}(X-c)$$

is equal to

$$\mathrm{tr}(Q) - (2r - r^2)(p-2)^2 E(\|Q^{-1/2}(X-c)\|^{-2}).$$

Solution. Without loss of generality, we assume that $c=0$. Define $\delta_r = \delta_{0,r}^Q$, $Y = Q^{1/2}X$, $\mu = Q^{1/2}\theta$,

$$h(\mu) = R_{\delta_r}(\theta) = E\left\|Y - \frac{r(p-2)}{\|Q^{-1}Y\|^2}Q^{-1}Y - \mu\right\|^2,$$

and

$$\begin{aligned} g(\mu) &= \mathrm{tr}(Q) - (2r-r^2)(p-2)^2 E(\|Q^{-1/2}X\|^{-2}) \\ &= \mathrm{tr}(Q) - (2r-r^2)(p-2)^2 E(\|Q^{-1}Y\|^{-2}). \end{aligned}$$

Let $\lambda_* > 0$ be the largest eigenvalue of Q^{-1}. Consider the following family of priors for μ:

$$\{N_p(0, (\alpha+\lambda_*)Q^2 - Q) : \alpha > 0\}.$$

Then the marginal distribution of Y is $N_p(0, (\alpha+\lambda_*)Q^2)$ and

$$\begin{aligned} E[g(\mu)] &= \mathrm{tr}(Q) - (2r-r^2)(p-2)^2 E(\|Q^{-1}Y\|^{-2}) \\ &= \mathrm{tr}(Q) - \frac{(2r-r^2)(p-2)}{\alpha+\lambda_*}. \end{aligned}$$

Note that the posterior distribution of μ given Y is

$$N_p\left(\left(I_p - \frac{Q^{-1}}{\alpha+\lambda_*}\right)Y, Q - \frac{1}{\alpha+\lambda_*}I_p\right).$$

Hence,

$$\begin{aligned} E[h(\mu)] &= E\left\|Y - \frac{r(p-2)}{\|Q^{-1}Y\|^2}Q^{-1}Y - E(\mu|Y) + E(\mu|Y) - \mu\right\|^2 \\ &= E\left\|Y - \frac{r(p-2)}{\|Q^{-1}Y\|^2}Q^{-1}Y - E(\mu|Y)\right\|^2 + E\|E(\mu|Y) - \mu\|^2 \\ &= \frac{p}{\alpha+\lambda_*} - \frac{(2r-r^2)(p-2)}{\alpha+\lambda_*} + \mathrm{tr}(Q) - \frac{p}{\alpha+\lambda_*} \\ &= \mathrm{tr}(Q) - \frac{(2r-r^2)(p-2)}{\alpha+\lambda_*}. \end{aligned}$$

This shows that $E[h(\mu)] = E[g(\mu)]$. Using the same argument as that in the proof of Theorem 4.15 in Shao (2003), we conclude that $h(\mu) = g(\mu)$ for any μ. ∎

Exercise 40 (#4.92). Suppose that X is distributed as $N_p(\theta, \sigma^2 D)$, where $\theta \in \mathcal{R}^p$ is unknown, $\sigma^2 > 0$ is unknown, and D is a known $p \times p$ positive definite matrix. Consider the estimation of θ under the loss $\|a - \theta\|^2$. Show that the risk of the estimator

$$\tilde{\delta}_{c,r} = X - \frac{r(p-2)\sigma^2}{\|D^{-1}(X-c)\|^2} D^{-1}(X-c)$$

is equal to

$$\sigma^2 \left[\mathrm{tr}(D) - (2r - r^2)(p-2)^2 \sigma^2 E(\|D^{-1}(X-c)\|^{-2}) \right].$$

Solution. Define $Z = \sigma^{-1} D^{-1/2}(X - c)$ and $\psi = \sigma^{-1} D^{-1/2}(\theta - c)$. Then Z is distributed as $N_p(\psi, I_p)$ and

$$\begin{aligned}
\tilde{\delta}_{c,r} - c &= \sigma D^{1/2} Z - \frac{r(p-2)\sigma D^{-1/2} Z}{\|D^{-1/2} Z\|^2} \\
&= \sigma D^{1/2} \left[Z - \frac{r(p-2)\sigma D^{-1}}{\|D^{-1/2} Z\|^2} Z \right] \\
&= \sigma D^{1/2} \delta^D_{0,r},
\end{aligned}$$

where $\delta^D_{0,r}$ is defined in the previous exercise with $Q = D$. Then the risk of $\tilde{\delta}_{c,r}$ is

$$\begin{aligned}
R_{\tilde{\delta}_{c,r}}(\theta) &= E\left[(\tilde{\delta}_{c,r} - \theta)^\tau (\tilde{\delta}_{c,r} - \theta) \right] \\
&= \sigma^2 E\left[(\delta^D_{c,r} - \psi)^\tau D (\delta^D_{c,r} - \psi) \right] \\
&= \sigma^2 \left[\mathrm{tr}(D) - (2r - r^2)(p-2)^2 E(\|D^{-1/2} Z\|^{-2}) \right] \\
&= \sigma^2 \left[\mathrm{tr}(D) - (2r - r^2)(p-2)^2 \sigma^2 E(\|D^{-1}(X-c)\|^{-2}) \right],
\end{aligned}$$

where the third equality follows from the result of the previous exercise. ∎

Exercise 41 (#4.96). Let $X = (X_1, ..., X_n)$ be a random sample of random variables with probability density f_θ. Find an MLE (maximum likelihood estimator) of θ in each of the following cases.
(i) $f_\theta(x) = \theta^{-1} I_{\{1,...,\theta\}}(x)$, θ is an integer between 1 and θ_0.
(ii) $f_\theta(x) = e^{-(x-\theta)} I_{(\theta,\infty)}(x)$, $\theta > 0$.
(iii) $f_\theta(x) = \theta(1-x)^{\theta-1} I_{(0,1)}(x)$, $\theta > 1$.
(iv) $f_\theta(x) = \frac{\theta}{1-\theta} x^{(2\theta-1)/(1-\theta)} I_{(0,1)}(x)$, $\theta \in (\frac{1}{2}, 1)$.

(v) $f_\theta(x) = 2^{-1}e^{-|x-\theta|}$, $\theta \in \mathcal{R}$.
(vi) $f_\theta(x) = \theta x^{-2} I_{(\theta,\infty)}(x)$, $\theta > 0$.
(vii) $f_\theta(x) = \theta^x(1-\theta)^{1-x} I_{\{0,1\}}(x)$, $\theta \in [\frac{1}{2}, \frac{3}{4}]$.
(viii) $f_\theta(x)$ is the density of $N(\theta, \theta^2)$, $\theta \in \mathcal{R}$, $\theta \neq 0$.
(ix) $f_\theta(x) = \sigma^{-n} e^{-(x-\mu)/\sigma} I_{(\mu,\infty)}(x)$, $\theta = (\mu, \sigma) \in \mathcal{R} \times (0, \infty)$.
(x) $f_\theta(x) = \frac{1}{\sqrt{2\pi}\sigma x} e^{-(\log x - \mu)^2/(2\sigma^2)} I_{(0,\infty)}(x)$, $\theta = (\mu, \sigma^2) \in \mathcal{R} \times (0, \infty)$.
(xi) $f_\theta(x) = I_{(0,1)}(x)$ if $\theta = 0$ and $f_\theta(x) = (2\sqrt{x})^{-1} I_{(0,1)}(x)$ if $\theta = 1$.
(xii) $f_\theta(x) = \beta^{-\alpha} \alpha x^{\alpha-1} I_{(0,\beta)}(x)$, $\theta = (\alpha, \beta) \in (0, \infty) \times (0, \infty)$.
(xiii) $f_\theta(x) = \binom{\theta}{x} p^x (1-p)^{\theta-x} I_{\{0,1,\dots,\theta\}}(x)$, $\theta = 1, 2, \dots$, where $p \in (0,1)$ is known.
(xiv) $f_\theta(x) = \frac{1}{2}(1-\theta^2) e^{\theta x - |x|}$, $\theta \in (-1, 1)$.

Solution. (i) Let $X_{(n)}$ be the largest order statistic. The likelihood function is $\ell(\theta) = \theta^{-n} I_{\{X_{(n)},\dots,\theta_0\}}(\theta)$, which is 0 when $\theta < X_{(n)}$ and decreasing on $\{X_{(n)}, \dots, \theta_0\}$. Hence, the MLE of θ is $X_{(n)}$.
(ii) Let $X_{(1)}$ be the smallest order statistic. The likelihood function is $\ell(\theta) = \exp\{-\sum_{i=1}^n (X_i - \theta)\} I_{(0, X_{(1)})}(\theta)$, which is 0 when $\theta > X_{(1)}$ and increasing on $(0, X_{(1)})$. Hence, the MLE of θ is $X_{(1)}$.
(iii) Note that $\ell(\theta) = \theta^n \prod_{i=1}^n (1 - X_i)^{\theta-1} I_{(0,1)}(X_i)$ and, when $\theta > 1$,

$$\frac{\partial \log \ell(\theta)}{\partial \theta} = \frac{n}{\theta} + \sum_{i=1}^n \log(1 - X_i) \quad \text{and} \quad \frac{\partial^2 \log \ell(\theta)}{\partial \theta^2} = -\frac{n}{\theta^2} < 0.$$

The equation $\frac{\partial \log \ell(\theta)}{\partial \theta} = 0$ has a unique solution $\hat{\theta} = -n / \sum_{i=1}^n \log(1 - X_i)$. If $\hat{\theta} > 1$, then it maximizes $\ell(\theta)$. If $\hat{\theta} \leq 1$, then $\ell(\theta)$ is decreasing on the interval $(1, \infty)$. Hence the MLE of θ is $\max\{1, \hat{\theta}\}$.
(iv) Note that

$$\frac{\partial \log \ell(\theta)}{\partial \theta} = \frac{n}{\theta(1-\theta)} + \frac{1}{(1-\theta)^2} \sum_{i=1}^n \log X_i$$

and $\frac{\partial \log \ell(\theta)}{\partial \theta} = 0$ has a unique solution $\hat{\theta} = (1 - n^{-1} \sum_{i=1}^n \log X_i)^{-1}$. Also, $\frac{\partial \log \ell(\theta)}{\partial \theta} < 0$ when $\theta > \hat{\theta}$ and $\frac{\partial \log \ell(\theta)}{\partial \theta} > 0$ when $\theta < \hat{\theta}$. Hence, the MLE of θ is $\max\{\hat{\theta}, \frac{1}{2}\}$.
(v) Note that $\ell(\theta) = 2^{-n} \exp\{-\sum_{i=1}^n |X_i - \theta|\}$. Let F_n be the distribution putting mass n^{-1} to each X_i. Then, by Exercise 11 in Chapter 1, any median of F_n is an MLE of θ.
(vi) Since $\ell(\theta) = \theta^n \prod_{i=1}^n X_i^{-2} I_{(0, X_{(1)})}(\theta)$, the same argument in part (ii) of the solution yields the MLE $X_{(1)}$.
(vii) Let $\bar{X}$ be the sample mean. Since

$$\frac{\partial \log \ell(\theta)}{\partial \theta} = \frac{n\bar{X}}{\theta} - \frac{n - n\bar{X}}{1 - \theta},$$

$\frac{\partial \log \ell(\theta)}{\partial \theta} = 0$ has a unique solution $\bar{X}$, $\frac{\partial \log \ell(\theta)}{\partial \theta} < 0$ when $\theta > \bar{X}$, and $\frac{\partial \log \ell(\theta)}{\partial \theta} > 0$ when $\theta < \bar{X}$. Hence, the same argument in part (iv) of the solution yields the MLE

$$\hat{\theta} = \begin{cases} \frac{1}{2} & \text{if } \bar{X} \in [0, \frac{1}{2}) \\ \bar{X} & \text{if } \bar{X} \in [\frac{1}{2}, \frac{3}{4}) \\ \frac{3}{4} & \text{if } \bar{X} \in (\frac{3}{4}, 1]. \end{cases}$$

(viii) Note that

$$\frac{\partial \log \ell(\theta)}{\partial \theta} = -\frac{1}{\theta^3}\left(n\theta^2 + \theta \sum_{i=1}^n X_i - \sum_{i=1}^n X_i^2\right).$$

The equation $\frac{\partial \log \ell(\theta)}{\partial \theta} = 0$ has two solutions

$$\theta_{\pm} = \frac{-\sum_{i=1}^n X_i \pm \sqrt{(\sum_{i=1}^n X_i)^2 + 4n\sum_{i=1}^n X_i^2}}{2n}.$$

Note that $\lim_{\theta \to 0} \log \ell(\theta) = -\infty$. By checking the sign change at the neighborhoods of $\theta_{\pm}$, we conclude that both θ_- and θ_+ are local maximum points. Therefore, the MLE of θ is

$$\hat{\theta} = \begin{cases} \theta_- & \text{if } \ell(\theta_-) \ge \ell(\theta_+) \\ \theta_+ & \text{if } \ell(\theta_-) < \ell(\theta_+). \end{cases}$$

(ix) The likelihood function

$$\ell(\theta) = \sigma^{-n} \exp\left\{-\frac{1}{\sigma}\sum_{i=1}^n (X_i - \mu)\right\} I_{(0, X_{(1)})}(\mu)$$

is 0 when $\mu > X_{(1)}$ and increasing on $(0, X_{(1)})$. Hence, the MLE of μ is $X_{(1)}$. Substituting $\mu = X_{(1)}$ into $\ell(\theta)$ and maximizing the resulting likelihood function yields that the MLE of σ is $n^{-1}\sum_{i=1}^n (X_i - X_{(1)})$.

(x) Let $Y_i = \log X_i$, $i = 1, ..., n$. Then

$$\ell(\theta) = \frac{1}{(\sqrt{2\pi}\sigma)^n} \exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^n (Y_i - \mu)^2 - \sum_{i=1}^n Y_i\right\}.$$

Solving $\frac{\partial \log \ell(\theta)}{\partial \theta} = 0$, we obtain the MLE of μ as $\bar{Y} = n^{-1}\sum_{i=1}^n Y_i$ and the MLE of σ^2 as $n^{-1}\sum_{i=1}^n (Y_i - \bar{Y})^2$.

(xi) Since $\ell(0) = 1$ and $\ell(1) = (2^n \prod_{i=1}^n \sqrt{X_i})^{-1}$, the MLE is equal to 0 if $2^n \prod_{i=1}^n \sqrt{X_i} < 1$ and is equal to 1 if $2^n \prod_{i=1}^n \sqrt{X_i} \ge 1$.

(xii) The likelihood function is $\ell(\theta) = \alpha^n \beta^{-n\alpha} \prod_{i=1}^n X_i^{\alpha-1} I_{(X_{(n)}, \infty)}(\beta)$,

which is 0 when $\beta < X_{(n)}$ and decreasing in β otherwise. Hence the MLE of β is $X_{(n)}$. Substituting $\beta = X_{(n)}$ into the likelihood function, we obtain the MLE of α as $n[\sum_{i=1}^n \log(X_{(n)}/X_i)]^{-1}$.
(xiii) Let $X_{(n)}$ be the largest X_i's and $T = \sum_{i=1}^n X_i$. Then

$$\ell(\theta) = \prod_{i=1}^n \binom{\theta}{X_i} p^T (1-p)^{n\theta - T} I_{\{X_{(n)}, X_{(n)}+1,...\}}(\theta).$$

For $\theta = X_{(n)}, X_{(n)} + 1, ...,$

$$\frac{\ell(\theta+1)}{\ell(\theta)} = (1-p)^n \prod_{i=1}^n \frac{\theta+1}{\theta+1-X_i}.$$

Since $(\theta+1)/(\theta+1-X_i)$ is decreasing in θ, the function $\ell(\theta+1)/\ell(\theta)$ is decreasing in θ. Also, $\lim_{\theta\to\infty} \ell(\theta+1)/\ell(\theta) = (1-p)^n < 1$. Therefore, the MLE of θ is $\max\{\theta : \theta \geq X_{(n)}, \ell(\theta+1)/\ell(\theta) \geq 1\}$.
(xiv) Let $\bar{X}$ be the sample mean. Then

$$\frac{\partial \log \ell(\theta)}{\partial \theta} = n\bar{X} - \frac{2n\theta}{1-\theta^2}.$$

The equation $\frac{\partial \log \ell(\theta)}{\partial \theta} = 0$ has two solutions $\theta_\pm = \pm\sqrt{1+\bar{X}^2} - 1$. Since $\theta_- < -1$ is not in the parameter space, we conclude that the MLE of θ is $\sqrt{1+\bar{X}^2} - 1$. ∎

Exercise 42. Let $(X_1, ..., X_n)$ be a random sample from the uniform distribution on the interval $(\theta, \theta + |\theta|)$. Find the MLE of θ when
(i) $\theta \in (0, \infty)$;
(ii) $\theta \in (-\infty, 0)$;
(iii) $\theta \in \mathcal{R}$, $\theta \neq 0$.
Solution. (i) When $\theta \in (0, \infty)$, the distribution of X_1 is uniform on $(\theta, 2\theta)$. The likelihood function is

$$\ell(\theta) = \theta^{-n} I_{(X_{(n)}/2, X_{(1)})}(\theta).$$

Since θ^{-n} is decreasing, the MLE of θ is $X_{(n)}/2$.
(ii) When $\theta \in (-\infty, 0)$, the distribution of X_1 is uniform on $(\theta, 0)$. Hence

$$\ell(\theta) = |\theta|^{-n} I_{(-\infty, X_{(1)})}(\theta).$$

Since $|\theta|^{-n}$ is decreasing, the MLE of θ is $X_{(1)}$.
(iii) Consider $\theta \neq 0$. If $\theta > 0$, then almost surely all X_i's are positive. If $\theta < 0$, then almost surely all X_i's are negative. Combining the results in (i)-(ii), we conclude that the MLE of θ is $X_{(n)}/2$ if $X_1 > 0$ and is $X_{(1)}$ if $X_1 < 0$. ∎

Exercise 43 (#4.98). Suppose that n observations are taken from $N(\mu, 1)$ with an unknown μ. Instead of recording all the observations, one records only whether the observation is less than 0. Find an MLE of μ.

Solution. Let $Y_i = 1$ if the ith observation is less than 0 and $Y_i = 0$ otherwise. Then $Y_1, ..., Y_n$ are the actual observations. Let $p = P(Y_i = 1) = \Phi(-\mu)$, where Φ is the cumulative distribution function of $N(0,1)$, $\ell(p)$ be the likelihood function in p, and $T = \sum_{i=1}^n Y_i$. Then

$$\frac{\partial \log \ell(p)}{\partial p} = \frac{T}{\theta} - \frac{n-T}{1-\theta}.$$

The likelihood equation has a unique solution T/n. Hence the MLE of p is T/n. Then, the MLE of μ is $-\Phi^{-1}(T/n)$. ∎

Exercise 44 (#4.100). Let $(Y_1, Z_1), ..., (Y_n, Z_n)$ be independent and identically distributed random 2-vectors such that Y_1 and Z_1 are independently distributed as the exponential distributions on $(0, \infty)$ with scale parameters $\lambda > 0$ and $\mu > 0$, respectively.
(i) Find the MLE of (λ, μ).
(ii) Suppose that we only observe $X_i = \min\{Y_i, Z_i\}$ and $\Delta_i = 1$ if $X_i = Y_i$ and $\Delta_i = 0$ if $X_i = Z_i$. Find the MLE of (λ, μ).

Solution. (i) Let $\ell(\lambda, \mu)$ be the likelihood function, $\bar{Y} = n^{-1}\sum_{i=1}^n Y_i$, and $\bar{Z} = n^{-1}\sum_{i=1}^n Z_i$. Since Y_i's and Z_i's are independent,

$$\frac{\partial \log \ell(\lambda, \mu)}{\partial \lambda} = -\frac{n}{\lambda} + \frac{n\bar{Y}}{\lambda^2} \quad \text{and} \quad \frac{\partial \log \ell(\lambda, \mu)}{\partial \mu} = -\frac{n}{\mu} + \frac{n\bar{Z}}{\mu^2}.$$

Hence, the MLE of (λ, μ) is $(\bar{Y}, \bar{Z})$.
(ii) The probability density of (X_i, Δ_i) is $\lambda^{-\Delta_i}\mu^{-(\Delta_i - 1)}e^{-(\lambda^{-1}+\mu^{-1})x_i}$. Let $T = \sum_{i=1}^n X_i$ and $D = \sum_{i=1}^n \Delta_i$. Then

$$\ell(\lambda, \mu) = \lambda^{-D}\mu^{D-n}e^{-(\lambda^{-1}+\mu^{-1})T}.$$

If $0 < D < n$, then

$$\frac{\partial \log \ell(\lambda, \mu)}{\partial \lambda} = -\frac{D}{\lambda} + \frac{T}{\lambda^2} \quad \text{and} \quad \frac{\partial \log \ell(\lambda, \mu)}{\partial \lambda} = \frac{D-n}{\mu} + \frac{T}{\mu^2}.$$

The likelihood equation has a unique solution $\hat{\lambda} = T/D$ and $\hat{\mu} = T/(n-D)$. The MLE of (λ, μ) is $(\hat{\lambda}, \hat{\mu})$.

If $D = 0$,

$$\ell(\lambda, \mu) = \mu^{-n}e^{-(\lambda^{-1}+\mu^{-1})T},$$

which is increasing in λ. Hence, there does not exist an MLE of λ. Similarly, when $D = n$, there does not exist an MLE of μ. ∎

Exercise 45 (#4.101). Let $(X_1, ..., X_n)$ be a random sample from the gamma distribution with shape parameter $\alpha > 0$ and scale parameter $\gamma > 0$. Show that almost surely the likelihood equation has a unique solution that is the MLE of $\theta = (\alpha, \gamma)$. Obtain the Newton-Raphson iteration equation and the Fisher-scoring iteration equation.

Solution. Let $\bar{X}$ be the sample mean and $Y = n^{-1}\sum_{i=1}^n \log X_i$. The log-likelihood function is

$$\log \ell(\theta) = -n\alpha \log \gamma - n \log \Gamma(\alpha) + (\alpha - 1)nY - \gamma^{-1}n\bar{X}.$$

Then, the likelihood equations are

$$-\log \gamma - \frac{\Gamma'(\alpha)}{\Gamma(\alpha)} + Y = 0 \quad \text{and} \quad -\frac{\alpha}{\gamma} + \frac{\bar{X}}{\gamma^2} = 0.$$

The second equation yields $\gamma = \bar{X}/\alpha$. Substituting $\gamma = \bar{X}/\alpha$ into the first equation we obtain that

$$h(\alpha) = \log \alpha - \frac{\Gamma'(\alpha)}{\Gamma(\alpha)} + Y - \log \bar{X} = 0.$$

From calculus,

$$\frac{\Gamma'(\alpha)}{\Gamma(\alpha)} = -C + \sum_{k=0}^{\infty} \left(\frac{1}{k+1} - \frac{1}{k+\alpha} \right)$$

and

$$\frac{d}{d\alpha} \left[\frac{\Gamma'(\alpha)}{\Gamma(\alpha)} \right] = \sum_{k=0}^{\infty} \frac{1}{(k+\alpha)^2},$$

where C is the Euler constant defined as

$$C = \lim_{m\to\infty} \left(\sum_{k=0}^{m-1} \frac{1}{k+1} - \log m \right).$$

Then

$$\begin{aligned} h'(\alpha) &= \frac{1}{\alpha} - \sum_{k=0}^{\infty} \frac{1}{(k+\alpha)^2} \\ &< \frac{1}{\alpha} - \sum_{k=0}^{\infty} \left(\frac{1}{k+\alpha} - \frac{1}{k+1+\alpha} \right) \\ &= \frac{1}{\alpha} - \frac{\Gamma'(\alpha)}{\Gamma(\alpha)} + \frac{\Gamma'(\alpha+1)}{\Gamma(\alpha+1)} \\ &= \frac{1}{\alpha} + \frac{d}{d\alpha} \log \frac{\Gamma(\alpha+1)}{\Gamma(\alpha)} \\ &= \frac{1}{\alpha} + \frac{d}{d\alpha} \log \frac{1}{\alpha} \\ &= 0. \end{aligned}$$

Hence, $h(\alpha)$ is decreasing. Also, it follows from the last two equalities of the previous expression that, for $m = 2, 3, ...,$

$$\frac{\Gamma'(m)}{\Gamma(m)} = \frac{1}{m-1} + \frac{1}{m-2} + \cdots + 1 + \frac{\Gamma'(1)}{\Gamma(1)} = \sum_{k=0}^{m-2} \frac{1}{k+1} - C.$$

Therefore,

$$\lim_{m\to\infty} \left[\log m - \frac{\Gamma'(m)}{\Gamma(m)}\right] = \lim_{m\to\infty} \left[\log m - \sum_{k=0}^{m-2} \frac{1}{k+1} + C\right] = 0$$

by the definition of C. Hence, $\lim_{\alpha\to\infty} h(\alpha) = Y - \log \bar{X}$, which is negative by Jensen's inequality when X_i's are not all the same. Since

$$\begin{aligned}
\lim_{\alpha\to 0} \left[\log \alpha - \frac{\Gamma'(\alpha)}{\Gamma(\alpha)}\right] &= \lim_{\alpha\to 0} \left[\log \alpha + C - \sum_{k=0}^{\infty} \left(\frac{1}{k+1} - \frac{1}{k+\alpha}\right)\right] \\
&= \lim_{\alpha\to 0} \left[\log \alpha + C + \frac{1}{\alpha} - 1 + \sum_{k=1}^{\infty} \frac{1-\alpha}{(k+1)(k+\alpha)}\right] \\
&= \lim_{\alpha\to 0} \left(\log \alpha + \frac{1}{\alpha}\right) + C - 1 + \sum_{k=1}^{\infty} \frac{1}{(k+1)k} \\
&= \infty,
\end{aligned}$$

we have $\lim_{\alpha\to 0} h(\alpha) = \infty$. Since h is continuous and decreasing, $h(\alpha) = 0$ has a unique solution. Thus, the likelihood equations have a unique solution, which is the MLE of θ.

Let

$$s(\theta) = \frac{\partial \log \ell(\theta)}{\partial \theta} = n\left(-\log\gamma - \frac{\Gamma'(\alpha)}{\Gamma(\alpha)} + Y, -\frac{\alpha}{\gamma} + \frac{\bar{X}}{\gamma^2}\right),$$

$$R(\theta) = \frac{\partial^2 \log \ell(\theta)}{\partial \theta \partial \theta^\tau} = n \begin{pmatrix} \left[\frac{\Gamma'(\alpha)}{\Gamma(\alpha)}\right]^2 - \frac{\Gamma''(\alpha)}{\Gamma(\alpha)} & -\frac{1}{\gamma} \\ -\frac{1}{\gamma} & \frac{\alpha}{\gamma^2} - \frac{2\bar{X}}{\gamma^3} \end{pmatrix},$$

and

$$F(\theta) = E[R(\theta)] = n \begin{pmatrix} \left[\frac{\Gamma'(\alpha)}{\Gamma(\alpha)}\right]^2 - \frac{\Gamma''(\alpha)}{\Gamma(\alpha)} & -\frac{1}{\gamma} \\ -\frac{1}{\gamma} & -\frac{\alpha}{\gamma^2} \end{pmatrix}.$$

Then the Newton-Raphson iteration equation is

$$\hat{\theta}^{(k+1)} = \hat{\theta}^{(k)} - [R(\hat{\theta}^{(k)})]^{-1} s(\hat{\theta}^{(k)}), \quad k = 0, 1, 2, ...$$

and the Fisher-scoring iteration equation is

$$\hat{\theta}^{(k+1)} = \hat{\theta}^{(k)} - [F(\hat{\theta}^{(k)})]^{-1} s(\hat{\theta}^{(k)}), \quad k = 0, 1, 2, ... \ \blacksquare$$

Exercise 46 (#4.102). Let $(X_1, ..., X_n)$ be a random sample from a population with discrete probability density $[x!(1-e^{-\theta})]^{-1}\theta^x e^{-\theta} I_{\{1,2,...\}}(x)$, where $\theta > 0$ is unknown. Show that the likelihood equation has a unique root when the sample mean $\bar{X} > 1$. Show whether this root is an MLE of θ.

Solution. Let $\ell(\theta)$ be the likelihood function and

$$h(\theta) = \frac{\partial \log \ell(\theta)}{\partial \theta} = n\left(\frac{\bar{X}}{\theta} - 1 - \frac{1}{e^\theta - 1}\right).$$

Obviously, $\lim_{\theta\to\infty} h(\theta) = -n$. Since $\lim_{\theta\to 0} \theta/(e^\theta - 1) = 1$,

$$\lim_{\theta\to 0} h(\theta) = n \lim_{\theta\to 0} \frac{1}{\theta}\left(\bar{X} - \frac{\theta}{e^\theta - 1}\right) - n = \infty$$

when $\bar{X} > 1$. Note that h is continuous. Hence, when $\bar{X} > 1$, $h(\theta) = 0$ has at least one solution. Note that

$$h'(\theta) = n\left[\frac{\bar{X}}{\theta^2} + \frac{e^\theta}{(e^\theta - 1)^2}\right] < 0$$

because $(e^\theta - 1)^2/e^\theta = (e^\theta - 1)(1 - e^{-\theta}) > \theta^2$. Hence, $h(\theta) = 0$ has a unique solution and $\log \ell(\theta)$ is convex. Therefore, the unique solution is the MLE of θ. ∎

Exercise 47 (#4.104). Let $(X_1, Y_1), ..., (X_n, Y_n)$ be independent and identically distributed as the bivariate normal distribution with $E(X_1) = E(Y_1) = 0$, $\mathrm{Var}(X_1) = \mathrm{Var}(Y_1) = 1$, and an unknown correlation coefficient $\rho \in (-1, 1)$. Show that the likelihood equation is a cubic in ρ and the probability that it has a unique root tends to 1 as $n \to \infty$.

Solution. Let $T = \sum_{i=1}^n (X_i^2 + Y_i^2)$ and $R = \sum_{i=1}^n X_i Y_i$. The likelihood function is

$$\ell(\rho) = (2\pi\sqrt{1-\rho^2})^{-n} \exp\left\{\frac{\rho R}{1-\rho^2} - \frac{T}{2(1-\rho^2)}\right\}.$$

Hence,

$$\frac{\partial \log \ell(\rho)}{\partial \rho} = \frac{n\rho}{1-\rho^2} + \frac{1+\rho^2}{(1-\rho^2)^2} R - \frac{\rho}{(1-\rho^2)^2} T$$

and the likelihood equation is $h(\rho) = 0$, where

$$h(\rho) = \rho(1-\rho^2) - n^{-1}T\rho + n^{-1}R(1+\rho^2)$$

is a cubic in ρ. Since h is continuous,

$$\lim_{\rho\to 1} h(\rho) = -n^{-1}(T - 2R) = -\frac{1}{n}\sum_{i=1}^n (X_i - Y_i)^2 < 0$$

and

$$\lim_{\rho \to -1} h(\rho) = n^{-1}(T + 2R) = \frac{1}{n}\sum_{i=1}^{n}(X_i + Y_i)^2 > 0,$$

the likelihood equation has at least one solution. Note that

$$h'(\rho) = 1 - 3\rho^2 - n^{-1}T + 2n^{-1}R\rho.$$

As $n \to \infty$, $n^{-1}T \to_p \text{Var}(X_1) + \text{Var}(Y_1) = 2$ and $n^{-1}R \to_p E(X_1 Y_1) = \rho$. Hence,

$$h'(\rho) \to_p 1 - 3\rho^2 - 2 + 2\rho^2 = -1 - \rho^2 < 0.$$

Therefore, the probability that $h(\rho) = 0$ has a unique solution tends to 1. ∎

Exercise 48 (#4.105). Let $(X_1, ..., X_n)$ be a random sample from the Weibull distribution with Lebesgue density $\alpha\theta^{-1}x^{\alpha-1}e^{-x^\alpha/\theta}I_{(0,\infty)}(x)$, where $\alpha > 0$ and $\theta > 0$ are unknown. Show that the likelihood equations are equivalent to $h(\alpha) = n^{-1}\sum_{i=1}^n \log X_i$ and $\theta = n^{-1}\sum_{i=1}^n X_i^\alpha$, where $h(\alpha) = (\sum_{i=1}^n X_i^\alpha)^{-1}\sum_{i=1}^n X_i^\alpha \log X_i - \alpha^{-1}$, and that the likelihood equations have a unique solution.

Solution. The log-likelihood function is

$$\log \ell(\alpha, \theta) = n\log\alpha - n\log\theta + (\alpha - 1)\sum_{i=1}^n \log X_i - \frac{1}{\theta}\sum_{i=1}^n X_i^\alpha.$$

Hence, the likelihood equations are

$$\frac{\partial \log \ell(\alpha,\theta)}{\partial \alpha} = \frac{n}{\alpha} + \sum_{i=1}^n \log X_i - \frac{1}{\theta}\sum_{i=1}^n X_i^\alpha \log X_i = 0$$

and

$$\frac{\partial \log \ell(\alpha,\theta)}{\partial \theta} = -\frac{n}{\theta} + \frac{1}{\theta^2}\sum_{i=1}^n X_i^\alpha = 0,$$

which are equivalent to $h(\alpha) = n^{-1}\sum_{i=1}^n \log X_i$ and $\theta = n^{-1}\sum_{i=1}^n X_i^\alpha$. Note that

$$h'(\alpha) = \frac{\sum_{i=1}^n X_i^\alpha (\log X_i)^2 \sum_{i=1}^n X_i^\alpha - (\sum_{i=1}^n X_i^\alpha \log X_i)^2}{(\sum_{i=1}^n X_i^\alpha)^2} + \frac{1}{\alpha^2} > 0$$

by the Cauchy-Schwarz inequality. Thus, $h(\alpha)$ is increasing. Since h is continuous, $\lim_{\alpha\to 0} h(\alpha) = -\infty$, and

$$\lim_{\alpha\to\infty} h(\alpha) = \lim_{\alpha\to\infty} \frac{\sum_{i=1}^n \left(\frac{X_i}{X_{(n)}}\right)^\alpha \log X_i}{\sum_{i=1}^n \left(\frac{X_i}{X_{(n)}}\right)^\alpha} = \log X_{(n)} > \frac{1}{n}\sum_{i=1}^n \log X_i,$$

where $X_{(n)}$ is the largest order statistic and the inequality holds as long as X_i's are not identical, we conclude that the likelihood equations have a unique solution. ∎

Exercise 49 (#4.106). Consider the one-way random effects model

$$X_{ij} = \mu + A_i + e_{ij}, \quad j = 1, ..., n, i = 1, ..., m,$$

where $\mu \in \mathcal{R}$, A_i's are independent and identically distributed as $N(0, \sigma_a^2)$, e_{ij}'s are independent and identically distributed as $N(0, \sigma^2)$, σ_a^2 and σ^2 are unknown, and A_i's and e_{ij}'s are independent.
(i) Find an MLE of (σ_a^2, σ^2) when $\mu = 0$.
(ii) Find an MLE of $(\mu, \sigma_a^2, \sigma^2)$.
Solution. (i) From the solution of Exercise 33 in Chapter 3, the likelihood function is

$$\ell(\mu, \sigma_a^2, \sigma^2) = \xi \exp\left\{-\frac{S_E}{2\sigma^2} - \frac{S_A}{2(\sigma^2 + n\sigma_a^2)} - \frac{nm}{2(\sigma^2 + n\sigma_a^2)} \sum_{i=1}^m (\bar{X}_{\cdot\cdot} - \mu)^2\right\},$$

where $S_A = n\sum_{i=1}^m (\bar{X}_{i\cdot} - \bar{X}_{\cdot\cdot})^2$, $S_E = \sum_{i=1}^m \sum_{j=1}^n (X_{ij} - \bar{X}_{i\cdot})^2$, $\bar{X}_{\cdot\cdot} = (nm)^{-1}\sum_{i=1}^m \sum_{j=1}^n X_{ij}$, $\bar{X}_{i\cdot} = n^{-1}\sum_{j=1}^n X_{ij}$, and ξ is a function of σ_a^2 and σ^2. In the case of $\mu = 0$, the likelihood function becomes

$$\ell(\sigma_a^2, \sigma^2) = \xi \exp\left\{-\frac{S_E}{2\sigma^2} - \frac{\tilde{S}_A}{2(\sigma^2 + n\sigma_a^2)}\right\},$$

where $\tilde{S}_A = n\sum_{i=1}^m \bar{X}_{i\cdot}^2$. The statistic $(\tilde{S}_A, S_E)$ is complete and sufficient for (σ_a^2, σ^2). Hence, $\ell(\sigma_a^2, \sigma^2)$ is proportional to the joint distribution of $\tilde{S}_A$ and S_E. Since X_{ij}'s are normal, S_E/σ^2 has the chi-square distribution $\chi^2_{m(n-1)}$ and $\tilde{S}_A/(\sigma^2 + n\sigma_a^2)$ has the chi-square distribution χ^2_m. Since the distribution of S_E does not depend on σ^2 and $\tilde{S}_A$ is complete and sufficient for σ_a^2 when σ^2 is known, by Basu's theorem, $\tilde{S}_A$ and S_E are independent. Therefore, the likelihood equations are

$$\frac{\partial \log \ell(\sigma_a^2, \sigma^2)}{\partial \sigma_a^2} = \frac{n\tilde{S}_A}{\sigma^2 + n\sigma_a^2} - \frac{nm}{\sigma^2 + n\sigma_a^2} = 0$$

and

$$\frac{\partial \log \ell(\sigma_a^2, \sigma^2)}{\partial \sigma^2} = \frac{n\tilde{S}_A}{\sigma^2 + n\sigma_a^2} - \frac{nm}{\sigma^2 + n\sigma_a^2} + \frac{S_E}{\sigma^4} - \frac{m(n-1)}{\sigma^2} = 0.$$

A unique solution is

$$\hat{\sigma}^2 = \frac{S_E}{m(n-1)} \quad \text{and} \quad \hat{\sigma}_a^2 = \frac{\tilde{S}_A}{nm} - \frac{S_E}{nm(n-1)}.$$

If $\hat{\sigma}_a^2 > 0$, then the MLE of (σ_a^2, σ^2) is $(\hat{\sigma}_a^2, \hat{\sigma}^2)$. If $\hat{\sigma}_a^2 \leq 0$, however, the maximum of $\ell(\sigma_a^2, \sigma^2)$ is achieved at the boundary of the parameter space when $\sigma_a^2 = 0$. Note that

$$\frac{\partial \log \ell(0, \sigma^2)}{\partial \sigma^2} = \frac{n\tilde{S}_A}{\sigma^2} - \frac{nm}{\sigma^2} + \frac{S_E}{\sigma^4} - \frac{m(n-1)}{\sigma^2} = 0$$

has a unique solution $\tilde{\sigma}^2 = (n\tilde{S}_A + S_E)/[nm + m(n-1)]$. Thus, the MLE of (σ_a^2, σ^2) is $(0, \tilde{\sigma}^2)$ when $\hat{\sigma}_a^2 \leq 0$.

(ii) It is easy to see that $\bar{X}_{\cdot\cdot}$ maximizes $\ell(\mu, \sigma_a^2, \sigma^2)$ for any σ_a^2 and σ^2. Hence, $\bar{X}_{\cdot\cdot}$ is the MLE of μ. To consider the MLE of (σ_a^2, σ^2), it suffices to consider

$$\ell(\bar{X}_{\cdot\cdot}, \sigma_a^2 \sigma^2) = \xi \exp\left\{ -\frac{S_E}{2\sigma^2} - \frac{S_A}{2(\sigma^2 + n\sigma_a^2)} - \frac{nm}{2(\sigma^2 + n\sigma_a^2)} \right\}.$$

Note that this is the same as $\ell(\sigma_a^2, \sigma^2)$ in the solution of part (i) with $\tilde{S}_A$ replaced by S_A and $S_A/(\sigma^2 + n\sigma_a^2)$ has the chi-square distribution χ^2_{m-1}. Using the same argument in part (i) of the solution, we conclude that the MLE of (σ_a^2, σ^2) is $(\hat{\sigma}_a^2, \hat{\sigma}^2)$ if $\hat{\sigma}_a^2 > 0$, where

$$\hat{\sigma}^2 = \frac{S_E}{m(n-1)} \quad \text{and} \quad \hat{\sigma}_a^2 = \frac{S_A}{n(m-1)} - \frac{S_E}{nm(n-1)},$$

and is $(0, \tilde{\sigma}^2)$ if $\hat{\sigma}_a^2 \leq 0$, where $\tilde{\sigma}^2 = (nS_A + S_E)/[n(m-1) + m(n-1)]$. ∎

Exercise 50 (#4.107). Let $(X_1, ..., X_n)$ be a random sample of random variables with Lebesgue density $\theta f(\theta x)$, where f is a Lebesgue density on $(0, \infty)$ or symmetric about 0, and $\theta > 0$ is an unknown parameter. Show that the likelihood equation has a unique root if $xf'(x)/f(x)$ is continuous and decreasing for $x > 0$. Verify that this condition is satisfied if $f(x) = \pi^{-1}(1 + x^2)^{-1}$.

Solution. Let $\ell(\theta)$ be the likelihood function and

$$h(\theta) = \sum_{i=1}^n \left[1 + \frac{\theta X_i f'(\theta X_i)}{f(\theta X_i)} \right].$$

Then

$$\frac{\partial \log \ell(\theta)}{\partial \theta} = \frac{1}{\theta} \sum_{i=1}^n \left[1 + \frac{\theta X_i f'(\theta X_i)}{f(\theta X_i)} \right] = 0$$

is the same as $h(\theta) = 0$. From the condition, $h(\theta)$ is decreasing in θ when $\theta > 0$. Hence, the likelihood equation has at most one solution. Define $g(t) = 1 + tf'(t)/f(t)$. Suppose that $g(t) \geq 0$ for all $t \in (0, \infty)$. Then $tf(t)$ is nondecreasing since its derivative is $f(t)g(t) \geq 0$. Let $t_0 \in (0, \infty)$. Then $tf(t) \geq t_0 f(t_0)$ for $t \in (t_0, \infty)$ and

$$1 \geq \int_{t_0}^{\infty} f(t)dt \geq \int_{t_0}^{\infty} \frac{t_0 f(t_0)}{t} dt = \infty,$$

which is impossible. Suppose that $g(t) \leq 0$ for all $t \in (0, \infty)$. Then $tf(t)$ is nonincreasing and $tf(t) \geq t_0 f(t_0)$ for $t \in (0, t_0)$. Then

$$1 \geq \int_0^{t_0} f(t)dt \geq \int_0^{t_0} \frac{t_0 f(t_0)}{t} dt = \infty,$$

which is impossible. Combining these results and the fact that $g(t)$ is nonincreasing, we conclude that

$$\lim_{\theta \to 0} \sum_{i=1}^n \left[1 + \frac{\theta X_i f'(\theta X_i)}{f(\theta X_i)} \right] > 0 > \lim_{\theta \to \infty} \sum_{i=1}^n \left[1 + \frac{\theta X_i f'(\theta X_i)}{f(\theta X_i)} \right]$$

and, therefore, $h(\theta) = 0$ has a unique solution.

For $f(x) = \pi^{-1}(1 + x^2)^{-1}$,

$$\frac{xf'(x)}{f(x)} = \frac{2x^2}{1 + x^2},$$

which is clearly continuous and decreasing for $x > 0$. ∎

Exercise 51 (#4.108). Let $(X_1, ..., X_n)$ be a random sample having Lebesgue density $f_\theta(x) = \theta f_1(x) + (1-\theta) f_2(x)$, where f_j's are two different known Lebesgue densities and $\theta \in (0, 1)$ is unknown.
(i) Provide a necessary and sufficient condition for the likelihood equation to have a unique solution and show that if there is a solution, it is the MLE of θ.
(ii) Derive the MLE of θ when the likelihood equation has no solution.
Solution. (i) Let $\ell(\theta)$ be the likelihood function. Note that

$$s(\theta) = \frac{\partial \log \ell(\theta)}{\partial \theta} = \sum_{i=1}^n \frac{f_1(X_i) - f_2(X_i)}{f_2(X_i) + \theta[f_1(X_i) - f_2(X_i)]},$$

which has derivative

$$s'(\theta) = -\sum_{i=1}^n \frac{[f_1(X_i) - f_2(X_i)]^2}{\{f_2(X_i) + \theta[f_1(X_i) - f_2(X_i)]\}^2} < 0.$$

Therefore, $s(\theta) = 0$ has at most one solution. The necessary and sufficient condition that $s(\theta) = 0$ has a solution (which is unique if it exists) is that $\lim_{\theta \to 0} s(\theta) > 0$ and $\lim_{\theta \to 1} s(\theta) < 0$, which is equivalent to

$$\sum_{i=1}^n \frac{f_1(X_i)}{f_2(X_i)} > n \quad \text{and} \quad \sum_{i=1}^n \frac{f_2(X_i)}{f_1(X_i)} > n.$$

The solution, if it exists, is the MLE since $s'(\theta) < 0$.
(ii) If $\sum_{i=1}^n \frac{f_2(X_i)}{f_1(X_i)} \leq n$, then $s(\theta) \geq 0$ and $\ell(\theta)$ is nondecreasing and, thus, the MLE of θ is 1. Similarly, if $\sum_{i=1}^n \frac{f_1(X_i)}{f_2(X_i)} \leq n$, then the MLE of θ is 0. ∎

Exercise 52 (#4.111). Let X_{ij}, $j = 1, ..., r > 1$, $i = 1, ..., n$, be independently distributed as $N(\mu_i, \sigma^2)$. Find the MLE of $\theta = (\mu_1, ..., \mu_n, \sigma^2)$. Show that the MLE of σ^2 is not a consistent estimator (as $n \to \infty$).
Solution. Let $\ell(\theta)$ be the likelihood function. Note that

$$\log \ell(\theta) = -\frac{nr}{2}\log(2\pi\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}\sum_{j=1}^{r}(X_{ij} - \mu_i)^2,$$

$$\frac{\partial \log \ell(\theta)}{\partial \mu_i} = \frac{1}{\sigma^2}\sum_{j=1}^{r}(X_{ij} - \mu_i),$$

and

$$\frac{\partial \log \ell(\theta)}{\partial \sigma^2} = -\frac{nr}{2\sigma^2} + \frac{1}{2\sigma^4}\sum_{i=1}^{n}\sum_{j=1}^{r}(X_{ij} - \mu_i)^2.$$

Hence, the MLE of μ_i is $\bar{X}_{i\cdot} = r^{-1}\sum_{j=1}^{r} X_{ij}$, $i = 1, ..., n$, and the MLE of σ^2 is

$$\hat{\sigma}^2 = \frac{1}{nr}\sum_{i=1}^{n}\sum_{j=1}^{r}(X_{ij} - \bar{X}_{i\cdot})^2.$$

Since

$$E\left[\sum_{j=1}^{r}(X_{ij} - \bar{X}_{i\cdot})^2\right] = (r-1)\sigma^2,$$

by the law of large numbers,

$$\hat{\sigma}^2 \to_p \frac{r-1}{r}\sigma^2$$

as $n \to \infty$. Hence, $\hat{\sigma}^2$ is inconsistent. ∎

Exercise 53 (#4.112). Let $(X_1, ..., X_n)$ be a random sample from the uniform distribution on $(0, \theta)$, where $\theta > 0$ is unknown. Let $\hat{\theta}$ be the MLE of θ and T be the UMVUE.
(i) Obtain the ratio of the mean squared error of T over the mean squared error of $\hat{\theta}$ and show that the MLE is inadmissible when $n \geq 2$.
(ii) Let $Z_{a,\theta}$ be a random variable having the exponential distribution on (a, ∞) with scale parameter θ. Prove $n(\theta - \hat{\theta}) \to_d Z_{0,\theta}$ and $n(\theta - T) \to_d Z_{-\theta,\theta}$. Obtain the asymptotic relative efficiency of $\hat{\theta}$ with respect to T.
Solution. (i) Let $X_{(n)}$ be the largest order statistic. Then $\hat{\theta} = X_{(n)}$ and $T(X) = \frac{n+1}{n}X_{(n)}$. The mean squared error of $\hat{\theta}$ is

$$E(X_{(n)} - \theta)^2 = \frac{2\theta^2}{(n+1)(n+2)}$$

and the mean squared error of T is

$$E(T-\theta)^2 = \frac{\theta^2}{n(n+2)}.$$

The ratio is $(n+1)/(2n)$. When $n \geq 2$, this ratio is less than 1 and, therefore, the MLE $\hat{\theta}$ is inadmissible.
(ii) From

$$\begin{aligned} P\left(n(\theta-\hat{\theta}) \leq x\right) &= P\left(X_{(n)} \geq \theta - \frac{x}{n}\right) \\ &= \theta^{-n} \int_{\theta-x/n}^{\theta} nt^{n-1} dt \\ &= 1 - \left(1 - \frac{x}{n\theta}\right)^n \\ &\to 1 - e^{-x/\theta} \end{aligned}$$

as $n \to \infty$, we conclude that $n(\theta-\hat{\theta}) \to_d Z_{0,\theta}$. From

$$n(\theta - T) = n(\theta - \hat{\theta}) - \hat{\theta}$$

and Slutsky's theorem, we conclude that $n(\theta - T) \to_d Z_{0,\theta} - \theta$, which has the same distribution as $Z_{-\theta,\theta}$. The asymptotic relative efficiency of $\hat{\theta}$ with respect to T is $E(Z^2_{-\theta,\theta})/E(Z^2_{0,\theta}) = \theta^2/(\theta^2+\theta^2) = \frac{1}{2}$. ∎

Exercise 54 (#4.113). Let $(X_1, ..., X_n)$ be a random sample from the exponential distribution on (a, ∞) with scale parameter θ, where $a \in \mathcal{R}$ and $\theta > 0$ are unknown. Obtain the asymptotic relative efficiency of the MLE of a (or θ) with respect to the UMVUE of a (or θ).
Solution. Let $X_{(1)}$ be the smallest order statistic. From Exercise 6 in Chapter 3, the UMVUE of a and θ are, respectively,

$$\tilde{a} = X_{(1)} - \frac{1}{n(n-1)} \sum_{i=1}^{n} (X_i - X_{(1)})$$

and

$$\tilde{\theta} = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - X_{(1)}).$$

From Exercise 41(ix), the MLE of (a, θ) is $(\hat{a}, \hat{\theta})$, where

$$\hat{a} = X_{(1)} \quad \text{and} \quad \hat{\theta} = \frac{1}{n} \sum_{i=1}^{n} (X_i - X_{(1)}).$$

From part (iii) of the solution of Exercise 7 in Chapter 2, $2(n-1)\tilde{\theta}/\theta$ has the chi-square distribution $\chi^2_{2(n-1)}$. Hence,

$$\sqrt{2(n-1)}\left(\frac{\tilde{\theta}}{\theta}-1\right) \to_d N(0,1),$$

i.e.,

$$\sqrt{n}\left(\tilde{\theta}-\theta\right) \to_d N(0,2\theta^2).$$

Since $\hat{\theta} = \frac{n-1}{n}\tilde{\theta}$, $\hat{\theta}$ has the same asymptotic distribution as $\tilde{\theta}$ and the asymptotic relative efficiency of $\hat{\theta}$ with respect to $\tilde{\theta}$ is 1.

Note that $n(\hat{a}-a) = n(X_{(1)}-a)$ has the same distribution as Z, where Z is a random variable having the exponential distribution on $(0,\infty)$ with scale parameter θ. Then

$$n(\tilde{a}-a) = n(X_{(1)}-a) - \frac{1}{n-1}\sum_{i=1}^n (X_i - X_{(1)}) \to_d Z-\theta,$$

since $\frac{1}{n-1}\sum_{i=1}^n (X_i - X_{(1)}) \to_p \theta$. Therefore, the asymptotic relative efficiency of $\hat{a}$ with respect to $\tilde{a}$ is $E(Z-\theta)^2/E(Z^2) = \frac{1}{2}$. ∎

Exercise 55 (#4.115). Let $(X_1,...,X_n)$, $n \geq 2$, be a random sample from a distribution having Lebesgue density $f_{\theta,j}$, where $\theta > 0$, $j = 1,2$, $f_{\theta,1}$ is the density of $N(0,\theta^2)$, and $f_{\theta,2}(x) = (2\theta)^{-1}e^{-|x|/\theta}$.
(i) Obtain an MLE of (θ, j).
(ii) Show whether the MLE of j in part (i) is consistent.
(iii) Show that the MLE of θ is consistent and derive its nondegenerated asymptotic distribution.

Solution. (i) Let $T_1 = \sum_{i=1}^n X_i^2$ and $T_2 = \sum_{i=1}^n |X_i|$. The likelihood function is

$$\ell(\theta,j) = \begin{cases} (2\pi)^{-n/2}\theta^{-n}e^{-T_1/(2\theta^2)} & j=1 \\ 2^{-n}\theta^{-n}e^{-T_2/\theta} & j=2. \end{cases}$$

Note that $\hat{\theta}_1 = \sqrt{T_1/n}$ maximizes $\ell(\theta,1)$ and $\hat{\theta}_2 = T_2/n$ maximizes $\ell(\theta,2)$. Define

$$\hat{j} = \begin{cases} 1 & \ell(\hat{\theta}_1,1) \geq \ell(\hat{\theta}_2,2) \\ 2 & \ell(\hat{\theta}_1,1) < \ell(\hat{\theta}_2,2), \end{cases}$$

which is the same as

$$\hat{j} = \begin{cases} 1 & \frac{\hat{\theta}_1}{\hat{\theta}_2} \leq \sqrt{\frac{2e}{\pi}} \\ 2 & \frac{\hat{\theta}_1}{\hat{\theta}_2} > \sqrt{\frac{2e}{\pi}}, \end{cases}$$

and define

$$\hat{\theta} = \begin{cases} \hat{\theta}_1 & \hat{j} = 1 \\ \hat{\theta}_2 & \hat{j} = 2. \end{cases}$$

Then

$$\ell(\hat{\theta}, \hat{j}) \geq \ell(\theta, j)$$

for any θ and j. This shows that $(\hat{\theta}, \hat{j})$ is an MLE of (θ, j).
(ii) The consistency of $\hat{j}$ means that

$$\lim_n P(\hat{j} = j) = 1,$$

which is equivalent to

$$\lim_n P\left(\frac{\hat{\theta}_1}{\hat{\theta}_2} \leq \sqrt{\frac{2e}{\pi}}\right) = \begin{cases} 1 & \text{if } j = 1 \\ 0 & \text{if } j = 2. \end{cases}$$

It suffices to consider the limit of $\hat{\theta}_1/\hat{\theta}_2$. When $j = 1$, $\hat{\theta}_1 \to_p \theta$ and $\hat{\theta}_2 \to_p \sqrt{\frac{2}{\pi}}\theta$ (since $E|X_1| = \sqrt{\frac{2}{\pi}}\theta$). Then $\hat{\theta}_1/\hat{\theta}_2 \to_p \sqrt{\frac{\pi}{2}} < \sqrt{\frac{2e}{\pi}}$ (since $\pi^2 < 4e$). When $j = 2$, $\hat{\theta}_1 \to_p \sqrt{2}\theta$ and $\hat{\theta}_2 \to_p \theta$. Then $\hat{\theta}_1/\hat{\theta}_2 \to_p \sqrt{2} > \sqrt{\frac{2e}{\pi}}$ (since $e < \pi$). Therefore, $\hat{j}$ is consistent.
(iii) When $j = 1$, by the result in part (ii) of the solution,

$$\lim_n P(\hat{\theta} = \hat{\theta}_1) = 1.$$

Hence, the asymptotic distribution of $\hat{\theta}$ is the same as that of $\hat{\theta}_1$ under the normal distribution assumption. By the central limit theorem and the δ-method, $\sqrt{n}(\hat{\theta}_1 - \theta) \to_d N(0, \theta^2/2)$. Similarly, when $j = 2$, the asymptotic distribution of $\hat{\theta}$ is the same as that of $\hat{\theta}_2$. By the central limit theorem, $\sqrt{n}(\hat{\theta}_2 - \theta) \to_d N(0, \theta^2)$. ∎

Exercise 56 (#4.115). Let $(X_1, ..., X_n)$, $n \geq 2$, be a random sample from a distribution with discrete probability density $f_{\theta,j}$, where $\theta \in (0, 1)$, $j = 1, 2$, $f_{\theta,1}$ is the Poisson distribution with mean θ, and $f_{\theta,2}$ is the binomial distribution with size 1 and probability θ.
(i) Obtain an MLE of (θ, j).
(ii) Show whether the MLE of j in part (i) is consistent.
(iii) Show that the MLE of θ is consistent and derive its nondegenerated asymptotic distribution.
Solution. (i) Let $X = (X_1, ..., X_n)$, $\bar{X}$ be the sample mean, $g(X) = (\prod_{i=1}^n X_i!)^{-1}$, and $h(X) = 1$ if all X_i's are not larger than 1 and $h(X) = 0$ otherwise. The likelihood function is

$$\ell(\theta, j) = \begin{cases} e^{-n\theta}\theta^{n\bar{X}} g(X) & j = 1 \\ \theta^{n\bar{X}}(1 - \theta)^{n - n\bar{X}} h(X) & j = 2. \end{cases}$$

Note that $\bar{X} = T/n$ maximizes both $\ell(\theta, 1)$ and $\ell(\theta, 2)$. Define

$$\hat{j} = \begin{cases} 1 & \ell(\bar{X}, 1) > \ell(\bar{X}, 2) \\ 2 & \ell(\bar{X}, 1) \leq \ell(\bar{X}, 2). \end{cases}$$

Then

$$\ell(\bar{X}, \hat{j}) \geq \ell(\theta, j)$$

for any θ and j and, hence, $(\bar{X}, \hat{j})$ is an MLE of (θ, j). We now simplify the formula for $\hat{j}$. If at least one X_i is larger than 1, then $h(X) = 0$, $\ell(\bar{X}, 1) > \ell(\bar{X}, 2)$, and $\hat{j} = 1$. If all X_i's are not larger than 1, then $h(X) = g(X) = 1$ and

$$\frac{\ell(\bar{X}, 2)}{\ell(\bar{X}, 1)} = (1 - \bar{X})^{n - n\bar{X}} e^{n\bar{X}} \geq 1$$

because of the inequality $(1-t)^{1-t} \geq e^{-t}$ for any $t \in [0, 1)$. This shows that

$$\hat{j} = \begin{cases} 1 & h(X) = 0 \\ 2 & h(X) = 1. \end{cases}$$

(ii) If $j = 2$, $h(X) = 1$ always holds. Therefore, $\hat{j}$ is consistent if we can show that if $j = 1$, $\lim_n P(h(X) = 1) = 0$. Since $P(X_1 = 0) = e^{-\theta}$ and $P(X_1 = 1) = e^{-\theta}\theta$,

$$\begin{aligned} P(h(X) = 1) &= \sum_{k=0}^{n} \binom{n}{k} (e^{-\theta}\theta)^k (e^{-\theta})^{n-k} \\ &= \sum_{k=0}^{n} \binom{n}{k} \theta^k e^{-n\theta} \\ &\leq \sum_{k=0}^{n} \binom{n}{k} \theta^k (1-\theta)^{n(1-\theta)}. \end{aligned}$$

For any fixed $\theta \in (0, 1)$ and any $\epsilon > 0$, there exists K such that $(1-\theta)^{k-\theta} < \epsilon$ whenever $k \geq K$. Then, when $n > K$,

$$\begin{aligned} P(h(X) = 1) &\leq \sum_{k=K}^{n} \binom{n}{k} \theta^k (1-\theta)^{n(1-\theta)} + \sum_{k=0}^{K} \binom{n}{k} \theta^k (1-\theta)^{n(1-\theta)} \\ &\leq \epsilon + \sum_{k=0}^{K} \binom{n}{k} \theta^k (1-\theta)^{n(1-\theta)}. \end{aligned}$$

For any fixed K,

$$\lim_n \sum_{k=0}^{K} \binom{n}{k} \theta^k (1-\theta)^{n(1-\theta)} = 0.$$

Hence, $\limsup_n P(h(X) = 1) \leq \epsilon$ and the result follows since ϵ is arbitrary.
(iii) By the central limit theorem, $\sqrt{n}(\bar{X} - \theta) \to_d N(0, \mathrm{Var}(X_1))$, since $E(\bar{X}) = \theta$ in any case. When $j = 1$, $\mathrm{Var}(X_1) = \theta$ and when $j = 2$, $\mathrm{Var}(X_1) = \theta(1-\theta)$. ∎

Exercise 57 (#4.122). Let $\hat{\theta}_n$ be an estimator of $\theta \in \mathcal{R}$ satisfying $\sqrt{n}(\hat{\theta}_n - \theta) \to_d N(0, v(\theta))$ as the sample size $n \to \infty$. Construct an estimator $\tilde{\theta}_n$ such that $\sqrt{n}(\tilde{\theta}_n - \theta) \to_d N(0, w(\theta))$ with $w(\theta) = v(\theta)$ for $\theta \neq \theta_0$ and $w(\theta_0) = t^2 v(\theta_0)$, where $t \in \mathcal{R}$ and θ_0 is a point in the parameter space.
Note. This is a generalized version of Hodges' superefficiency example (e.g., Example 4.38 in Shao, 2003).
Solution. Consider

$$\tilde{\theta}_n = \begin{cases} \hat{\theta}_n & \text{if } |\hat{\theta}_n - \theta_0| \geq n^{-1/4} \\ t\hat{\theta}_n + (1-t)\theta_0 & \text{if } |\hat{\theta}_n - \theta_0| < n^{-1/4}. \end{cases}$$

We now show that $\tilde{\theta}_n$ has the desired property. If $\theta \neq \theta_0$, then $\hat{\theta}_n - \theta_0 \to_p \theta - \theta_0 \neq 0$ and, therefore, $\lim_n P(|\hat{\theta}_n - \theta_0| < n^{-1/4}) = 0$. On the event $\{|\hat{\theta}_n - \theta| \geq n^{-1/4}\}$, $\tilde{\theta}_n = \hat{\theta}_n$. Hence the asymptotic distribution of $\sqrt{n}(\tilde{\theta}_n - \theta)$ is the same as that of $\sqrt{n}(\hat{\theta}_n - \theta)$ when $\theta \neq \theta_0$.

Consider now $\theta = \theta_0$. Then

$$\begin{aligned} \lim_n P(|\hat{\theta}_n - \theta_0| < n^{-1/4}) &= \lim_n P(\sqrt{n}|\hat{\theta}_n - \theta_0| < n^{1/4}) \\ &= \lim_n [\Phi(n^{1/4}) - \Phi(-n^{1/4})] \\ &= 1, \end{aligned}$$

where Φ is the cumulative distribution function of $N(0,1)$. On the event $\{|\hat{\theta}_n - \theta| < n^{-1/4}\}$, $\sqrt{n}(\tilde{\theta}_n - \theta_0) = \sqrt{n}t(\hat{\theta}_n - \theta_0) \to_d N(0, t^2 v(\theta_0))$. ∎

Exercise 58 (#4.123). Let $(X_1, ..., X_n)$ be a random sample from a distribution with probability density f_θ with respect to a σ-finite measure ν on $(\mathcal{R}, \mathcal{B})$, where $\theta \in \Theta$ and Θ is an open set in $\mathcal{R}$. Suppose that for every x in the range of X_1, $f_\theta(x)$ is twice continuously differentiable in θ and satisfies

$$\frac{\partial}{\partial \theta} \int \psi_\theta(x) d\nu = \int \frac{\partial}{\partial \theta} \psi_\theta(x) d\nu$$

for $\psi_\theta(x) = f_\theta(x)$ and $= \partial f_\theta(x)/\partial\theta$; the Fisher information

$$I_1(\theta) = E\left[\frac{\partial}{\partial \theta} \log f_\theta(X_1)\right]^2$$

is finite and positive; and for any given $\theta \in \Theta$, there exists a positive number c_θ and a positive function h_θ such that $E[h_\theta(X_1)] < \infty$ and

$\sup_{|\gamma-\theta|<c_\theta} \left|\frac{\partial^2 \log f_\gamma(x)}{\partial\gamma^2}\right| \leq h_\theta(x)$ for all x in the range of X_1. Show that

$$\frac{\log \ell(\theta+n^{-1/2}) - \log \ell(\theta) + I_1(\theta)/2}{\sqrt{I_1(\theta)}} \to_d N(0,1),$$

where $\ell(\theta)$ is the likelihood function.
Solution. Let $L(\gamma) = \log \ell(\gamma)$. By Taylor's expansion,

$$L(\theta + n^{-1/2}) - L(\theta) = n^{-1/2}L'(\theta) + (2n)^{-1}L''(\xi_n),$$

where ξ_n satisfies $|\xi_n - \theta| \leq n^{-1/2}$. Let $f'_\theta(x) = \partial f_\theta(x)/\partial\theta$. Then

$$L'(\theta) = \sum_{i=1}^n \frac{f'_\theta(X_i)}{f_\theta(X_i)}.$$

Note that $\frac{f'_\theta(X_i)}{f_\theta(X_i)}$, $i = 1, ..., n$, are independent and identically distributed with

$$E\frac{f'_\theta(X_i)}{f_\theta(X_i)} = 0 \quad \text{and} \quad \text{Var}\left(\frac{f'_\theta(X_i)}{f_\theta(X_i)}\right) = I_1(\theta)$$

(under the given condition). Hence, by the central limit theorem,

$$n^{-1/2}L'(\theta) \to_d N(0, I_1(\theta)).$$

By the law of large numbers and the given condition,

$$n^{-1}L''(\theta) = \frac{1}{n}\sum_{i=1}^n \frac{\partial^2 \log f_\theta(X_i)}{\partial\theta^2} \to_p E\left(\frac{\partial^2 \log f_\theta(X_1)}{\partial\theta^2}\right) = -I_1(\theta).$$

Since $\partial^2 \log f_\theta(x)/\partial\theta^2$ is continuous in θ, $n^{-1}[L''(\xi_n) - L''(\theta)] \to 0$ for any fixed $X_1, X_2, ...$, i.e., $n^{-1}L''(\xi_n) \to_p -I_1(\theta)$ for any fixed $X_1, X_2, ...$ Under the given condition,

$$n^{-1}|L''(\xi_n)| \leq n^{-1}\sum_{i=1}^n h_\theta(X_i) \to_p E[h_\theta(X_1)].$$

Hence, by the dominated convergence theorem, $n^{-1}E[L''(\xi_n) - L''(\theta)] \to 0$, which implies that $n^{-1}L''(\xi_n) \to_p -I_1(\theta)$. Then, the result follows from Slutsky's theorem. ∎

Exercise 59 (#4.124). Let $(X_1, ..., X_n)$ be a random sample from $N(\mu, 1)$ truncated at two known points $\alpha < \beta$, i.e., the Lebesgue density of X_1 is

$$\{\sqrt{2\pi}[\Phi(\beta-\mu) - \Phi(\alpha-\mu)]\}^{-1}e^{-(x-\mu)^2/2}I_{(\alpha,\beta)}(x),$$

where Φ is the cumulative distribution function of $N(0,1)$. Show that the sample mean $\bar{X}$ is the unique MLE of $\theta = EX_1$ and is asymptotically

efficient.
Solution. The log-likelihood function is

$$\log \ell(\mu) = -n \log(\Phi(\beta-\mu) - \Phi(\alpha-\mu)) - \frac{1}{2}\sum_{i=1}^{n}(X_i-\mu)^2.$$

Hence,

$$\frac{\partial \log \ell(\mu)}{\partial \mu} = n[\bar{X} - g(\mu)],$$

where

$$g(\mu) = \mu - \frac{\Phi'(\beta-\mu) - \Phi'(\alpha-\mu)}{\Phi(\beta-\mu) - \Phi(\alpha-\mu)}.$$

Since the inverse function g^{-1} exists, $g^{-1}(\bar{X})$ is the unique solution of the likelihood equation and, hence, it is the unique MLE of μ. Note that

$$\Phi(\beta-\mu) - \Phi(\alpha-\mu) = \frac{1}{\sqrt{2\pi}}\int_\alpha^\beta e^{-(x-\mu)^2/2}dx,$$

$$-[\Phi'(\beta-\mu) - \Phi'(\alpha-\mu)] = \frac{1}{\sqrt{2\pi}}\int_\alpha^\beta (x-\mu)e^{-(x-\mu)^2/2}dx,$$

and

$$\begin{aligned}\theta &= \frac{1}{\sqrt{2\pi}[\Phi(\beta-\mu) - \Phi(\alpha-\mu)]}\int_\alpha^\beta xe^{-(x-\mu)^2/2}dx \\ &= \mu + \frac{1}{\sqrt{2\pi}[\Phi(\beta-\mu) - \Phi(\alpha-\mu)]}\int_\alpha^\beta (x-\mu)e^{-(x-\mu)^2/2}dx \\ &= g(\mu).\end{aligned}$$

Hence, $\bar{X}$ is the unique MLE of θ. By the asymptotic property of the MLE (e.g., Theorem 4.17 in Shao, 2003), $\bar{X}$ is asymptotically efficient. ∎

Exercise 60 (#4.127). Let $(X_1, ..., X_n)$ be a random sample such that $\log X_i$ is distributed as $N(\theta, \theta)$ with an unknown $\theta > 0$. Show that one of the solutions of the likelihood equation is the unique MLE of θ. Obtain the asymptotic distribution of the MLE of θ.
Solution. Let $Y_i = \log X_i$, $T = n^{-1}\sum_{i=1}^n Y_i^2$, and $\ell(\theta)$ be the likelihood function. Then

$$\log \ell(\theta) = -\frac{n\log\theta}{2} - \frac{1}{2\theta}\sum_{i=1}^{n}(Y_i-\theta)^2,$$

$$\frac{\partial \log \ell(\theta)}{\partial \theta} = \frac{n}{2}\left(\frac{T}{\theta^2} - \frac{1}{\theta} - 1\right)$$

and

$$\frac{\partial^2 \log \ell(\theta)}{\partial \theta^2} = \frac{n}{2\theta^2} - \frac{nT}{\theta^3}.$$

The likelihood equation $\frac{\partial \log \ell(\theta)}{\partial \theta} = 0$ has two solutions

$$\frac{\pm\sqrt{1+4T} - 1}{2}.$$

At each solution, $T = \theta + \theta^2$ and $\frac{\partial^2 \log \ell(\theta)}{\partial \theta^2} = -n(\frac{1}{2\theta^2} + \frac{1}{\theta}) < 0$. Hence, both solutions are maximum points of the likelihood function. Since $\theta > 0$, the only positive solution, $\hat{\theta} = (\sqrt{1+4T} - 1)/2$, is the unique MLE of θ. Since $EY_1^2 = \theta + \theta^2$, the Fisher information is

$$I_n(\theta) = -E\left(\frac{\partial^2 \log \ell(\theta)}{\partial \theta^2}\right) = E\left(\frac{n}{2\theta^2} - \frac{nT}{\theta^3}\right) = \frac{(2\theta+1)n}{2\theta^2}.$$

Thus, $\sqrt{n}(\hat{\theta} - \theta) \to_d N(0, 2\theta^2/(2\theta+1))$. ∎

Exercise 61 (#4.131). Let $(X_1, ..., X_n)$ be a random sample from the distribution $P(X_1 = 0) = 6\theta^2 - 4\theta + 1$, $P(X_1 = 1) = \theta - 2\theta^2$, and $P(X_1 = 2) = 3\theta - 4\theta^2$, where $\theta \in (0, \frac{1}{2})$ is unknown. Obtain the asymptotic distribution of an RLE (root of likelihood equation) of θ.

Solution. Let Y be the number of X_i's that are 0 and Z be the number of X_i's that are 1. Then, the likelihood function is

$$\ell(\theta) = (6\theta^2 - 4\theta + 1)^Y (\theta - 2\theta^2)^Z (3\theta - 4\theta^2)^{n-Y-Z},$$

$$\frac{\partial \log \ell(\theta)}{\partial \theta} = \frac{(12\theta - 4)Y}{6\theta^2 - 4\theta + 1} + \frac{(1-4\theta)Z}{\theta - 2\theta^2} + \frac{(3 - 8\theta)(n - Y - Z)}{3\theta - 4\theta^2},$$

and

$$\frac{\partial^2 \log \ell(\theta)}{\partial \theta^2} = -\frac{(72\theta^2 - 48\theta + 4)Y}{(6\theta^2 - 4\theta + 1)^2} - \frac{(8\theta^2 - 4\theta + 1)Z}{(\theta - 2\theta^2)^2} - \frac{(32\theta^2 - 24\theta + 9)(n - Y - Z)}{(3\theta - 4\theta^2)^2}.$$

By the theorem for RLE (e.g., Theorem 4.17 in Shao, 2003), there exists an RLE $\hat{\theta}$ such that $\sqrt{n}(\hat{\theta} - \theta) \to_d N(0, I_1^{-1}(\theta))$, where

$$\begin{aligned} I_1(\theta) &= -\frac{1}{n} E\left[\frac{\partial^2 \log \ell(\theta)}{\partial \theta^2}\right] \\ &= \frac{72\theta^2 - 48\theta + 4}{6\theta^2 - 4\theta + 1} + \frac{8\theta^2 - 4\theta + 1}{\theta - 2\theta^2} + \frac{32\theta^2 - 24\theta + 9}{3\theta - 4\theta^2}. \end{aligned}$$ ∎

Exercise 62 (#4.132). Let $(X_1, ..., X_n)$ be a random sample from $N(\mu, 1)$, where $\mu \in \mathcal{R}$ is unknown. Let $\theta = P(X_1 \leq c)$, where c is a known constant. Find the asymptotic relative efficiencies of the MLE of θ with respect to the UMVUE of θ and the estimator $n^{-1}\sum_{i=1}^n I_{(-\infty,c]}(X_i)$.
Solution. The MLE of μ is the sample mean $\bar{X}$. Let Φ be the cumulative distribution function of $N(0,1)$. Then $\theta = \Phi(c-\mu)$ and the MLE of θ is $\hat{\theta} = \Phi(c - \bar{X})$. From the central limit theorem and the δ-method, $\sqrt{n}(\hat{\theta} - \theta) \to_d N(0, [\Phi'(c-\mu)]^2)$. From Exercise 3(ii) in Chapter 3, the UMVUE of θ is $\tilde{\theta} = \Phi((c-\bar{X})/\sqrt{1-n^{-1}})$. Note that

$$\begin{aligned}
\sqrt{n}(\tilde{\theta} - \hat{\theta}) &= \sqrt{n}\left[\Phi\left(\frac{c-\bar{X}}{\sqrt{1-n^{-1}}}\right) - \Phi(c-\bar{X})\right] \\
&= \sqrt{n}\Phi'(\xi_n)(c-\bar{X})\left(\frac{1}{\sqrt{1-n^{-1}}} - 1\right) \\
&= \frac{\Phi'(\xi_n)(c-\bar{X})}{\sqrt{n}\sqrt{1-n^{-1}}(1+\sqrt{1-n^{-1}})} \\
&\to_p 0,
\end{aligned}$$

where ξ_n is a point between $c - \bar{X}$ and $(c-\bar{X})/\sqrt{1-n^{-1}}$. Hence, the asymptotic relative efficiency of $\hat{\theta}$ with respect to $\tilde{\theta}$ is 1. For the estimator $T = n^{-1}\sum_{i=1}^n I_{(-\infty,c]}(X_i)$, by the central limit theorem, $\sqrt{n}(T-\theta) \to_d N(0, \theta(1-\theta))$. Hence, the asymptotic relative efficiency of $\hat{\theta}$ with respect to T is $\theta(1-\theta)/[\Phi'(c-\mu)]^2$. ∎

Exercise 63 (#4.135). Let $(X_1, ..., X_n)$ be a random sample from a population having the Lebesgue density

$$f_{\theta_1,\theta_2}(x) = \begin{cases} (\theta_1+\theta_2)^{-1}e^{-x/\theta_1} & x > 0 \\ (\theta_1+\theta_2)^{-1}e^{x/\theta_2} & x \leq 0, \end{cases}$$

where $\theta_1 > 0$ and $\theta_2 > 0$ are unknown.
(i) Find the MLE of $\theta = (\theta_1, \theta_2)$.
(ii) Obtain a nondegenerated asymptotic distribution of the MLE of θ.
(iii) Obtain the asymptotic relative efficiencies of the MLE's with respect to the moment estimators.
Solution. (i) Let $T_1 = \sum_{i=1}^n X_i I_{(0,\infty)}(X_i)$ and $T_2 = -\sum_{i=1}^n X_i I_{(-\infty,0]}(X_i)$. Then, the log-likelihood function is

$$\log \ell(\theta) = -n\log(\theta_1+\theta_2) - \frac{T_1}{\theta_1} - \frac{T_2}{\theta_2},$$

$$\frac{\partial \log \ell(\theta)}{\partial \theta} = \left(-\frac{n}{\theta_1+\theta_2} + \frac{T_1}{\theta_1^2}, -\frac{n}{\theta_1+\theta_2} + \frac{T_2}{\theta_2^2}\right),$$

and

$$\frac{\partial^2 \log \ell(\theta)}{\partial \theta \partial \theta^\tau} = \begin{pmatrix} \frac{n}{(\theta_1+\theta_2)^2} - \frac{2T_1}{\theta_1^3} & \frac{n}{(\theta_1+\theta_2)^2} \\ \frac{n}{(\theta_1+\theta_2)^2} & \frac{n}{(\theta_1+\theta_2)^2} - \frac{2T_2}{\theta_2^3} \end{pmatrix}.$$

Since the likelihood equation has a unique solution, the MLE of θ is

$$\hat{\theta} = (\hat{\theta}_1, \hat{\theta}_2) = n^{-1}(\sqrt{T_1T_2} + T_1, \sqrt{T_1T_2} + T_2).$$

(ii) By the asymptotic theorem for MLE (e.g., Theorem 4.17 in Shao, 2003), $\sqrt{n}(\hat{\theta} - \theta) \to_d N_2(0, [I_1(\theta)]^{-1})$, where

$$I_1(\theta) = -\frac{1}{n}E\left[\frac{\partial^2 \log \ell(\theta)}{\partial\theta\partial\theta^\tau}\right] = \frac{1}{(\theta_1+\theta_2)^2}\begin{pmatrix} 1+\frac{2\theta_2}{\theta_1} & -1 \\ -1 & 1+\frac{2\theta_1}{\theta_2} \end{pmatrix}$$

and

$$[I_1(\theta)]^{-1} = \frac{(\theta_1+\theta_2)^2}{\left(1+\frac{2\theta_2}{\theta_1}\right)\left(1+\frac{2\theta_1}{\theta_2}\right)-1}\begin{pmatrix} 1+\frac{2\theta_1}{\theta_2} & 1 \\ 1 & 1+\frac{2\theta_2}{\theta_1} \end{pmatrix}.$$

(iii) Let $\mu_j = EX_1^j$, $j = 1, 2$. From Exercise 52 in Chapter 3,

$$\theta_1 + \theta_2 = \sqrt{2\mu_2 - 3\mu_1^2}$$

and

$$\frac{\theta_2}{\theta_1} = \frac{\sqrt{2\mu_2 - 3\mu_1^2} - \mu_1}{\sqrt{2\mu_2 - 3\mu_1^2} + \mu_1}.$$

Let $\tau_j(\mu_1, \mu_2)$ be the jth diagonal element of the matrix $\Lambda\Sigma\Lambda^\tau$ in the solution of Exercise 52 in Chapter 3. Then, the asymptotic relative efficiency of $\hat{\theta}_1$ with respect to the moment estimator of θ_1 given in the solution of Exercise 52 in Chapter 3 is

$$\frac{\tau_1(\mu_1,\mu_2)\left[\left(1+2\frac{\sqrt{2\mu_2-3\mu_1^2}-\mu_1}{\sqrt{2\mu_2-3\mu_1^2}+\mu_1}\right)\left(1+2\frac{\sqrt{2\mu_2-3\mu_1^2}+\mu_1}{\sqrt{2\mu_2-3\mu_1^2}-\mu_1}\right)-1\right]}{(2\mu_2-3\mu_1)\left(1+2\frac{\sqrt{2\mu_2-3\mu_1^2}+\mu_1}{\sqrt{2\mu_2-3\mu_1^2}-\mu_1}\right)}$$

and the asymptotic relative efficiency of $\hat{\theta}_2$ with respect to the moment estimator of θ_2 is

$$\frac{\tau_2(\mu_1,\mu_2)\left[\left(1+2\frac{\sqrt{2\mu_2-3\mu_1^2}-\mu_1}{\sqrt{2\mu_2-3\mu_1^2}+\mu_1}\right)\left(1+2\frac{\sqrt{2\mu_2-3\mu_1^2}+\mu_1}{\sqrt{2\mu_2-3\mu_1^2}-\mu_1}\right)-1\right]}{(2\mu_2-3\mu_1)\left(1+2\frac{\sqrt{2\mu_2-3\mu_1^2}-\mu_1}{\sqrt{2\mu_2-3\mu_1^2}+\mu_1}\right)}. \blacksquare$$

Exercise 64 (#4.136). In Exercise 47, show that the RLE $\hat{\rho}$ of ρ satisfies $\sqrt{n}(\hat{\rho} - \rho) \to_d N\big(0, (1-\rho^2)^2/(1+\rho^2)\big)$.

Solution. Let $\ell(\rho)$ be the likelihood function. From Exercise 47,

$$\frac{\partial \log \ell(\rho)}{\partial \rho} = n(1-\rho^2)^2 h(\rho),$$

where $h(\rho) = \rho(1-\rho^2) - n^{-1}T\rho + n^{-1}R(1+\rho^2)$, $T = \sum_{i=1}^n (X_i^2 + Y_i^2)$, and $R = \sum_{i=1}^n X_i Y_i$. Then,

$$\frac{\partial^2 \log \ell(\rho)}{\partial \rho^2} = \frac{nh'(\rho)}{(1-\rho^2)^2} + \frac{4n\rho h(\rho)}{(1-\rho^2)^3}.$$

Since $E[h(\rho)] = 0$ and $E[h'(\rho)] = -(1+\rho^2)$,

$$I_1(\rho) = -\frac{1}{n} E\left[\frac{\partial^2 \log \ell(\rho)}{\partial \rho^2}\right] = \frac{1+\rho^2}{(1-\rho^2)^2}.$$

By the asymptotic theorem for RLE, $\sqrt{n}(\hat{\rho} - \rho) \to_d N\big(0, [I_1(\rho)]^{-1}\big)$. The result follows. ∎

Exercise 65 (#4.137). In Exercise 50, obtain a nondegenerated asymptotic distribution of the RLE $\hat{\theta}$ of θ when $f(x) = \pi^{-1}(1+x^2)^{-1}$.
Solution. For $f(x) = \pi^{-1}(1+x^2)^{-1}$ and $f_\theta(x) = \theta f(\theta x)$,

$$\frac{\partial f_\theta(x)}{\partial \theta} = \frac{1-\theta^2x^2}{\pi(1+\theta^2x^2)^2} \quad \text{and} \quad \frac{\partial^2 f_\theta(x)}{\partial \theta^2} = \frac{2\theta x^2(\theta^2x^2-3)}{\pi(1+\theta^2x^2)^3},$$

which are continuous functions of θ. For any $\theta_0 \in (0,\infty)$,

$$\begin{aligned}\sup_{\theta\in[\theta_0/2,2\theta_0]} \left|\frac{\partial f_\theta(x)}{\partial \theta}\right| &= \sup_{\theta\in[\theta_0/2,2\theta_0]} \left|\frac{1-\theta^2x^2}{\pi(1+\theta^2x^2)^2}\right| \\ &\leq \sup_{\theta\in[\theta_0/2,2\theta_0]} \frac{1}{\pi(1+\theta^2x^2)} \\ &= \frac{1}{\pi[1+(\theta_0/2)^2x^2]},\end{aligned}$$

which is integrable with respect to the Lebesgue measure on $(-\infty,\infty)$. Also,

$$\begin{aligned}\sup_{\theta\in[\theta_0/2,2\theta_0]} \left|\frac{\partial^2 f_\theta(x)}{\partial^2 \theta}\right| &= \sup_{\theta\in[\theta_0/2,2\theta_0]} \left|\frac{2\theta x^2(\theta^2x^2-3)}{\pi(1+\theta^2x^2)^3}\right| \\ &\leq \sup_{\theta\in[\theta_0/2,2\theta_0]} \frac{2C}{\pi\theta(1+\theta^2x^2)} \\ &= \frac{2C}{\pi(\theta_0/2)[1+(\theta_0/2)^2x^2]},\end{aligned}$$

which is integrable, where C does not depend on θ. Therefore, by the dominated convergence theorem,

$$\frac{\partial}{\partial \theta}\int_{-\infty}^{\infty} \psi_\theta(x)dx = \int_{-\infty}^{\infty} \frac{\partial}{\partial \theta}\psi_\theta(x)dx$$

holds for both $\psi_\theta(x) = f_\theta(x)$ and $\frac{\partial f_\theta(x)}{\partial \theta}$. Note that

$$\begin{aligned}
\sup_{\theta\in[\theta_0/2,2\theta_0]} \left| \frac{\partial^2 \log f_\theta(x)}{\partial \theta^2} \right| &= \sup_{\theta\in[\theta_0/2,2\theta_0]} \left| \frac{1}{\theta^2} + \frac{2x^2(1-\theta^2x^2)}{(1+\theta^2x^2)^2} \right| \\
&\leq \sup_{\theta\in[\theta_0/2,2\theta_0]} \left| \frac{1}{\theta^2} + \frac{2x^2}{1+\theta^2x^2} \right| \\
&= \frac{4}{\theta_0^2} + \frac{2x^2}{1+(\theta_0/2)^2x^2}
\end{aligned}$$

and, when X_1 has density $\theta_0 f(\theta_0 x)$,

$$E\left[\frac{4}{\theta_0^2} + \frac{2X_1^2}{1+(\theta_0/2)^2X_1^2}\right] < \infty.$$

Therefore, the regularity conditions in Theorem 4.17 of Shao (2003) are satisfied and

$$\sqrt{n}(\hat{\theta} - \theta) \to_d N(0, [I_1(\theta)]^{-1}).$$

It remains to compute $I_1(\theta)$. Note that

$$\begin{aligned}
I_1(\theta) &= -E\left[\frac{\partial^2 \log f_\theta(X_1)}{\partial \theta^2}\right] \\
&= \frac{1}{\theta^2} + \frac{2\theta}{\pi} \int_{-\infty}^{\infty} \frac{x^2(1-\theta^2x^2)}{(1+\theta^2x^2)^3} dx \\
&= \frac{1}{\theta^2} + \frac{2}{\theta^2\pi} \int_{-\infty}^{\infty} \frac{x^2(1-x^2)}{(1+x^2)^3} dx \\
&= \frac{1}{\theta^2} + \frac{2}{\theta^2\pi} \int_{-\infty}^{\infty} \frac{3(1+x^2)-2-(1+x^2)^2}{(1+x^2)^3} dx \\
&= \frac{1}{\theta^2} + \frac{2}{\theta^2\pi} \left[3\int_{-\infty}^{\infty} \frac{1}{(1+x^2)^2} dx \right. \\
&\qquad \left. -\ 2\int_{-\infty}^{\infty} \frac{1}{(1+x^2)^3} dx - \int_{-\infty}^{\infty} \frac{1}{1+x^2} dx \right] \\
&= \frac{1}{\theta^2} + \frac{2}{\theta^2}\left(\frac{3}{2} - \frac{3}{4} - 1\right) \\
&= \frac{1}{2\theta^2},
\end{aligned}$$

where the identity

$$\int_{-\infty}^{\infty} \frac{1}{(1+x^2)^k} dx = \frac{\sqrt{\pi}\,\Gamma(k-\frac{1}{2})}{\Gamma(k)}$$

is used. ∎

Exercise 66 (#4.138). Let $(X_1, ..., X_n)$ be a random sample from the logistic distribution on $\mathcal{R}$ with Lebesgue density

$$f_\theta(x) = \sigma^{-1} e^{-(x-\mu)/\sigma} / [1 + e^{-(x-\mu)/\sigma}]^2,$$

where $\mu \in \mathcal{R}$ and $\sigma > 0$ are unknown. Obtain a nondegenerated asymptotic distribution of the RLE $\hat{\theta}$ of $\theta = (\mu, \sigma)$.

Solution. Using the same argument as that in the solution for the previous exercise, we can show that the conditions in Theorem 4.17 of Shao (2003) are satisfied. Then

$$\sqrt{n}(\hat{\theta} - \theta) \to_d N(0, [I_1(\theta)]^{-1}).$$

To obtain $I_1(\theta)$, we calculate

$$\frac{\partial \log f_\theta(x)}{\partial \mu} = -\frac{1}{\sigma} + \frac{2}{\sigma[1 + e^{-(x-\mu)/\sigma}]},$$

$$\frac{\partial^2 \log f_\theta(x)}{\partial \mu^2} = -\frac{2e^{-(x-\mu)/\sigma}}{\sigma^2[1 + e^{-(x-\mu)/\sigma}]^2},$$

and

$$\begin{aligned}
-E\left[\frac{\partial^2 \log f_\theta(X_1)}{\partial \mu^2}\right] &= \frac{2}{\sigma^2}\int_{-\infty}^{\infty} \frac{e^{2(x-\mu)/\sigma}}{\sigma[1 + e^{(x-\mu)/\sigma}]^4} dx \\
&= \frac{2}{\sigma^2}\int_{-\infty}^{\infty} \frac{e^{2y}}{(1 + e^y)^4} dy \\
&= \frac{2}{\sigma^2}\int_0^1 \frac{(1-t)^2}{t^2} t^4 \left(\frac{1}{1-t} + \frac{1}{t}\right) dt \\
&= \frac{2}{\sigma^2}\int_0^1 (1-t) t dt \\
&= \frac{1}{3\sigma^2},
\end{aligned}$$

where the second equality follows from the transformation $y = -(x-\mu)/\sigma$ and the third equality follows from the transformation $t = (1 + e^y)^{-1}$ $(y = \log(1-t) - \log t)$. Note that

$$\frac{\partial^2 \log f_\theta(x)}{\partial \mu \partial \sigma} = \frac{e^{-(x-\mu)/\sigma} - 1}{\sigma^2[1 + e^{-(x-\mu)/\sigma}]} - \frac{2(x-\mu)e^{-(x-\mu)/\sigma}}{\sigma^2[1 + e^{-(x-\mu)/\sigma}]^2}$$

is odd symmetric about μ and $f_\theta(x)$ is symmetric about μ. Hence,

$$E\left[\frac{\partial^2 \log f_\theta(X_1)}{\partial \mu \partial \sigma}\right] = 0.$$

Differentiating $\log f_\theta(x)$ with respect to σ gives

$$\frac{\partial \log f_\theta(x)}{\partial \sigma} = -\frac{1}{\sigma} - \frac{x-\mu}{\sigma} + \frac{2(x-\mu)}{\sigma^2[1+e^{-(x-\mu)/\sigma}]}$$

and

$$\frac{\partial^2 \log f_\theta(x)}{\partial \sigma^2} = \frac{1}{\sigma^2} + \frac{2(x-\mu)}{\sigma^3} - \frac{4(x-\mu)}{\sigma^3[1+e^{-(x-\mu)/\sigma}]} - \frac{2(x-\mu)^2 e^{-(x-\mu)/\sigma}}{\sigma^4[1+e^{-(x-\mu)/\sigma}]^2}.$$

From $E(X_1-\mu)=0$ and the transformation $y=-(x-\mu)/\sigma$, we obtain that

$$-E\left[\frac{\partial^2 \log f_\theta(x)}{\partial \sigma^2}\right] = -\frac{1}{\sigma^2} + \frac{4}{\sigma^2}\int_{-\infty}^{\infty} \frac{ye^{2y}}{(1+e^y)^3}dy + \frac{2}{\sigma^2}\int_{-\infty}^{\infty}\frac{y^2e^{2y}}{(1+e^y)^4}dy.$$

From integration by parts,

$$\begin{aligned}\int_{-\infty}^{\infty} \frac{ye^{2y}}{(1+e^y)^3}dy &= \frac{ye^y}{2(1+e^y)^2}\bigg|_{\infty}^{-\infty} + \frac{1}{2}\int_{-\infty}^{\infty}\frac{e^y+ye^y}{(1+e^y)^2}dy \\ &= \frac{1}{2}\int_{-\infty}^{\infty}\frac{e^y}{(1+e^y)^2}dy + \frac{1}{2}\int_{-\infty}^{\infty}\frac{ye^y}{(1+e^y)^2}dy \\ &= \frac{1}{2}\end{aligned}$$

and

$$\begin{aligned}\int_{-\infty}^{\infty}\frac{y^2e^{2y}}{(1+e^y)^4}dy &= 2\int_0^{\infty}\frac{y^2e^{2y}}{(1+e^y)^4}dy \\ &= \frac{2y^2e^y}{3(1+e^y)^3}\bigg|_{\infty}^{0} + \frac{2}{3}\int_0^{\infty}\frac{(2y+y^2)e^y}{(1+e^y)^3}dy \\ &= \frac{(2y+y^2)}{3(1+e^y)^2}\bigg|_{\infty}^{0} + \frac{2}{3}\int_0^{\infty}\frac{1+y}{(1+e^y)^2}dy \\ &= \frac{2(1+y)e^{-y}}{3(1+e^y)}\bigg|_{\infty}^{0} - \frac{2}{3}\int_0^{\infty}\frac{ye^{-y}}{1+e^y}dy \\ &= \frac{1}{3} - \frac{2}{3}\int_0^{\infty}\frac{ye^{-2y}}{1+e^{-y}}dy.\end{aligned}$$

Using the series $(1+e^{-y})^{-1} = \sum_{j=0}^{\infty}(-e^{-y})^j$, we obtain that

$$\begin{aligned}\int_0^{\infty}\frac{ye^{-2y}}{1+e^{-y}}dy &= \int_0^{\infty} ye^{-2y}\sum_{j=0}^{\infty}(-1)^j e^{-jy}dy \\ &= \sum_{j=0}^{\infty}(-1)^j\int_0^{\infty} ye^{-(j+2)y}dy \\ &= \sum_{j=0}^{\infty}(-1)^j\frac{1}{(j+2)^2}.\end{aligned}$$

Noting that $(-1)^j = 1$ when j is even and $(-1)^j = -1$ when j is odd, we obtain that

$$\begin{aligned}
\sum_{j=0}^{\infty}(-1)^j \frac{1}{(j+2)^2} &= \sum_{k=1}^{\infty} \frac{1}{(2k)^2} - \sum_{k=1}^{\infty} \frac{1}{(2k+1)^2} \\
&= \sum_{k=1}^{\infty} \frac{1}{(2k)^2} - \sum_{k=1}^{\infty} \left[\frac{1}{(2k+1)^2} + \frac{1}{(2k)^2} - \frac{1}{(2k)^2}\right] \\
&= 2\sum_{k=1}^{\infty} \frac{1}{(2k)^2} - \sum_{j=2}^{\infty} \frac{1}{j^2} \\
&= \frac{1}{2}\sum_{k=1}^{\infty} \frac{1}{k^2} - \sum_{j=1}^{\infty} \frac{1}{j^2} + 1 \\
&= 1 - \frac{1}{2}\sum_{k=1}^{\infty} \frac{1}{k^2} \\
&= 1 - \frac{\pi^2}{12}.
\end{aligned}$$

Combining these results, we obtain that

$$-E\left[\frac{\partial^2 \log f_\theta(x)}{\partial \sigma^2}\right] = -\frac{1}{\sigma^2} + \frac{2}{\sigma^2} + \frac{2}{\sigma^2}\left[\frac{1}{3} - \frac{2}{3}\left(1 - \frac{\pi^2}{12}\right)\right] = \frac{1}{3\sigma^2} + \frac{\pi^2}{9\sigma^2}.$$

Therefore,

$$I_1(\theta) = \frac{1}{\sigma^2}\begin{pmatrix} \frac{1}{3} & 0 \\ 0 & \frac{1}{3} + \frac{\pi^2}{9} \end{pmatrix}. \blacksquare$$

Exercise 67 (#4.140). Let $(X_1, ..., X_n)$ be a random sample of binary random variables with $P(X_1 = 1) = p$, where $p \in (0,1)$ is unknown. Let $\hat{\theta}$ be the MLE of $\theta = p(1-p)$.
(i) Show that $\hat{\theta}$ is asymptotically normal when $p \neq \frac{1}{2}$.
(ii) When $p = \frac{1}{2}$, derive a nondegenerated asymptotic distribution of $\hat{\theta}$ with an appropriate normalization.
Solution. (i) Since the sample mean $\bar{X}$ is the MLE of p, the MLE of $\theta = p(1-p)$ is $\bar{X}(1-\bar{X})$. From the central limit theorem, $\sqrt{n}(\bar{X} - p) \to_d N(0, \theta)$. Using the δ-method with $g(x) = x(1-x)$ and $g'(x) = 1 - 2x$, we obtain that $\sqrt{n}(\hat{\theta} - \theta) \to_d N(0, (1-2p)^2\theta)$. Note that this asymptotic distribution is degenerate when $p = \frac{1}{2}$.
(ii) When $p = \frac{1}{2}$, $\sqrt{n}(\bar{X} - \frac{1}{2}) \to_d N(0, \frac{1}{4})$. Hence,

$$4n\left(\frac{1}{4} - \hat{\theta}\right) = 4n\left(\bar{X} - \frac{1}{2}\right)^2 \to_d \chi_1^2. \blacksquare$$

Exercise 68 (#4.141). Let $(X_1, Y_1), ..., (X_n, Y_n)$ be independent and identically distributed random 2-vectors satisfying $0 \le X_1 \le 1$, $0 \le Y_1 \le 1$, and

$$P(X_1 > x, Y_1 > y) = (1-x)(1-y)(1-\max\{x,y\})^\theta$$

for $0 \le x \le 1$, $0 \le y \le 1$, where $\theta \ge 0$ is unknown.
(i) Obtain the likelihood function and the likelihood equation.
(ii) Obtain the asymptotic distribution of the MLE of θ.
Solution. (i) Let $X = X_1$ and $Y = Y_1$. Note that $F(x,y)$ is differentiable when $x \ne y$ but not differentiable when $x = y$. Hence, when $x \ne y$, (X,Y) has Lebesgue density

$$f_\theta(x,y) = \begin{cases} (\theta+1)(1-x)^\theta & x > y \\ (\theta+1)(1-y)^\theta & x < y \end{cases}$$

and

$$\begin{aligned} P(X > t, Y > t, X \ne Y) &= 2P(X > t, Y > t, X > Y) \\ &= 2(\theta+1)\int_t^1 \int_t^x (1-x)^\theta dy dx \\ &= 2(\theta+1)\int_t^1 (x-t)(1-x)^\theta dx \\ &= \frac{2(1-t)^{\theta+2}}{\theta+2}. \end{aligned}$$

Also, $P(X > t, Y > t) = (1-t)^{\theta+2}$. Hence,

$$\begin{aligned} P(X > t, X = Y) &= P(X > t, Y > t, X = Y) \\ &= P(X > t, Y > t) - P(X > t, Y > t, X \ne Y) \\ &= \frac{\theta(1-t)^{\theta+2}}{\theta+2}. \end{aligned}$$

This means that on the line $x = y$, (X,Y) has Lebesgue density $\theta(1-t)^{\theta+1}$. Let ν be the sum of the Lebesgue measure on $\mathcal{R}^2$ and the Lebesgue measure on $\{(x,y) \in \mathcal{R}^2 : x = y\}$. Then the probability density of (X,Y) with respect to ν is

$$f_\theta(x,y) = \begin{cases} (\theta+1)(1-x)^\theta & x > y \\ (\theta+1)(1-y)^\theta & x < y \\ \theta(1-x)^{\theta+1} & x = y. \end{cases}$$

Let T be the number of (X_i, Y_i)'s with $X_i = Y_i$ and $Z_i = \max\{X_i, Y_i\}$. Then the likelihood function is

$$\ell(\theta) = (\theta+1)^{n-T}\theta^T \prod_{i=1}^n (1-Z_i)^\theta \prod_{i:X_i=Y_i} (1-Z_i)$$

and the likelihood equation is

$$\frac{\partial \log \ell(\theta)}{\partial \theta} = \frac{n-T}{\theta+1} + \frac{T}{\theta} + \sum_{i=1}^{n} \log(1-Z_i) = 0,$$

which has a unique solution (in the parameter space)

$$\hat{\theta} = \frac{\sqrt{(n-W)^2 + 4WT} - (n-W)}{2W},$$

where $W = -\sum_{i=1}^{n} \log(1-Z_i)$.
(ii) Since

$$\frac{\partial^2 \log \ell(\theta)}{\partial \theta^2} = -\frac{n-T}{(\theta+1)^2} - \frac{T}{\theta^2} < 0,$$

$\hat{\theta}$ is the MLE of θ. Since $E(T) = n\theta/(\theta+2)$, we obtain that

$$I_1(\theta) = -\frac{1}{n} E\left[\frac{\partial^2 \log \ell(\theta)}{\partial \theta^2}\right] = \frac{\theta^2 + 4\theta + 1}{\theta(\theta+2)(\theta+1)^2}.$$

Hence,

$$\sqrt{n}(\hat{\theta} - \theta) \to_d N\left(0, \frac{\theta(\theta+2)(\theta+1)^2}{\theta^2 + 4\theta + 1}\right). \blacksquare$$

Exercise 69. Consider the one-way random effects model

$$X_{ij} = \mu + A_i + e_{ij}, \quad j = 1, ..., n, i = 1, ..., m,$$

where $\mu \in \mathcal{R}$, A_i's are independent and identically distributed as $N(0, \sigma_a^2)$, e_{ij}'s are independent and identically distributed as $N(0, \sigma^2)$, σ_a^2 and σ^2 are unknown, and A_i's and e_{ij}'s are independent. Obtain nondegenerate asymptotic distributions of the MLE's of μ, σ_a^2, and σ^2.
Solution. From Exercise 49(ii), the MLE of μ is $\bar{X}_{\cdot\cdot}$, which is always normally distributed with mean μ and variance $m^{-1}(\sigma_a^2 + n^{-1}\sigma^2)$.

From Exercise 49(ii), the MLE of σ^2 is $\hat{\sigma}^2 = S_E/[m(n-1)]$ and the MLE of σ_a^2 is $\hat{\sigma}_a^2 = S_A/[n(m-1)] - S_E/[nm(n-1)]$, provided that $\hat{\sigma}_a^2 > 0$. We now show that as long as $nm \to \infty$, $P(\hat{\sigma}_a^2 \leq 0) \to 0$, which implies that, for the asymptotic distributions of the MLE's, we may assume that $\hat{\sigma}_a^2 > 0$. Since S_E/σ^2 has the chi-square distribution $\chi^2_{m(n-1)}$, $S_E/[m(n-1)] \to_p \sigma^2$ as $nm \to \infty$ (either $n \to \infty$ or $m \to \infty$). Since $S_A/(\sigma^2 + n\sigma_a^2)$ has the chi-square distribution χ^2_{m-1}, the distribution of $S_A/[n(m-1)]$ is the same as that of $(\sigma_a^2 + n^{-1}\sigma^2)W_{m-1}/(m-1)$, where W_{m-1} is a random variable having the chi-square distribution χ^2_{m-1}. We need to consider three different cases.
Case 1: $m \to \infty$ and $n \to \infty$. In this case, $S_E/[nm(n-1)] \to_p 0$ and

$(\sigma_a^2 + n^{-1}\sigma^2)W_{m-1}/(m-1) \to_p \sigma_a^2 > 0$. Hence, $\hat{\sigma}_a^2 \to_p \sigma_a^2 > 0$, which implies $P(\hat{\sigma}_a^2 \leq 0) \to 0$.
Case 2: $m \to \infty$ but n is fixed. In this case, $S_E/[nm(n-1)] \to_p n^{-1}\sigma^2$ and $(\sigma_a^2 + n^{-1}\sigma^2)W_{m-1}/(m-1) \to_p (\sigma_a^2 + n^{-1}\sigma^2)$. We still have $\hat{\sigma}_a^2 \to_p \sigma_a^2 > 0$.
Case 3: $n \to \infty$ but m is fixed. In this case, $S_E/[nm(n-1)] \to_p 0$ and $(\sigma_a^2 + n^{-1}\sigma^2)W_{m-1}/(m-1) \to_d \sigma_a^2 W_{m-1}/(m-1)$. Hence, by Slutsky's theorem, $\hat{\sigma}_a^2 \to_d \sigma_a^2 W_{m-1}/(m-1)$, which is a nonnegative random variable. Hence, $P(\hat{\sigma}_a^2 \leq 0) \to 0$.

Therefore, the asymptotic distributions of MLE's are the same as those of $\hat{\sigma}_a^2$ and $\hat{\sigma}^2$. Since S_E/σ^2 has the chi-square distribution $\chi^2_{m(n-1)}$,

$$\sqrt{nm}\left(\hat{\sigma}^2 - \sigma^2\right) \to_d N(0, 2\sigma^4)$$

as $nm \to \infty$ (either $n \to \infty$ or $m \to \infty$). For $\hat{\sigma}_a^2$, we need to consider the three cases previously discussed.
Case 1: $m \to \infty$ and $n \to \infty$. In this case,

$$\sqrt{m}\left[\frac{S_E}{nm(n-1)} - \frac{\sigma^2}{n}\right] \to_p 0$$

and

$$\sqrt{m}\left(\frac{W_{m-1}}{m-1} - 1\right) \to_d N(0, 2).$$

Since $S_A/[n(m-1)]$ and $(\sigma_a^2 + n^{-1}\sigma^2)W_{m-1}/(m-1)$ have the same distribution,

$$\sqrt{m}(\hat{\sigma}_a^2 - \sigma_a^2) = \sqrt{m}\left[\frac{S_A}{n(m-1)} - \left(\sigma_a^2 + \frac{\sigma^2}{n}\right) + \frac{\sigma^2}{n} - \frac{S_E}{nm(n-1)}\right]$$

has the same asymptotic distribution as that of

$$\sqrt{m}\left(\sigma_a^2 + \frac{\sigma^2}{n}\right)\left(\frac{W_{m-1}}{m-1} - 1\right).$$

Thus,

$$\sqrt{m}(\hat{\sigma}_a^2 - \sigma_a^2) \to_d N(0, 2\sigma_a^4).$$

Case 2: $m \to \infty$ but n is fixed. In this case,

$$\sqrt{m}\left[\frac{S_E}{nm(n-1)} - \frac{\sigma^2}{n}\right] \to_d N(0, 2\sigma^4 n^{-3}).$$

From the argument in the previous case and the fact that S_A and S_E are independent, we obtain that

$$\sqrt{m}(\hat{\sigma}_a^2 - \sigma_a^2) \to_d N\left(0, 2(\sigma_a^2 + n^{-1}\sigma^2)^2 + 2\sigma^4 n^{-3}\right).$$

Case 3: $n \to \infty$ but m is fixed. In this case, $S_E/[nm(n-1)] - \sigma^2/n \to_p 0$ and

$$\left(\sigma_a^2 + \frac{\sigma^2}{n}\right)\left(\frac{W_{m-1}}{m-1} - 1\right) \to_d \sigma_a^2 \left(\frac{W_{m-1}}{m-1} - 1\right).$$

Therefore,

$$\hat{\sigma}_a^2 - \sigma_a^2 \to_d \sigma_a^2 \left(\frac{W_{m-1}}{m-1} - 1\right). \ \blacksquare$$

Exercise 70 (#4.151). Let $(X_1, ..., X_n)$ be a random sample from the logistic distribution on $\mathcal{R}$ with Lebesgue density

$$f_\theta(x) = \sigma^{-1} e^{-(x-\mu)/\sigma}/[1 + e^{-(x-\mu)/\sigma}]^2,$$

where $\mu \in \mathcal{R}$ and $\sigma > 0$ are unknown. Using Newton-Raphson and Fisher-scoring methods, find
(i) one-step MLE's of μ when σ is known;
(ii) one-step MLE's of σ when μ is known;
(iii) one-step MLE's of (μ, σ);
(iv) $\sqrt{n}$-consistent initial estimators in (i)-(iii).
Solution. (i) Let $\ell(\mu)$ be the likelihood function when σ is known. From Exercise 66,

$$s_\sigma(\mu) = \frac{\partial \log \ell(\mu)}{\partial \mu} = \frac{n}{\sigma} - \frac{2}{\sigma}\sum_{i=1}^n \frac{e^{-(X_i-\mu)/\sigma}}{1 + e^{-(X_i-\mu)/\sigma}},$$

$$s'_\sigma(\mu) = \frac{\partial^2 \log \ell(\mu)}{\partial \mu^2} = -\frac{2}{\sigma^2}\sum_{i=1}^n \frac{e^{-(X_i-\mu)/\sigma}}{[1 + e^{-(X_i-\mu)/\sigma}]^2},$$

and

$$-E\left[\frac{\partial^2 \log \ell(\mu)}{\partial \mu^2}\right] = \frac{n}{3\sigma^2}.$$

Hence, the one-step MLE of μ is

$$\hat{\mu}^{(1)} = \hat{\mu}^{(0)} - [s'_\sigma(\hat{\mu}^{(0)})]^{-1} s_\sigma(\hat{\mu}^{(0)})$$

by the Newton-Raphson method, where $\hat{\mu}^{(0)}$ is an initial estimator of μ, and is

$$\hat{\mu}^{(1)} = \hat{\mu}^{(0)} + 3\sigma^2 n^{-1} s_\sigma(\hat{\mu}^{(0)})$$

by the Fisher-scoring method.
(ii) Let $\ell(\sigma)$ be the likelihood function when μ is known. From Exercise 66,

$$s_\mu(\sigma) = \frac{\partial \log \ell(\sigma)}{\partial \sigma} = -\frac{n}{\sigma} - \sum_{i=1}^n \frac{X_i - \mu}{\sigma} + \sum_{i=1}^n \frac{2(X_i - \mu)}{\sigma^2[1 + e^{-(X_i-\mu)/\sigma}]},$$

$$
\begin{aligned}
s'_\mu(\sigma) &= \frac{\partial^2 \log \ell(\sigma)}{\partial \sigma^2} \\
&= \frac{n}{\sigma^2} + \sum_{i=1}^n \frac{2(X_i-\mu)}{\sigma^3} - \sum_{i=1}^n \frac{4(X_i-\mu)}{\sigma^3[1+e^{-(X_i-\mu)/\sigma}]} \\
&\quad - \sum_{i=1}^n \frac{2(X_i-\mu)^2 e^{-(X_i-\mu)/\sigma}}{\sigma^4[1+e^{-(X_i-\mu)/\sigma}]^2},
\end{aligned}
$$

and

$$
-E\left[\frac{\partial^2 \log \ell(\sigma)}{\partial \sigma^2}\right] = \frac{1}{\sigma^2}\left(\frac{1}{3}+\frac{\pi^2}{9}\right).
$$

Hence, the one-step MLE of σ is

$$
\hat{\sigma}^{(1)} = \hat{\sigma}^{(0)} - [s'_\mu(\hat{\sigma}^{(0)})]^{-1} s_\mu(\hat{\sigma}^{(0)})
$$

by the Newton-Raphson method, where $\hat{\sigma}^{(0)}$ is an initial estimator of σ, and is

$$
\hat{\sigma}^{(1)} = \hat{\sigma}^{(0)} + \frac{(\hat{\sigma}^{(0)})^2}{\pi^2/9+1/3} s_\mu(\hat{\sigma}^{(0)})
$$

by the Fisher-scoring method.
(iii) Let $\ell(\mu,\sigma)$ be the likelihood function when both μ and σ are unknown. From parts (i)-(ii) of the solution,

$$
s(\mu,\sigma) = \frac{\partial \log \ell(\mu,\sigma)}{\partial(\mu,\sigma)} = \begin{pmatrix} s_\sigma(\mu) \\ s_\mu(\sigma) \end{pmatrix}
$$

and

$$
s'(\mu,\sigma) = \frac{\partial^2 \log \ell(\theta)}{\partial(\mu,\sigma)\partial(\mu,\sigma)^\tau} = \begin{pmatrix} s'_\sigma(\mu) & s'_{\mu,\sigma} \\ s'_{\mu,\sigma} & s'_\mu(\sigma) \end{pmatrix},
$$

where

$$
s'_{\mu,\sigma} = \frac{\partial^2 \log \ell(\mu,\sigma)}{\partial\mu\partial\sigma} = \sum_{i=1}^n \frac{e^{-(X_i-\mu)/\sigma}-1}{\sigma^2[1+e^{-(X_i-\mu)/\sigma}]} - \sum_{i=1}^n \frac{2(X_i-\mu)e^{-(X_i-\mu)/\sigma}}{\sigma^2[1+e^{-(X_i-\mu)/\sigma}]^2}.
$$

Hence, by the Newton-Raphson method, the one-step MLE of (μ,σ) is

$$
\begin{pmatrix} \hat{\mu}^{(1)} \\ \hat{\sigma}^{(1)} \end{pmatrix} = \begin{pmatrix} \hat{\mu}^{(0)} \\ \hat{\sigma}^{(0)} \end{pmatrix} - \left[s'(\hat{\mu}^{(0)},\hat{\sigma}^{(0)})\right]^{-1} s(\hat{\mu}^{(0)},\hat{\sigma}^{(0)}).
$$

From Exercise 66,

$$
-E[s'(\mu,\sigma)] = \frac{n}{\sigma^2}\begin{pmatrix} \frac{1}{3} & 0 \\ 0 & \frac{1}{3}+\frac{\pi^2}{9} \end{pmatrix}.
$$

Hence, by the Fisher-scoring method, the one-step MLE of (μ, σ) is

$$\begin{pmatrix} \hat{\mu}^{(1)} \\ \hat{\sigma}^{(1)} \end{pmatrix} = \begin{pmatrix} \hat{\mu}^{(0)} \\ \hat{\sigma}^{(0)} \end{pmatrix} + \frac{[\hat{\sigma}^{(0)}]^2}{n} \begin{pmatrix} 3s_{\hat{\sigma}^{(0)}}(\hat{\mu}^{(0)}) \\ s_{\hat{\mu}^{(0)}}(\hat{\sigma}^{(0)}) / \left(\frac{1}{3} + \frac{\pi^2}{9}\right) \end{pmatrix}.$$

(iv) Note that logistic distribution has mean μ and variance $\sigma^2\pi^2/3$. Thus, in (i)-(iii), we may take $\hat{\mu}^{(0)} = \bar{X}$ (the sample mean) and

$$\hat{\sigma}^{(0)} = \frac{\sqrt{3}}{\pi} \sqrt{\frac{1}{n} \sum_{i=1}^{n} (X_i - \bar{X})^2},$$

which are $\sqrt{n}$-consistent. ∎

Chapter 5

Estimation in Nonparametric Models

Exercise 1 (#5.3). Let $p \geq 1$ and $\mathcal{F}_p$ be the set of cumulative distribution functions on $\mathcal{R}$ having finite pth moment. Mallows' distance between F and G in $\mathcal{F}_p$ is defined to be

$$\varrho_{M_p}(F, G) = \inf(E\|X - Y\|^p)^{1/p},$$

where the infimum is taken over all pairs of random variables X and Y having marginal distributions F and G, respectively. Show that ϱ_{M_p} is a distance on $\mathcal{F}_p$.

Solution. Let U be a random variable having the uniform distribution on the interval $(0,1)$ and $F^{-1}(t) = \inf\{x : F(x) \geq t\}$ for any cumulative distribution function F. We first show that

$$\varrho_{M_p}(F, G) = [E|F^{-1}(U) - G^{-1}(U)|^p]^{1/p}.$$

Since $F^{-1}(U)$ is distributed as F and $G^{-1}(U)$ is distributed as G, we have

$$\varrho_{M_p}(F, G) \leq [E|F^{-1}(U) - G^{-1}(U)|^p]^{1/p}.$$

Let X and Y be any random variables whose marginal distributions are F and G, respectively. From Jensen's inequality for conditional expectations,

$$E|X - Y|^p = E[E(|X - Y|^p|X)] \geq E|X - E(Y|X)|^p.$$

Since X and $F^{-1}(U)$ have the same distribution, we conclude that

$$\varrho_{M_p}(F, G) \geq [E|F^{-1}(U) - E(Y|F^{-1}(U))|^p]^{1/p}.$$

Then the result follows if we can show that $E(Y|F^{-1}(U)) = G^{-1}(U)$. Clearly, $G^{-1}(U)$ is a Borel function of U. Since Y and $G^{-1}(U)$ have the same distribution, $\int_B Y dP = \int_B G^{-1}(U)dP$ for any $B \in \sigma(F^{-1}(U))$. Hence, $E(Y|F^{-1}(U)) = G^{-1}(U)$ a.s.

It is clear that $\varrho_{M_p}(F,G) \geq 0$ and $\varrho_{M_p}(F,G) = \varrho_{M_p}(G,F)$. If $\varrho_{M_p}(F,G) = 0$, then, by the established result, $E|F^{-1}(U) - F^{-1}(U)|^p = 0$. Thus, $F^{-1}(t) = G^{-1}(t)$ a.e. with respect to Lebesgue measure. Hence, $F = G$. Finally, for F, G, and H in $\mathcal{F}_p$,

$$\begin{aligned}
\varrho_{M_p}(F,G) &= \left[E|F^{-1}(U) - G^{-1}(U)|^p\right]^{1/p} \\
&\leq \left[E|F^{-1}(U) - H^{-1}(U)|^p\right]^{1/p} \\
&\quad + \left[E|H^{-1}(U) - G^{-1}(U)|^p\right]^{1/p} \\
&= \varrho_{M_p}(F,H) + \varrho_{M_p}(H,G),
\end{aligned}$$

where the inequality follows from Minkowski's inequality. This proves that ϱ_{M_p} is a distance. ∎

Exercise 2 (#5.5). Let $\mathcal{F}_1$ be the collection of cumulative distribution functions on $\mathcal{R}$ with finite means and ϱ_{M_1} be as defined in Exercise 1. Show that
(i) $\varrho_{M_1}(F,G) = \int_0^1 |F^{-1}(t) - G^{-1}(t)|dt$;
(ii) $\varrho_{M_1}(F,G) = \int_{-\infty}^{\infty} |F(x) - G(x)|dx$.
Solution. (i) Let U be a random variable having the uniform distribution on the interval $(0,1)$. From the solution of the previous exercise, $\varrho_{M_p}(F,G) = E|F^{-1}(U) - G^{-1}(U)|$. The result follows from

$$E|F^{-1}(U) - G^{-1}(U)| = \int_0^1 |F^{-1}(t) - G^{-1}(t)|dt.$$

(ii) From part (i), it suffices to show that

$$\int_{-\infty}^{\infty} |F(x) - G(x)|dx = \int_0^1 |F^{-1}(t) - G^{-1}(t)|dt.$$

Note that $\int_{-\infty}^{\infty} |F(x) - G(x)|dx$ is equal to the area on $\mathcal{R}^2$ bounded by two curves $F(x)$ and $G(x)$ and $\int_0^1 |F^{-1}(t) - G^{-1}(t)|dt$ is equal to the area on $\mathcal{R}^2$ bounded by two curves $F^{-1}(t)$ and $G^{-1}(t)$. Hence, they are the same and the result follows. ∎

Exercise 3. Let ϱ_{M_p} be the Mallows' distance defined in Exercise 1 and $\{G, G_1, G_2, ...\} \subset \mathcal{F}_p$. Show that $\lim_n \varrho_{M_p}(G_n, G) = 0$ if and only if $\lim_n \int |x|^p dG_n(x) = \int |x|^p dG(x)$ and $\lim_n G_n(x) = G(x)$ for any x at which G is continuous.

Solution. Let U be a random variable having the uniform distribution on $(0,1)$. By the solution of Exercise 1,

$$\varrho_{M_p}(G_n,G) = [E|G_n^{-1}(U) - G^{-1}(U)|^p]^{1/p}.$$

Assume that $\lim_n \varrho_{M_p}(G_n,G) = 0$. Then $\lim_n E|G_n^{-1}(U) - G^{-1}(U)|^p = 0$, which implies that $\lim_n E|G_n^{-1}(U)|^p = E|G^{-1}(U)|^p$ and $G_n^{-1}(U) \to_d G^{-1}(U)$. Since $G_n^{-1}(U)$ has distribution G_n and $G^{-1}(U)$ has distribution G, we conclude that $\lim_n \int |x|^p dG_n(x) = \int |x|^p dG(x)$ and $\lim_n G_n(x) = G(x)$ for any x at which G is continuous.

Assume now that $\lim_n \int |x|^p dG_n(x) = \int |x|^p dG(x)$ and $\lim_n G_n(x) = G(x)$ for any x at which G is continuous. Using the same argument in the solution of Exercise 54 in Chapter 1, we can show that $G_n^{-1}(U) \to_p G^{-1}(U)$. Since $\lim_n \int |x|^p dG_n(x) = \int |x|^p dG(x)$, by Theorem 1.8(viii) in Shao (2003), the sequence $\{|G_n^{-1}(U)|^p\}$ is uniformly integrable and, hence, $\lim_n E|G_n^{-1}(U) - G^{-1}(U)|^p = 0$, which means $\lim_n \varrho_{M_p}(G_n,G) = 0$. ∎

Exercise 4 (#5.6). Find an example of cumulative distribution functions G, G_1, G_2,... on $\mathcal{R}$ such that

(i) $\lim_n \varrho_\infty(G_n,G) = 0$ but $\varrho_{M_p}(G_n,G)$ does not converge to 0, where ϱ_{M_p} is the distance defined in Exercise 1 and ϱ_∞ is the sup-norm distance defined as $\varrho_\infty(F,G) = \sup_x |F(x) - G(x)|$ for any cumulative distribution functions F and G;

(ii) $\lim_n \varrho_{M_p}(G_n,G) = 0$ but $\varrho_\infty(G_n,G)$ does not converge to 0.

Solution. Let U be a random variable having the uniform distribution on the interval $(0,1)$.

(i) Let G be the cumulative distribution function of U and G_n be the cumulative distribution function of

$$U_n = \begin{cases} U & \text{if } U \geq n^{-1} \\ n^2 & \text{if } U < n^{-1}. \end{cases}$$

Then $\lim_n P(|U_n - U| > \epsilon) = \lim_n n^{-1} = 0$ for any $\epsilon > 0$ and, hence, $U_n \to_d U$. Since the distribution of U is continuous, by Pólya's theorem (e.g., Proposition 1.16 in Shao, 2003), $\lim_n \varrho_\infty(G_n,G) = 0$. But $E|U_n| \geq n^2 P(U < n^{-1}) = n$ and $E|U| = \frac{1}{2}$. Hence $\lim_n E|U_n| \neq E|U|$. By Exercise 3, $\varrho_{M_p}(G_n,G)$ does not converge to 0.

(ii) Let G_n be the cumulative distribution function of U/n and $G(x) = I_{[0,\infty)}(x)$ (the degenerate distribution at 0). Then $\lim_n E|U/n|^p = 0$ for any p and $U/n \to_d 0$. Thus, by Exercise 3, $\lim_n \varrho_{M_p}(G_n,G) = 0$. But $G_n(0) = P(U \leq 0) = 0$ for all n and $G(0) = 1$, i.e., $G_n(0)$ does not converge to $G(0)$. Hence $\varrho_\infty(G_n,G)$ does not converge to 0. ∎

Exercise 5 (#5.8). Let X be a random variable having cumulative dis-

tribution function F. Show that
(i) $E|X|^2 < \infty$ implies $\int \{F(t)[1 - F(t)]\}^{p/2} dt < \infty$ for $p > 1$;
(ii) $E|X|^{2+\delta} < \infty$ with some $\delta > 0$ implies $\int \{F(t)[1 - F(t)]\}^{1/2} dt < \infty$.
Solution. (i) If $E|X|^2 < \infty$, then, by Exercise 23 in Chapter 1,

$$E|X|^2 = \int_0^\infty P(|X|^2 > t)dt = 2\int_0^\infty sP(|X| > s)ds,$$

which implies that $\lim_{s\to\infty} s^2[1 - F(s)] = 0$ and $\lim_{s\to-\infty} s^2 F(s) = 0$. Then, $\lim_{s\to\infty} s^p[1 - F(s)]^{p/2} = 0$ and $\lim_{s\to-\infty} s^p[F(s)]^{p/2} = 0$. Since $p > 1$, we conclude that $\int_{-\infty}^0 [F(s)]^{p/2} ds < \infty$ and $\int_0^\infty [1 - F(s)]^{p/2} ds < \infty$, which implies that

$$\int_{-\infty}^\infty \{F(t)[1 - F(t)]\}^{p/2} dt \le \int_{-\infty}^0 [F(t)]^{p/2} dt + \int_0^\infty [1 - F(t)]^{p/2} dt < \infty.$$

(ii) Similarly, when $E|X|^{2+\delta} < \infty$,

$$E|X|^{2+\delta} = \int_0^\infty P(|X|^{2+\delta} > t)dt = (2 + \delta)\int_0^\infty P(|X| > s)s^{1+\delta} ds,$$

which implies $\lim_{s\to\infty} s^{2+\delta}[1-F(s)] = 0$ and $\lim_{s\to-\infty} s^{2+\delta}F(s) = 0$. Then, $\lim_{s\to\infty} s^{1+\delta/2}[1 - F(s)]^{1/2} = 0$ and $\lim_{s\to-\infty} s^{1+\delta/2}[F(s)]^{1/2} = 0$. Since $\delta > 0$, this implies that

$$\int_{-\infty}^\infty \{F(t)[1 - F(t)]\}^{1/2} dt \le \int_{-\infty}^0 [F(t)]^{1/2} dt + \int_0^\infty [1 - F(t)]^{1/2} dt < \infty. \blacksquare$$

Exercise 6 (#5.10). Show that $p_i = c/n$, $i = 1, ..., n$, $\lambda = -(c/n)^{n-1}$ is a maximum of the function

$$H(p_1, ..., p_n, \lambda) = \prod_{i=1}^n p_i + \lambda\left(\sum_{i=1}^n p_i - c\right)$$

over $p_i > 0$, $i = 1, ..., n$, $\sum_{i=1}^n p_i = c$.
Note. This exercise shows that the empirical distribution function (which puts mass n^{-1} to each of n observations) is a maximum likelihood estimator.
Solution. It suffices to show that

$$\prod_{i=1}^n p_i \le \left(\frac{c}{n}\right)^n$$

for any $p_i > 0$, $i = 1, ..., n$, $\sum_{i=1}^n p_i = c$. Let X be a random variable taking value p_i with probability n^{-1}, $i = 1, ..., n$. From Jensen's inequality,

$$\frac{1}{n}\sum_{i=1}^n \log p_i = E(\log X) \le \log E(X) = \log\left(\frac{1}{n}\sum_{i=1}^n p_i\right) = \log\left(\frac{c}{n}\right),$$

which establishes the result. ∎

Exercise 7 (#5.11). Consider the problem of maximizing $\prod_{i=1}^n p_i$ over

$$p_i > 0, \quad i = 1, ..., n, \quad \sum_{i=1}^n p_i = 1, \quad \text{and} \quad \sum_{i=1}^n p_i u_i = 0,$$

where u_i's are s-vectors. Show that the solution is

$$\hat{p}_i = \frac{1}{n(1 + \lambda^\tau u_i)}, \qquad i = 1, ..., n,$$

and $\lambda \in \mathcal{R}^s$ satisfying

$$\sum_{i=1}^n \frac{u_i}{1 + \lambda^\tau u_i} = 0.$$

Solution. Consider the Lagrange multiplier method with

$$H(p_1, ..., p_n, \tau, \lambda) = \sum_{i=1}^n \log p_i + \tau \left(\sum_{i=1}^n p_i - 1 \right) - n\lambda^\tau \sum_{i=1}^n p_i u_i.$$

Taking the derivatives of H and setting them to 0, we obtain that

$$\frac{1}{p_i} + \tau - n\lambda^\tau u_i = 0, \quad i = 1, ..., n, \quad \sum_{i=1}^n p_i = 1, \quad \text{and} \quad \sum_{i=1}^n p_i u_i = 0.$$

The solution to these equations is $\tau = -n$ and

$$\hat{p}_i = \frac{1}{n(1 + \lambda^\tau u_i)}, \quad i = 1, ..., n.$$

Substituting $\hat{p}_i$ into $\sum_{i=1}^n p_i u_i = 0$, we conclude that λ is the solution of

$$\sum_{i=1}^n \frac{u_i}{1 + \lambda^\tau u_i} = 0. \ ∎$$

Exercise 8 (#5.15). Let $\delta_1, ..., \delta_n$ be n observations from a binary random variable and

$$\ell(p_1, ..., p_{n+1}) = \prod_{i=1}^n p_i^{\delta_i} \left(\sum_{j=i+1}^{n+1} p_j \right)^{1-\delta_i}.$$

(i) Show that maximizing $\ell(p_1, ..., p_{n+1})$ subject to $p_i \geq 0$, $i = 1, ..., n+1$, $\sum_{i=1}^{n+1} p_i = 1$ is equivalent to maximizing

$$\prod_{i=1}^n q_i^{\delta_i} (1 - q_i)^{n-i+1-\delta_i},$$

where $q_i = p_i / \sum_{j=i}^{n+1} p_j$, $i = 1, ..., n$.
(ii) Show that

$$\hat{p}_i = \frac{\delta_i}{n-i+1} \prod_{j=1}^{i-1} \left(1 - \frac{\delta_j}{n-j+1}\right), \quad i = 1, ..., n, \quad \hat{p}_{n+1} = 1 - \sum_{i=1}^{n} \hat{p}_i$$

maximizes $\ell(p_1, ..., p_{n+1})$ subject to $p_i \geq 0$, $i = 1, ..., n+1$, $\sum_{i=1}^{n+1} p_i = 1$.
(iii) For any $x_1 \leq x_2 \leq \cdots \leq x_n$, show that

$$\hat{F}(t) = \sum_{i=1}^{n+1} \hat{p}_i I_{(0,t]}(x_i) = 1 - \prod_{x_i \leq t} \left(1 - \frac{\delta_i}{n-i+1}\right).$$

(iv) When $\delta_i = 1$ for all i, show that $\hat{p}_i = n^{-1}$, $i = 1, ..., n$, and $\hat{p}_{n+1} = 0$.
Note. This exercise shows that the well-known Kaplan-Meier product-limit estimator is a maximum likelihood estimator.
Solution. (i) Since

$$1 - q_i = 1 - \frac{p_i}{\sum_{j=i}^{n+1} p_j} = \frac{\sum_{j=i+1}^{n+1} p_j}{\sum_{j=i}^{n+1} p_j},$$

$$\prod_{i=1}^{n} q_i^{\delta_i} (1-q_i)^{1-\delta_i} = \prod_{i=1}^{n} p_i^{\delta_i} \left(\sum_{j=i+1}^{n+1} p_j\right)^{1-\delta_i} \left(\sum_{j=i}^{n+1} p_j\right)^{-1}.$$

From

$$\begin{aligned}\prod_{i=1}^{n}(1-q_i)^{n-i} &= \left(\frac{\sum_{j=2}^{n+1} p_j}{\sum_{j=1}^{n+1} p_j}\right)^{n-1} \left(\frac{\sum_{j=3}^{n+1} p_j}{\sum_{j=2}^{n+1} p_j}\right)^{n-2} \cdots \frac{\sum_{j=n}^{n+1} p_j}{\sum_{j=n-1}^{n+1} p_j} \\ &= \sum_{j=2}^{n+1} p_j \sum_{j=3}^{n+1} p_j \cdots \sum_{j=n}^{n+1} p_j \\ &= \prod_{i=1}^{n} \sum_{j=i}^{n+1} p_j,\end{aligned}$$

we obtain that

$$\prod_{i=1}^{n} q_i^{\delta_i} (1-q_i)^{n-i+1-\delta_i} = \prod_{i=1}^{n} p_i^{\delta_i} \left(\sum_{j=i+1}^{n+1} p_j\right)^{1-\delta_i}.$$

The result follows since $q_1, ..., q_n$ are n free variables.
(ii) From part (i),

$$\log \ell(p_1, ..., p_{n+1}) = \sum_{i=1}^{n} [\delta_i \log q_i + (n-i+1-\delta_i) \log(1-q_i)].$$

Then
$$\frac{\partial \log \ell}{\partial q_i} = \frac{\delta_i}{q_i} - \frac{n-i+1-\delta_i}{1-q_i} = 0, \quad i = 1, ..., n,$$
have the solution
$$\hat{q}_i = \frac{\delta_i}{n-i+1}, \quad i = 1, ..., n,$$
which maximizes $\ell(p_1, ..., p_{n+1})$ since
$$\frac{\partial^2 \log \ell}{\partial q_i^2} = -\frac{\delta_i}{q_i^2} - \frac{n-i+1-\delta_i}{(1-q_i)^2} < 0 \quad \text{and} \quad \frac{\partial^2 \log \ell}{\partial q_i \partial q_k} = 0$$
for any i and $k \neq i$. Since
$$\prod_{j=1}^{i}(1-q_j) = \frac{\sum_{k=2}^{n+1} p_k}{\sum_{k=1}^{n+1} p_k} \frac{\sum_{k=3}^{n+1} p_k}{\sum_{k=2}^{n+1} p_k} \cdots \frac{\sum_{k=i}^{n+1} p_k}{\sum_{k=i-1}^{n+1} p_k} = \sum_{k=i}^{n+1} p_k,$$
we obtain that
$$q_i \prod_{j=1}^{i}(1-q_j) = p_i, \quad i = 1, ..., n.$$
Hence, by (i), $\ell(p_1, ..., p_{n+1})$ is maximized by
$$\hat{p}_i = \hat{q}_i \prod_{j=1}^{i}(1-\hat{q}_j) = \frac{\delta_i}{n-i+1} \prod_{j=1}^{i}\left(1 - \frac{\delta_j}{n-j+1}\right), \quad i = 1, ..., n,$$
and $\hat{p}_{n+1} = 1 - \sum_{i=1}^{n} \hat{p}_i$.
(iii) Define $x_0 = 0$ and $x_{n+1} = \infty$. Let $t \in (x_i, x_{i+1}]$, $i = 0, 1, ..., n$. Then
$$\prod_{x_i < t}\left(1 - \frac{\delta_i}{n-i+1}\right) = \prod_{j=1}^{i}(1-\hat{q}_j) = \frac{\hat{p}_i}{\hat{q}_i} = \sum_{j=i}^{n+1} \hat{p}_j = 1 - \sum_{j=1}^{i-1} \hat{p}_j.$$
Hence,
$$\sum_{i=1}^{n+1} \hat{p}_i I_{(0,t]}(x_i) = 1 - \prod_{x_i < t}\left(1 - \frac{\delta_i}{n-i+1}\right).$$
(iv) When $\delta_i = 1$ for all i,
$$\begin{aligned} \hat{p}_i &= \frac{1}{n-i+1} \prod_{j=1}^{i-1} \frac{n-j}{n-j+1} \\ &= \frac{1}{n-i+1} \frac{n-1}{n} \frac{n-2}{n-1} \cdots \frac{n-i+1}{n-i+2} \\ &= \frac{1}{n}. \ \blacksquare \end{aligned}$$

Exercise 9 (#5.16). Let $(X_1, ..., X_n)$ be a random sample of random variables with Lebesgue density f,

$$F_n(t) = \frac{1}{n}\sum_{i=1}^n I_{(-\infty,t]}(X_i), \quad t \in \mathcal{R},$$

be the empirical distribution, and

$$f_n(t) = \frac{F_n(t+\lambda_n) - F_n(t-\lambda_n)}{2\lambda_n}, \qquad t \in \mathcal{R},$$

where $\{\lambda_n\}$ is a sequence of positive constants.
(i) Show that f_n is a Lebesgue density on $\mathcal{R}$.
(ii) Suppose that f is continuously differentiable at t, $\lim_n \lambda_n = 0$, and $\lim_n n\lambda_n = \infty$. Show that the mean squared error of $f_n(t)$ as an estimator of $f(t)$ equals

$$\frac{f(t)}{2n\lambda_n} + o\left(\frac{1}{n\lambda_n}\right) + O(\lambda_n^2)$$

as $n \to \infty$.
(iii) Under $\lim_n n\lambda_n^3 = 0$ and the conditions in (ii), show that

$$\sqrt{n\lambda_n}[f_n(t) - f(t)] \to_d N(0, f(t)/2).$$

(iv) Suppose that f is continuous on $[a, b]$, $-\infty < a < b < \infty$, $\lim_n \lambda_n = 0$, and $\lim_n n\lambda_n = \infty$. Show that

$$\int_a^b f_n(t)dt \to_p \int_a^b f(t)dt.$$

Solution. (i) Clearly, $f_n(t) \geq 0$. Note that

$$f_n(t) = \frac{1}{2n\lambda_n}\sum_{i=1}^n I_{(X_i-\lambda_n, X_i+\lambda_n]}(t).$$

Therefore

$$\int_{-\infty}^{\infty} f_n(t)dt = \frac{1}{2n\lambda_n}\sum_{i=1}^n \int_{X_i-\lambda_n}^{X_i+\lambda_n} dt = \frac{1}{2n\lambda_n}\sum_{i=1}^n 2\lambda_n = 1.$$

(ii) Note that $2n\lambda_n f_n(t)$ has the binomial distribution with size n and probability $F(t+\lambda_n) - F(t-\lambda_n)$, where F is the cumulative distribution function of X_1. Then

$$E[f_n(t)] = \frac{F(t+\lambda_n) - F(t-\lambda_n)}{2\lambda_n} = f(t) + O(\lambda_n)$$

and

$$\begin{aligned}\mathrm{Var}(f_n(t)) &= \frac{F(t+\lambda_n) - F(t-\lambda_n) - [F(t+\lambda_n) - F(t-\lambda_n)]^2}{4n\lambda_n^2} \\ &= \frac{f(t)}{2n\lambda_n} + O\left(\frac{1}{n}\right) + \frac{[O(\lambda_n) + O(\lambda_n^2)]^2}{4n\lambda_n^2} \\ &= \frac{f(t)}{2n\lambda_n} + o\left(\frac{1}{n\lambda_n}\right),\end{aligned}$$

since $\lim_n \lambda_n = 0$. Therefore, the mean squared error is

$$\mathrm{Var}(f_n(t)) + \{E[f_n(t)] - f(t)\}^2 = \frac{f(t)}{2n\lambda_n} + o\left(\frac{1}{n\lambda_n}\right) + O(\lambda_n^2).$$

(iii) Since $2n\lambda_n f_n(t)$ has the binomial distribution, by the central limit theorem,

$$\frac{f_n(t) - E[f_n(t)]}{\sqrt{\mathrm{Var}(f_n(t))}} \to_d N(0,1).$$

From part (ii) of the solution, $n\lambda_n \mathrm{Var}(f_n(t)) = f(t)/2 + o(1)$. Hence,

$$\sqrt{n\lambda_n}\{f_n(t) - E[f_n(t)]\} \to_d N(0, f(t)/2).$$

From part (ii) of the solution,

$$\sqrt{n\lambda_n}\{E[f_n(t)] - f(t)\} = O\left(\sqrt{n\lambda_n^3}\right) = o(1)$$

under the given condition. Hence,

$$\sqrt{n\lambda_n}[f_n(t) - f(t)] \to_d N(0, f(t)/2).$$

(iv) Note that

$$E\int_a^b f_n(t)dt = \int_a^b \frac{F(t+\lambda_n) - F(t-\lambda_n)}{2\lambda_n}dt = \int_a^b f(\xi_{t,n})dt$$

by the mean value theorem, where $|\xi_{t,n} - t| \leq \lambda_n$ and, hence, $\lim_n \xi_{t,n} = t$. From the continuity of f, $f(\xi_{t,n})$ is bounded and $\lim_n f(\xi_{t,n}) = f(t)$. By the dominated convergence theorem,

$$\lim_n E\int_a^b f_n(t)dt = \int_a^b f(t)dt.$$

Because

$$f_n(t) = \frac{1}{2n\lambda_n}\sum_{i=1}^n I_{[t-\lambda_n, t+\lambda_n)}(X_i),$$

for any $t < s$,

$$\begin{aligned} f_n(t)f_n(s) &= \frac{1}{4n^2\lambda_n^2}\sum_{i=1}^n\sum_{j=1}^n I_{[t-\lambda_n,t+\lambda_n)}(X_i)I_{[s-\lambda_n,s+\lambda_n)}(X_j) \\ &= \frac{1}{4n^2\lambda_n^2}\sum_{i=1}^n I_{[s-\lambda_n,t+\lambda_n)}(X_i) \\ &\quad + \frac{1}{4n^2\lambda_n^2}\sum_{i\neq j} I_{[t-\lambda_n,t+\lambda_n)}(X_i)I_{[s-\lambda_n,s+\lambda_n)}(X_j). \end{aligned}$$

Then,

$$\begin{aligned} E[f_n(t)f_n(s)] &= \frac{E[I_{[s-\lambda_n,t+\lambda_n)}(X_1)]}{4n\lambda_n^2} \\ &\quad + \frac{(n-1)[F(t+\lambda_n)-F(t-\lambda_n)][F(s+\lambda_n)-F(s-\lambda_n)]}{4n\lambda_n^2} \\ &= \frac{\max\{0, F(t+\lambda_n)-F(s-\lambda_n)\}}{4n\lambda_n^2} + \frac{(n-1)f(\xi_{t,n})f(\eta_{s,n})}{n}, \end{aligned}$$

where $|\xi_{t,n} - t| \le \lambda_n$ and $|\eta_{s,n} - s| \le \lambda_n$. By the continuity of f and the fact that $t < s$,

$$\lim_n E[f_n(t)f_n(s)] = f(t)f(s).$$

Then, by Fubini's theorem and the dominated convergence theorem,

$$\begin{aligned} \lim_n E\left[\int_a^b f_n(t)dt\right]^2 &= \lim_n E\left[\int_a^b f_n(t)dt\right]\left[\int_a^b f_n(s)ds\right] \\ &= \lim_n \int_a^b\int_a^b E[f_n(t)f_n(s)]dtds \\ &= \int_a^b\int_a^b f(t)f(s)dtds \\ &= \left[\int_a^b f(t)dt\right]^2. \end{aligned}$$

Combining this result and the previous result, we conclude that

$$\lim_n \mathrm{Var}\left(\int_a^b f_n(t)dt\right) = 0$$

and, therefore, $\int_a^b f_n(t)dt \to_p \int_a^b f(t)dt$. ∎

Exercise 10 (#5.17). Let $(X_1, ..., X_n)$ be a random sample of random variables with Lebesgue density f, w be a known Lebesgue density on $\mathcal{R}$, and

$$\hat{f}(t) = \frac{1}{n\lambda_n}\sum_{i=1}^{n} w\left(\frac{t-X_i}{\lambda_n}\right),$$

where $\{\lambda_n\}$ is a sequence of positive constants.
(i) Show that $\hat{f}$ is a Lebesgue density on $\mathcal{R}$.
(ii) Show that if $\lambda_n \to 0$, $n\lambda_n \to \infty$, and f is bounded and continuous at t and $w_0 = \int_{-\infty}^{\infty}[w(t)]^2 dt < \infty$, then

$$\sqrt{n\lambda_n}\{\hat{f}(t) - E[\hat{f}(t)]\} \to_d N\big(0, w_0 f(t)\big).$$

(iii) Assume that $\lim_n n\lambda_n^3 = 0$, $\int_{-\infty}^{\infty}|t|w(t)dt < \infty$, f' is bounded and continuous at t, and the conditions in (ii) hold. Show that

$$\sqrt{n\lambda_n}[\hat{f}(t) - f(t)] \to_d N\big(0, w_0 f(t)\big).$$

(iv) Suppose that $\lambda_n \to 0$, $n\lambda_n \to \infty$, w is bounded, and f is bounded and continuous on $[a, b]$, $-\infty < a < b < \infty$. Show that $\int_a^b \hat{f}(t)dt \to_p \int_a^b f(t)dt$.
Solution. (i) The result follows from

$$\int_{-\infty}^{\infty} \hat{f}(t)dt = \frac{1}{n\lambda_n}\sum_{i=1}^{n}\int_{-\infty}^{\infty} w\left(\frac{t-x}{\lambda_n}\right)dt = \int_{-\infty}^{\infty} w(y)dy = 1.$$

(ii) Let $Y_{in} = w\left(\frac{t-X_i}{\lambda_n}\right)$. Then $Y_{1n}, ..., Y_{nn}$ are independent and identically distributed with

$$E(Y_{1n}) = \int_{-\infty}^{\infty} w\left(\frac{t-x}{\lambda_n}\right)f(x)dx = \lambda_n\int_{-\infty}^{\infty} w(y)f(t-\lambda_n y)dy = O\left(\lambda_n\right)$$

and

$$\begin{aligned}\mathrm{Var}(Y_{1n}) &= \int_{-\infty}^{\infty}\left[w\left(\frac{t-x}{\lambda_n}\right)\right]^2 f(x)dx - \left[\int_{-\infty}^{\infty} w\left(\frac{t-x}{\lambda_n}\right)f(x)dx\right]^2 \\ &= \lambda_n\int_{-\infty}^{\infty}[w(y)]^2 f(t-\lambda_n y)dy + O(\lambda_n^2) \\ &= \lambda_n w_0 f(t) + o(\lambda_n),\end{aligned}$$

since f is bounded and continuous at t and $w_0 = \int_{-\infty}^{\infty}[w(t)]^2 dt < \infty$. Then

$$\mathrm{Var}\big(\hat{f}(t)\big) = \frac{1}{n^2\lambda_n^2}\sum_{i=1}^{n}\mathrm{Var}(Y_{in}) = \frac{w_0 f(t)}{n\lambda_n} + o\left(\frac{1}{n\lambda_n}\right).$$

Note that $\hat{f}(t) - E\hat{f}(t) = \sum_{i=1}^n [Y_{in} - E(Y_{in})]/(n\lambda_n)$. To apply Lindeberg's central limit theorem to $\hat{f}(t)$, we note that, for any $\epsilon > 0$,

$$\frac{E(Y_{1n}^2 I_{\{|Y_{1n}-E(Y_{1n})|>\epsilon\sqrt{n\lambda_n}\}})}{\lambda_n} = \int_{|w(y)-E(Y_{1n})|>\epsilon\sqrt{n\lambda_n}} [w(y)]^2 f(t-\lambda_n y)dy,$$

which converges to 0 under the given conditions.
(iii) Note that

$$\begin{aligned} E[\hat{f}(t)] - f(t) &= \lambda_n^{-1} E(Y_{1n}) - f(t) \\ &= \int_{-\infty}^{\infty} w(y)[f(t-\lambda_n y) - f(t)]dy \\ &= \lambda_n \int_{-\infty}^{\infty} y w(y) f'(\xi_{t,y,n})dy, \end{aligned}$$

where $|\xi_{t,y,n} - t| \leq \lambda_n$. Under the condition that f' is bounded and continuous at t and $\int_{-\infty}^{\infty} |y|w(y)dy < \infty$,

$$\lim_n \sqrt{n\lambda_n}\{E[\hat{f}(t)] - f(t)\} = \lim_n \sqrt{n\lambda_n} O(\lambda_n) = 0.$$

Hence the result follows from the result in part (ii).
(iv) Since f is bounded and continuous,

$$\begin{aligned} \lim_n E \int_a^b \hat{f}(t)dt &= \lim_n \int_a^b E[\hat{f}(t)]dt \\ &= \lim_n \int_a^b \int_{-\infty}^{\infty} w(y) f(t-\lambda_n y)dydt \\ &= \int_a^b \int_{-\infty}^{\infty} w(y) f(t)dydt \\ &= \int_a^b f(t)dt. \end{aligned}$$

For $t \neq s$,

$$\begin{aligned} E[\hat{f}(t)\hat{f}(s)] &= \frac{1}{n^2\lambda_n^2} E\left[\sum_{i=1}^n \sum_{j=1}^n w\left(\frac{t-X_i}{\lambda_n}\right) w\left(\frac{s-X_j}{\lambda_n}\right)\right] \\ &= \frac{1}{n\lambda_n^2} E\left[w\left(\frac{t-X_1}{\lambda_n}\right) w\left(\frac{s-X_1}{\lambda_n}\right)\right] \\ &\quad + \frac{n-1}{n\lambda_n^2} E\left[w\left(\frac{t-X_1}{\lambda_n}\right)\right] E\left[w\left(\frac{s-X_1}{\lambda_n}\right)\right] \\ &= \frac{1}{n\lambda_n} \int_{-\infty}^{\infty} w\left(\frac{t-s+\lambda_n y}{\lambda_n}\right) w(y) f(s-\lambda_n y)dy \\ &\quad + \frac{n-1}{n} \int_{-\infty}^{\infty} w(y) f(t-\lambda_n y)dy \int_{-\infty}^{\infty} w(y) f(s-\lambda_n y)dy, \end{aligned}$$

which converges to $f(t)f(s)$ under the given conditions. Hence,

$$\lim_n \mathrm{Var}\left(\int_a^b \hat{f}(t)dt\right) = 0$$

and, therefore, $\int_a^b \hat{f}(t)dt \to_p \int_a^b f(t)dt$. ∎

Exercise 11 (#5.20). Let $\ell(\theta,\xi)$ be a likelihood function. Show that a maximum profile likelihood estimator $\hat{\theta}$ of θ is an MLE if $\xi(\theta)$, the maximum of $\sup_\xi \ell(\theta,\xi)$ for a fixed θ, does not depend on θ.

Note. A maximum profile likelihood estimator $\hat{\theta}$ maximizes the profile likelihood function $\ell_P(\theta) = \ell(\theta,\xi(\theta))$, where $\ell(\theta,\xi(\theta)) = \sup_\xi \ell(\theta,\xi)$ for each fixed θ.

Solution. Suppose that $\hat{\xi}$ satisfies $\ell(\theta,\hat{\xi}) = \sup_\xi \ell(\theta,\xi)$ for any θ. Then the profile likelihood function is $\ell_P(\theta) = \ell(\theta,\hat{\xi})$. If $\hat{\theta}$ satisfies $\ell_P(\hat{\theta}) = \sup_\theta \ell_P(\theta)$, then $\ell(\hat{\theta},\hat{\xi}) = \ell_P(\hat{\theta}) \geq \ell_P(\theta) = \ell(\theta,\hat{\xi}) \geq \ell(\theta,\xi)$ for any θ and ξ. Hence, $(\hat{\theta},\hat{\xi})$ is an MLE of (θ,ξ). ∎

Exercise 12 (#5.21). Let $(X_1,...,X_n)$ be a random sample from $N(\mu,\sigma^2)$. Derive the profile likelihood function for μ or σ^2. Discuss in each case whether the maximum profile likelihood estimator is the same as the MLE.

Solution. The likelihood function is

$$\ell(\mu,\sigma^2) = (2\pi)^{-n/2}(\sigma^2)^{-n/2}\exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^n (X_i-\mu)^2\right\}.$$

For fixed σ^2, $\ell(\mu,\sigma^2) \leq \ell(\bar{X},\sigma^2)$, since $\sum_{i=1}^n (X_i-\mu)^2 \geq \sum_{i=1}^n (X_i-\bar{X})^2$, where $\bar{X}$ is the sample mean. Hence the maximum does not depend on σ^2 and the profile likelihood function is

$$\ell(\bar{X},\sigma^2) = (2\pi)^{-n/2}(\sigma^2)^{-n/2}\exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^n (X_i-\bar{X})^2\right\}.$$

By the result in the previous exercise, the profile MLE of σ^2 is the same as the MLE of σ^2. This can also be shown by directly verifying that $\ell(\bar{X},\sigma^2)$ is maximized at $\hat{\sigma}^2 = n^{-1}\sum_{i=1}^n (X_i-\bar{X})^2$.

For fixed μ, $\ell(\mu,\sigma^2)$ is maximized at $\sigma^2(\mu) = n^{-1}\sum_{i=1}^n (X_i-\mu)^2$. Then the profile likelihood function is

$$\ell(\mu,\sigma^2(\mu)) = (2\pi)^{-n/2}e^{-n/2}\left[\frac{n}{\sum_{i=1}^n (X_i-\mu)^2}\right]^{n/2}.$$

Since $\sum_{i=1}^n (X_i-\mu)^2 \geq \sum_{i=1}^n (X_i-\bar{X})^2$, $\ell(\mu,\sigma^2(\mu))$ is maximized at $\bar{X}$, which is the same as the MLE of μ (although $\sigma^2(\mu)$ depends on μ). ∎

Exercise 13 (#5.23). Let $(X_1,...,X_n)$ be a random sample from a distribution F and let $\pi(x) = P(\delta_i = 1|X_i = x)$, where $\delta_i = 1$ if X_i is observed and $\delta_i = 0$ if X_i is missing. Assume that $0 < \pi = \int \pi(x)dF(x) < 1$.
(i) Let $F_1(x) = P(X_i \le x|\delta_i = 1)$. Show that F and F_1 are the same if and only if $\pi(x) \equiv \pi$.
(ii) Let $\hat{F}$ be the empirical distribution putting mass r^{-1} to each observed X_i, where r is the number of observed X_i's. Show that $\hat{F}(x)$ is unbiased and consistent for $F_1(x)$, $x \in \mathcal{R}$.
(iii) When $\pi(x) \equiv \pi$, show that $\hat{F}(x)$ in part (ii) is unbiased and consistent for $F(x)$, $x \in \mathcal{R}$. When $\pi(x)$ is not constant, show that $\hat{F}(x)$ is biased and inconsistent for $F(x)$ for some $x \in \mathcal{R}$.
Solution. (i) If $\pi(x) \equiv \pi$, then X_i and δ_i are independent. Hence, $F_1(x) = P(X_i \le x|\delta_i = 1) = P(X_i \le x) = F(x)$ for any x. If $F_1(x) = F(x)$ for any x, then $P(X_i \le x, \delta_i = 1) = P(X_i \le x)P(\delta_i = 1)$ for any x and, hence, X_i and δ_i are independent. Thus, $\pi(x) \equiv \pi$.
(ii) Note that

$$\hat{F}(x) = \frac{\sum_{i=1}^n \delta_i I_{(-\infty,x]}(X_i)}{\sum_{i=1}^n \delta_i}.$$

Since $E[\delta_i I_{(-\infty,x]}(X_i)|\delta_i] = \delta_i F_1(x)$, we obtain that

$$\begin{aligned} E[\hat{F}(x)] &= E\{E[\hat{F}(x)|\delta_1,...,\delta_n]\} \\ &= E\left\{\frac{\sum_{i=1}^n E[\delta_i I_{(-\infty,x]}(X_i)|\delta_i]}{\sum_{i=1}^n \delta_i}\right\} \\ &= E\left\{\frac{\sum_{i=1}^n \delta_i F_1(x)}{\sum_{i=1}^n \delta_i}\right\} \\ &= F_1(x), \end{aligned}$$

i.e., $\hat{F}(x)$ is unbiased. From the law of large numbers,

$$\frac{1}{n}\sum_{i=1}^n \delta_i I_{(-\infty,x]}(X_i) \to_p E[\delta_1 I_{(-\infty,x]}(X_1)] = E[\delta_1 F_1(x)] = \pi F_1(x)$$

and

$$\frac{1}{n}\sum_{i=1}^n \delta_i \to_p E(\delta_1) = \pi.$$

Hence, $\hat{F}(x) \to_p F_1(x)$.
(iii) When $\pi(x) \equiv \pi$, $F(x) = F_1(x)$ (part (i)). Hence, $\hat{F}(x)$ is unbiased and consistent for $F(x)$ (part (ii)). When $\pi(x)$ is not constant, $F_1(x) \ne F(x)$ for some x (part (i)). Since $\hat{F}(x)$ is unbiased and consistent for $F_1(x)$ (part (ii)), it is biased and inconsistent for $F(x)$ for x at which $F(x) \ne F_1(x)$. ∎

Exercise 14 (#5.25). Let $\mathcal{F}$ be a collection of distributions on $\mathcal{R}^d$. A functional T defined on $\mathcal{F}$ is Gâteaux differentiable at $G \in \mathcal{F}$ if and only if there is a linear functional L_G on $\mathcal{D} = \{c(G_1 - G_2) : c \in \mathcal{R}, G_j \in \mathcal{F}, j = 1, 2\}$ (i.e., $L_G(c_1\Delta_1 + c_2\Delta_2) = c_1 L_G(\Delta_1) + c_2 L_G(\Delta_2)$ for any $\Delta_j \in \mathcal{D}$ and $c_j \in \mathcal{R}$) such that $\Delta \in \mathcal{D}$ and $G + t\Delta \in \mathcal{F}$ imply

$$\lim_{t\to 0}\left[\frac{T(G+t\Delta) - T(G)}{t} - L_G(\Delta)\right] = 0.$$

Assume that the functional L_F is continuous in the sense that $\|\Delta_j - \Delta\|_\infty \to 0$ implies $L_F(\Delta_j) \to L_F(\Delta)$, where $D \in \mathcal{D}$, $D_j \in \mathcal{D}$, and $\|D\|_\infty = \sup_x |D(x)|$ for any $D \in \mathcal{D}$ is the sup-norm. Show that $\phi_F(x) = L_F(\delta_x - F)$ is a bounded function of x, where δ_x is the degenerated distribution at x.
Solution. Suppose that ϕ_F is unbounded. Then, there exists a sequence $\{x_n\}$ of numbers such that $\lim_n |\phi_F(x_n)| = \infty$. Let $t_n = |\phi_F(x_n)|^{-1/2}$ and $H_n = t_n(\delta_{x_n} - F)$. Then $H_n \in \mathcal{D}$ and by the linearity of L_F,

$$|L_F(H_n)| = t_n|L(\delta_{x_n} - F)| = t_n|\phi_F(x_n)| = |\phi_F(x_n)|^{1/2} \to \infty$$

as $n \to \infty$. On the other hand, $\|H_n\|_\infty \le t_n \to 0$ implies $L_F(H_n) \to L(0)$ if L_F is continuous. This contradiction shows that ϕ_F is bounded. ∎

Exercise 15 (#5.26). Suppose that a functional T is Gâteaux differentiable at F with a bounded and continuous influence function $\phi_F(x) = L_F(\delta_x - F)$, where δ_x is the degenerated distribution at x. Show that L_F is continuous in the sense described in the previous exercise.
Solution. From the linearity of L_F,

$$\int \phi_F(x)dG = \int L_F(\delta_x - F)dG = L_F\left(\int \delta_x dG - F\right) = L_F(G - F)$$

for any distribution G. Hence,

$$L_F(D) = \int \phi_F(x)dD, \quad D \in \mathcal{D}.$$

If $\|D_j - D\|_\infty \to 0$, then, since ϕ_F is bounded and continuous, $\int \phi_F(x)dD_j \to \int \phi_F(x)dD$. Hence, $L_F(D_j) \to L_F(D)$. ∎

Exercise 16 (#5.29). Let $\mathcal{F}$ be the collection of all distributions on $\mathcal{R}$ and z be a fixed real number. Define

$$T(G) = \int G(z-y)dG(y), \quad G \in \mathcal{F}.$$

Obtain the influence function ϕ_F for T and show that ϕ_F is continuous if and only if F is continuous.
Solution. For $G \in \mathcal{F}$ and $\Delta \in \mathcal{D} = \{c(G_1 - G_2) : c \in \mathcal{R}, G_j \in \mathcal{F}, j = 1, 2\}$,

$$\begin{aligned} T(G+t\Delta) - T(G) &= \int (G+t\Delta)(z-y)d(G+t\Delta)(y) - \int G(x-y)dG(y) \\ &= 2t\int \Delta(z-y)dG(y) + t^2\int \Delta(z-y)d\Delta(y). \end{aligned}$$

Hence,

$$\lim_{t\to 0}\frac{T(G+t\Delta)-T(G)}{t}=2\int\Delta(z-y)dG(y)$$

and the influence function is

$$\phi_F(x)=2\int(\delta_x-F)(z-y)dF(y)=2\left[F(z-x)-\int F(z-y)dF(y)\right],$$

where δ_x is the degenerated distribution at x. Hence ϕ_F is continuous if and only if F is continuous. ∎

Exercise 17 (#5.34). An L-functional is defined as

$$T(G)=\int xJ(G(x))dG(x),\quad G\in\mathcal{F}_0,$$

where $\mathcal{F}_0$ contains all distributions on $\mathcal{R}$ for which T is well defined and $J(t)$ is a Borel function on $[0,1]$.
(i) Show that the influence function is

$$\phi_F(x)=-\int_{-\infty}^{\infty}(\delta_x-F)(y)J(F(y))dy,$$

where δ_x is the degenerated distribution at x.
(ii) Show that $\int\phi_F(x)dF(x)=0$ and, if J is bounded and F has a finite second moment, then $\int[\phi_F(x)]^2dF(x)<\infty$.
Solution. (i) For F and G in $\mathcal{F}_0$,

$$\begin{aligned}T(G)-T(F)&=\int xJ(G(x))dG(x)-\int xJ(F(x))dF(x)\\&=\int_0^1[G^{-1}(t)-F^{-1}(t)]J(t)dt\\&=\int_0^1\int_{F^{-1}(t)}^{G^{-1}(t)}dxJ(t)dt\\&=\int_{-\infty}^{\infty}\int_{G(x)}^{F(x)}J(t)dtdx\\&=\int_{-\infty}^{\infty}[F(x)-G(x)]J(F(x))dx\\&\quad-\int_{-\infty}^{\infty}U_G(x)[G(x)-F(x)]J(F(x))dx,\end{aligned}$$

where

$$U_G(x)=\begin{cases}\frac{\int_{F(x)}^{G(x)}J(t)dt}{[G(x)-F(x)]J(F(x))}-1 & G(x)\neq F(x),J(F(x))\neq 0\\ 0 & \text{otherwise}\end{cases}$$

and the fourth equality follows from Fubini's theorem and the fact that the region in $\mathcal{R}^2$ between curves $F(x)$ and $G(x)$ is the same as the region in $\mathcal{R}^2$ between curves $G^{-1}(t)$ and $F^{-1}(t)$. Then, for any $\Delta \in \mathcal{D} = \{c(G_1 - G_2) : c \in \mathcal{R}, G_j \in \mathcal{F}, j = 1, 2\}$,

$$\lim_{t\to 0} \frac{T(F + t\Delta) - T(F)}{t} = -\int_{-\infty}^{\infty} \Delta(x) J(F(x))dx,$$

since $\lim_{t\to 0} U_{F+t\Delta}(x) = 0$ and, by the dominated convergence theorem,

$$\lim_{t\to 0} \int_{-\infty}^{\infty} U_{F+t\Delta}\Delta(x)J(F(x))dx = 0.$$

Letting $\Delta = \delta_x - F$, we obtain the influence function as claimed.
(ii) By Fubini's theorem,

$$\int \phi_F(x)dF(x) = -\int_{-\infty}^{\infty} \left[\int (\delta_x - F)(y)dF(x)\right] J(F(y))dy = 0,$$

since $\int \delta_x(y)dF(x) = F(y)$. Suppose now that $|J| < C$ for a constant C. Then

$$\begin{aligned} |\phi_F(x)| &\le C \int_{-\infty}^{\infty} |\delta_x(y) - F(y)|dy \\ &= C\left(\int_{-\infty}^{x} F(y)dy + \int_{x}^{\infty} [1 - F(y)]dy\right) \\ &\le C\left(|x| + \int_{-\infty}^{0} F(y)dy + \int_{0}^{\infty} [1 - F(y)]dy\right) \\ &= C(|x| + E|X|), \end{aligned}$$

where X is the random variable having distribution F. Thus,

$$[\phi_F(x)]^2 \le C^2(|x| + E|X|)^2$$

and $\int [\phi_F(x)]^2 dF(x) < \infty$ when $EX^2 < \infty$. ∎

Exercise 18 (#5.37). Obtain explicit forms of the influence functions for L-functionals (Exercise 17) in the following cases and discuss which of them are bounded and continuous.
(i) $J \equiv 1$.
(ii) $J(t) = 4t - 2$.
(iii) $J(t) = (\beta - \alpha)^{-1} I_{(\alpha,\beta)}(t)$ for some constants $\alpha < \beta$.
Solution. (i) When $J \equiv 1$, $T(G) = \int x dG(x)$ is the mean functional ($\mathcal{F}_0$ is the collection of all distributions with finite means). From the previous

exercise, the influence function is

$$\begin{aligned}\phi_F(x) &= -\int_{-\infty}^{\infty}[\delta_x(y) - F(y)]dy \\ &= \int_{-\infty}^{x} F(y)dy - \int_{x}^{\infty}[1 - F(y)]dy \\ &= \int_{0}^{x} dy + \int_{-\infty}^{0} F(y)dy - \int_{0}^{\infty}[1 - F(y)]dy \\ &= x - \int_{-\infty}^{\infty} ydF(y).\end{aligned}$$

This influence function is continuous, but not bounded.
(ii) When $J(t) = 4t - 2$,

$$\phi_F(x) = 2\int_{-\infty}^{x} F(y)[2F(y) - 1]dy - 2\int_{x}^{\infty}[1 - F(y)][2F(y) - 1]dy.$$

Clearly, ϕ_F is continuous. Since

$$\lim_{x\to\infty}\int_{-\infty}^{x} F(y)[2F(y) - 1]dy = \int_{-\infty}^{\infty} F(y)[2F(y) - 1]dy = \infty$$

and

$$\lim_{x\to\infty}\int_{x}^{\infty}[1 - F(y)][2F(y) - 1]dy = 0,$$

we conclude that $\lim_{x\to\infty}\phi_F(x) = \infty$. Similarly, $\lim_{x\to-\infty}\phi_F(x) = -\infty$. Hence, ϕ_F is not bounded.
(iii) When $J(t) = (\beta - \alpha)^{-1}I_{(\alpha,\beta)}(t)$,

$$\phi_F(x) = -\frac{1}{\beta - \alpha}\int_{F^{-1}(\alpha)}^{F^{-1}(\beta)}[\delta_x(y) - F(y)]dy,$$

which is continuous. ϕ_F is also bounded, since

$$|\phi_F(x)| \leq \frac{1}{\beta - \alpha}\int_{F^{-1}(\alpha)}^{F^{-1}(\beta)}|\delta_x(y) - F(y)|dy \leq \frac{F^{-1}(\beta) - F^{-1}(\alpha)}{\beta - \alpha}. \blacksquare$$

Exercise 19. In part (iii) of the previous exercise, show that if F is continuous at $F^{-1}(\alpha)$ and $F^{-1}(\beta)$, then

$$\phi_F(x) = \begin{cases} \frac{F^{-1}(\alpha)(1-\alpha) - F^{-1}(\beta)(1-\beta)}{\beta-\alpha} - T(F) & x < F^{-1}(\alpha) \\ \frac{x - F^{-1}(\alpha)\alpha - F^{-1}(\beta)(1-\beta)}{\beta-\alpha} - T(F) & F^{-1}(\alpha) \leq x \leq F^{-1}(\beta) \\ \frac{F^{-1}(\beta)\beta - F^{-1}(\alpha)\alpha}{\beta-\alpha} - T(F) & x > F^{-1}(\beta). \end{cases}$$

Solution. When $x < F^{-1}(\alpha)$,

$$\begin{aligned}\phi_F(x) &= -\frac{1}{\beta-\alpha}\int_{F^{-1}(\alpha)}^{F^{-1}(\beta)}[1-F(y)]dy \\ &= -\frac{y[1-F(y)]}{\beta-\alpha}\bigg|_{F^{-1}(\alpha)}^{F^{-1}(\beta)} - \frac{1}{\beta-\alpha}\int_{F^{-1}(\alpha)}^{F^{-1}(\beta)} ydF(y) \\ &= \frac{F^{-1}(\alpha)(1-\alpha)-F^{-1}(\beta)(1-\beta)}{\beta-\alpha} - T(F),\end{aligned}$$

since $F(F^{-1}(\alpha)) = \alpha$ and $F(F^{-1}(\beta)) = \beta$. Similarly, when $x > F^{-1}(\beta)$,

$$\begin{aligned}\phi_F(x) &= \frac{1}{\beta-\alpha}\int_{F^{-1}(\alpha)}^{F^{-1}(\beta)} F(y)]dy \\ &= \frac{yF(y)}{\beta-\alpha}\bigg|_{F^{-1}(\alpha)}^{F^{-1}(\beta)} - \frac{1}{\beta-\alpha}\int_{F^{-1}(\alpha)}^{F^{-1}(\beta)} ydF(y) \\ &= \frac{F^{-1}(\beta)\beta - F^{-1}(\alpha)\alpha}{\beta-\alpha} - T(F).\end{aligned}$$

Finally, consider the case of $F^{-1}(\alpha) \leq x \leq F^{-1}(\beta)$. Then

$$\begin{aligned}\phi_F(x) &= \frac{1}{\beta-\alpha}\int_{F^{-1}(\alpha)}^{x} F(y)dy - \frac{1}{\beta-\alpha}\int_{x}^{F^{-1}(\beta)}[1-F(y)]dy \\ &= \frac{yF(y)}{\beta-\alpha}\bigg|_{F^{-1}(\alpha)}^{x} - \frac{1}{\beta-\alpha}\int_{F^{-1}(\alpha)}^{x} ydF(y) \\ &\quad + \frac{y[1-F(y)]}{\beta-\alpha}\bigg|_{x}^{F^{-1}(\beta)} - \frac{1}{\beta-\alpha}\int_{x}^{F^{-1}(\beta)} ydF(y) \\ &= \frac{x - F^{-1}(\alpha)\alpha - F^{-1}(\beta)(1-\beta)}{\beta-\alpha} - T(F). \quad \blacksquare\end{aligned}$$

Exercise 20 (#5.67, #5.69, #5.74). Let T be an L-functional defined in Exercise 17.
(i) Show that $T(F) = \theta$ if F is symmetric about θ, J is symmetric about $\frac{1}{2}$, and $\int_0^1 J(t)dt = 1$.
(ii) Assume that

$$\sigma_F^2 = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} J(F(x))J(F(y))[F(\min\{x,y\}) - F(x)F(y)]dxdy$$

is finite. Show that $\sigma^2 = \int[\phi_F(x)]^2 dF(x)$, where ϕ_F is the influence function of T.

(iii) Show that if $J \equiv 1$, then σ_F^2 in (ii) is equal to the variance of F.

Solution. (i) If F is symmetric about θ, then $F(x) = F_0(x-\theta)$, where F_0 is a cumulative distribution function that is symmetric about 0, i.e., $F_0(x) = 1 - F_0(-x)$. Also, $J(t) = J(1-t)$. Then

$$\begin{aligned}\int xJ(F_0(x))dF_0(x) &= \int xJ(1-F_0(-x))dF_0(x)\\ &= \int xJ(F_0(-x))dF_0(x)\\ &= -\int yJ(F_0(y))dF_0(y),\end{aligned}$$

i.e., $\int xJ(F_0(x))dF_0(x) = 0$. Hence,

$$\begin{aligned}T(F) &= \int xJ(F(x))dF(x)\\ &= \theta\int J(F(x))dF(x) + \int (x-\theta)J(F_0(x-\theta))dF_0(x-\theta)\\ &= \theta\int_0^1 J(t)dt + \int yJ(F_0(y))dF_0(y)\\ &= \theta.\end{aligned}$$

(ii) From Exercise 17, $\phi_F(x) = -\int_{-\infty}^{\infty}[\delta_x(y) - F(y)]J(F(y))dy$. Then

$$\begin{aligned}[\phi_F(t)]^2 &= \left[\int_{-\infty}^{\infty}(\delta_t - F)(y)J(F(y))dy\right]^2\\ &= \int_{-\infty}^{\infty}(\delta_t - F)(y)J(F(y))dy\int_{-\infty}^{\infty}(\delta_t - F)(x)J(F(x))dx\\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(\delta_t - F)(x)(\delta_t - F)(y)J(F(x))J(F(y))dxdy.\end{aligned}$$

Then the result follows from Fubini's theorem and the fact that

$$\int (\delta_t - F)(x)(\delta_t - F)(y)dF(t) = F(\min\{x,y\}) - F(x)F(y).$$

(iii) When $J \equiv 1$, by part (i) of the solution to the previous exercise, $\phi_F(x) = x - \int ydF(y)$. Hence, $\int[\phi_F(x)]^2dF(x)$ is the variance of F when $J \equiv 1$. The result follows from part (ii). ∎

Exercise 21 (#5.65, #5.72, #5.73). Let T be an L-functional given in Exercise 18(iii) with $\beta = 1-\alpha$ and $\alpha \in (0, \frac{1}{2})$, and let F_n be the empirical distribution based on a random sample from a distribution F.

(i) Let $X_{(1)} \le \cdots \le X_{(n)}$ be the order statistics. Show that

$$T(F_n) = \bar{X}_\alpha = \frac{1}{(1-2\alpha)n}\sum_{j=m_\alpha+1}^{n-m_\alpha} X_{(j)},$$

which is called the α-trimmed sample mean.
(ii) Assume that F is continuous at $F^{-1}(\alpha)$ and $F^{-1}(1-\alpha)$ and is symmetric about θ. Show that

$$\sigma_\alpha^2 = \frac{2}{(1-2\alpha)^2}\left\{\int_0^{F_0^{-1}(1-\alpha)} x^2 dF_0(x) + \alpha[F_0^{-1}(1-\alpha)]^2\right\},$$

where $F_0(x-\theta) = F(x)$, is equal to the σ_F^2 in part (ii) of the previous exercise with $J(t) = (1-2\alpha)^{-1}I_{(\alpha,1-\alpha)}(t)$.
(iii) Show that if $F_0'(0)$ exists and is positive, then $\lim_{\alpha\to\frac{1}{2}}\sigma_\alpha^2 = 1/[2F_0'(0)]^2$.
(iv) Show that if $\sigma^2 = \int x^2 dF_0(x) < \infty$, then $\lim_{\alpha\to 0}\sigma_\alpha^2 = \sigma^2$.
Solution. (i) Note that

$$T(F_n) = \int xJ(F_n(x))dF_n(x) = \frac{1}{n}\sum_{i=1}^n J\left(\tfrac{i}{n}\right)X_{(i)},$$

since $F_n(X_{(i)}) = i/n$, $i = 1, ..., n$. The result follows from the fact that $J(\frac{i}{n})$ is not 0 if and only if $m_\alpha \le i \le n - m_\alpha$.
(ii) Note that J is symmetric about $\frac{1}{2}$. If F is symmetric about θ, then $T(F) = \theta$ (Exercise 20) and $F^{-1}(\alpha) + F^{-1}(1-\alpha) = 2\theta$. From Exercise 19 with $\beta = 1-\alpha$, we conclude that

$$\phi_F(x) = \begin{cases} \frac{F_0^{-1}(\alpha)}{1-2\alpha} & x < F^{-1}(\alpha) \\ \frac{x-\theta}{1-2\alpha} & F^{-1}(\alpha) \le x \le F^{-1}(1-\alpha) \\ \frac{F_0^{-1}(1-\alpha)}{1-2\alpha} & x > F^{-1}(1-\alpha), \end{cases}$$

where $F_0^{-1}(\alpha) = F^{-1}(\alpha) - \theta$ and $F_0^{-1}(1-\alpha) = F^{-1}(1-\alpha) - \theta$. Because $F_0^{-1}(\alpha) = -F_0^{-1}(1-\alpha)$, we obtain that

$$\begin{aligned} \int [\phi_F(x)]^2 dF(x) &= \frac{[F_0^{-1}(\alpha)]^2}{(1-2\alpha)^2}\alpha + \frac{[F_0^{-1}(1-\alpha)]^2}{(1-2\alpha)^2}\alpha \\ &\quad + \int_{F^{-1}(\alpha)}^{F^{-1}(1-\alpha)} \frac{(x-\theta)^2}{(1-2\alpha)^2} dF(x) \\ &= \frac{2\alpha[F_0^{-1}(1-\alpha)]^2}{(1-2\alpha)^2} + \int_{F_0^{-1}(\alpha)}^{F_0^{-1}(1-\alpha)} \frac{x^2}{(1-2\alpha)^2} dF_0(x) \\ &= \sigma_\alpha^2. \end{aligned}$$

By Exercise 20, we conclude that $\sigma_\alpha^2 = \sigma_F^2$.
(iii) Note that $F_0^{-1}(\frac{1}{2}) = 0$. Since $F_0'(0)$ exists and $F_0'(0) > 0$,

$$F_0^{-1}(\alpha) = \frac{\alpha - \frac{1}{2}}{F_0'(0)} + R_\alpha,$$

where $\lim_{\alpha\to\frac{1}{2}} R_\alpha/(\alpha-\frac{1}{2})=0$. Then

$$[F_0^{-1}(\alpha)]^2 = \frac{(\alpha-\frac{1}{2})^2}{[F_0'(0)]^2} + U_\alpha,$$

where $\lim_{\alpha\to\frac{1}{2}} U_\alpha/(\alpha-\frac{1}{2})^2=0$. Hence,

$$\lim_{\alpha\to\frac{1}{2}} \frac{2\alpha[F_0^{-1}(1-\alpha)]^2}{(1-2\alpha)^2} = \lim_{\alpha\to\frac{1}{2}} \frac{2\alpha[F_0^{-1}(\alpha)]^2}{4(\alpha-\frac{1}{2})^2} = \frac{1}{[2F_0'(0)]^2}.$$

Note that

$$\int_0^{F_0^{-1}(1-\alpha)} x^2 dF_0(x) = \int_{\frac{1}{2}}^{1-\alpha} [F_0^{-1}(t)]^2 dt$$

and, by l'Hôpital's rule,

$$\lim_{\alpha\to\frac{1}{2}} \frac{\int_{\frac{1}{2}}^{1-\alpha}[F_0^{-1}(t)]^2 dt}{(1-2\alpha)^2} = \lim_{\alpha\to\frac{1}{2}} \frac{[F_0^{-1}(1-\alpha)]^2}{4(1-2\alpha)} = 0.$$

Hence, $\lim_{\alpha\to\frac{1}{2}} \sigma_\alpha^2 = 1/[2F_0'(0)]^2$.
(iv) Note that $\lim_{\alpha\to 0} F_0^{-1}(1-\alpha)=\infty$. Since $\int x^2 dF_0(x)<\infty$, we have $\lim_{x\to\infty} x^2[1-F(x)]=0$. Hence, $\lim_{\alpha\to 0}\alpha[F_0^{-1}(1-\alpha)]^2=0$. Then,

$$\begin{aligned}\lim_{\alpha\to 0}\sigma_\alpha^2 &= \lim_{\alpha\to 0} 2\left\{\int_0^{F_0^{-1}(1-\alpha)} x^2 dF_0(x) + \alpha[F_0^{-1}(1-\alpha)]^2\right\} \\ &= 2\int_0^\infty x^2 dF_0(x) \\ &= \sigma^2,\end{aligned}$$

since F_0 is symmetric about 0. ∎

Exercise 22 (#5.75). Calculate σ_F^2 defined in Exercise 20(ii) with $J(t) = 4t-2$ and F being the double exponential distribution with location parameter $\mu\in\mathcal{R}$ and scale parameter 1.
Solution. Note that F is symmetric about μ. Using the result in part (ii) of the solution to Exercise 18 and changing variable $z=y-\mu$, we obtain that

$$\phi_F(x) = 2\int_{-\infty}^{x-\mu} F_0(y)[2F_0(y)-1]dy - 2\int_{x-\mu}^{\infty}[1-F_0(y)][2F_0(y)-1]dy,$$

where F_0 is the double exponential distribution with location parameter 0 and scale parameter 1. Let X be a random variable having distribution F. Then $X-\mu$ has distribution F_0 and, therefore, the distribution of $\phi_F(X)$

does not depend on μ and we may solve the problem by assuming that $\mu = 0$. Note that

$$\phi'_F(x) = F(x)[4F(x) - 2] + [1 - F(x)][4F(x) - 2] = 4F(x) - 2.$$

Hence,

$$\phi_F(x) = \int_0^x [4F(y) - 2]dy + c,$$

where c is a constant. For the double exponential distribution with location parameter 0 and scale parameter 1,

$$F(y) = \begin{cases} \frac{1}{2}e^y & y < 0 \\ \frac{1}{2} + \frac{1}{2}(1 - e^{-y}) & y \geq 0 \end{cases}$$

and, hence,

$$\int_0^x [4F(y) - 2]dy = 2\int_0^x (1 - e^{-y})dy = 2(x - 1 + e^{-x})$$

when $x \geq 0$ and

$$\int_0^x [4F(y) - 2]dy = 2\int_0^x (e^y - 1)dy = 2(e^x - 1 - x)$$

when $x < 0$. Thus,

$$\phi_F(x) = 2(|x| - 1 + e^{-|x|}) + c.$$

From the property of influence function, $E[\phi_F(X)] = 0$. Hence,

$$c = -2E(|X| - 1 + e^{-|X|}) = -1,$$

since $|X|$ has the exponential distribution on $(0, \infty)$ with scale parameter 1. Then,

$$\begin{aligned} \sigma_F^2 &= E[\phi_F(X)]^2 \\ &= E(2|X| - 3 + 2e^{-|X|})^2 \\ &= E(4|X|^2 + 9 + 4e^{-2|X|} - 12|X| + 8|X|e^{-|X|} - 12e^{-|X|}) \\ &= 8 + 9 + \frac{4}{3} - 12 + 2 - 6 \\ &= \frac{7}{3}. \end{aligned}$$ ▮

Exercise 23 (#5.59). Let $T(G) = G^{-1}(p)$ be the pth quantile functional. Suppose that F has a positive derivative F' in a neighborhood of $\theta =$

$F^{-1}(p)$. Show that T is Gâteaux differentiable at F and obtain the influence function.

Solution. Let $H_t = F + t(G-F)$. Differentiating the identity $H_t(H_t^{-1}(p)) = p$ with respect to t at $t = 0$, we obtain that

$$G(F^{-1}(p)) - F(F^{-1}(p)) + F'(F^{-1}(p))H_t'(0) = 0.$$

Hence,

$$H_t'(0) = \frac{p - G(F^{-1}(p))}{F'(F^{-1}(p))}.$$

Let $G = \delta_x$, the degenerated distribution at x. Then the influence function is

$$\phi_F(x) = \frac{p - I_{[x,\infty)}(F^{-1}(p))}{F'(F^{-1}(p))}. \quad ∎$$

Exercise 24 (#5.51). Let F_n be the empirical distribution based on a random sample of size n from a distribution F on $\mathcal{R}$ having Lebesgue density f. Let $\varphi_n(t)$ be the Lebesgue density of the pth sample quantile $F_n^{-1}(p)$. Prove that

$$\varphi_n(t) = n\binom{n-1}{l_p-1}[F(t)]^{l_p-1}[1-F(t)]^{n-l_p}f(t),$$

where $l_p = np$ if np is an integer and $l_p = 1+$ the integer part of np if np is not an integer, by
(i) using the fact that $nF_n(t)$ has a binomial distribution;
(ii) using the fact that $F_n^{-1}(p) = c_{np}X_{(m_p)} + (1-c_{np})X_{(m_p+1)}$, where $X_{(j)}$ is the jth order statistic, m_p is the integer part of np, $c_{np} = 1$ if np is an integer, and $c_{np} = 0$ if np is not an integer.

Solution. (i) Since $nF_n(t)$ has the binomial distribution with size n and probability $F(t)$, for any $t \in \mathcal{R}$,

$$\begin{aligned} P\big(F_n^{-1}(p) \le t\big) &= P\big(F_n(t) \ge p\big) \\ &= \sum_{i=l_p}^{n} \binom{n}{i}[F(t)]^i[1-F(t)]^{n-i}. \end{aligned}$$

Differentiating term by term leads to

$$\begin{aligned} \varphi_n(t) = &\sum_{i=l_p}^{n} \binom{n}{i} i[F(t)]^{i-1}[1-F(t)]^{n-i}f(t) \\ &- \sum_{i=l_p}^{n} \binom{n}{i}(n-i)[F(t)]^i[1-F(t)]^{n-i-1}f(t) \end{aligned}$$

$$
\begin{aligned}
&= \binom{n}{l_p} l_p [F(t)]^{l_p-1}[1-F(t)]^{n-l_p} f(t) \\
&\quad + n \sum_{i=l_p+1}^{n} \binom{n-1}{i-1} [F(t)]^{i-1}[1-F(t)]^{n-i} f(t) \\
&\quad - \sum_{i=l_p}^{n-1} \binom{n-1}{i} [F(t)]^{i}[1-F(t)]^{n-i-1} f(t) \\
&= n\binom{n-1}{l_p-1} [F(t)]^{l_p-1}[1-F(t)]^{n-l_p} f(t).
\end{aligned}
$$

(ii) The Lebesgue density of the jth order statistic is

$$
n\binom{n-1}{j-1} [F(t)]^{j-1}[1-F(t)]^{n-j} f(t).
$$

Then, the result follows from the fact that

$$
F_n^{-1}(p) = \begin{cases} X_{(m_p)} & \text{if } np \text{ is an integer} \\ X_{(m_p+1)} & \text{if } np \text{ is not an integer} \end{cases}
$$

and $l_p = m_p$ if np is an integer and $l_p = m_p + 1$ if np is not an integer. ▮

Exercise 25 (#5.52). Let F_n be the empirical distribution based on a random sample from a distribution F on $\mathcal{R}$ with a finite mean. Show that the pth sample quantile $F_n^{-1}(p)$ has a finite jth moment for sufficiently large n, $j = 1, 2,...,$ where $p \in (0, 1)$.

Solution. From the previous exercise, the cumulative distribution function of $F_n^{-1}(p)$ is

$$
G_n(x) = \sum_{i=l_p}^{n} \binom{n}{i} [F(t)]^{i}[1-F(t)]^{n-i},
$$

where $l_p = np$ if np is an integer and $l_p = 1+$ the integer part of np if np is not an integer. When $n \to \infty$, $l_p \to \infty$ and $n - l_p \to \infty$. Hence, $j \leq l_p$ and $j \leq n - l_p + 1$ for sufficiently large n. Since $\int_0^\infty [1 - F(x)]dx < \infty$ (F has a finite mean),

$$
\lim_{x\to\infty} x^{j-1}[1-F(x)]^{n-i-1} \leq \lim_{x\to\infty} \max\{1-F(x), x^{j-1}[1-F(x)]^{j-1}\} = 0
$$

for $i = 1, ..., l_p - 1$. Thus,

$$
\int_0^\infty x^{j-1}[F(x)]^i[1-F(x)]^{n-i}dx \leq \int_0^\infty x^{j-1}[1-F(x)]^{n-i}dx < \infty
$$

for $i = 1, ..., l_p - 1$, which implies that

$$
\int_0^\infty x^{j-1}[1-G_n(x)]dx = \sum_{i=1}^{l_p-1} \binom{n}{i} \int_0^\infty x^{j-1}[F(x)]^i[1-F(x)]^{n-i}dx < \infty.
$$

Similarly, for $i = l_p, l_p + 1, ..., n$,

$$\int_{-\infty}^{0} |x|^{j-1}[F(x)]^i[1 - F(x)]^{n-i}dx \le \int_{-\infty}^{0} |x|^{j-1}[F(x)]^i dx < \infty$$

and, thus,

$$\int_{-\infty}^{0} |x|^{j-1}G_n(x)dx = \sum_{i=l_p}^{n} \binom{n}{i} \int_{-\infty}^{0} |x|^{j-1}[F(x)]^i[1 - F(x)]^{n-i}dx < \infty.$$

Therefore,

$$\begin{aligned}\int_{-\infty}^{\infty} |x|^j dG_n(x) &= \int_{0}^{\infty} x^j dG_n(x) + \int_{-\infty}^{0} |x|^j dG_n(x) \\ &= j\int_{0}^{\infty} x^{j-1}[1 - G_n(x)]dx + j\int_{-\infty}^{0} |x|^{j-1}G_n(x)dx \\ &< \infty. \ \blacksquare\end{aligned}$$

Exercise 26 (#5.54). Let F_n be the empirical distribution based on a random sample from a distribution F on $\mathcal{R}$ with Lebesgue density f that is positive and continuous at $\theta = F^{-1}(p)$, $p \in (0, 1)$. Using Scheffé's theorem (e.g., Proposition 1.18 in Shao, 2003), prove that

$$\sqrt{n}[F_n^{-1}(p) - \theta] \to_d N\left(0, \frac{p(1-p)}{[f(\theta)]^2}\right).$$

Solution. From Exercise 24, the Lebesgue density of $\sqrt{n}[F_n^{-1}(p) - \theta]$ is

$$\sqrt{n}\binom{n-1}{l_p-1}\left[F\left(\theta + \frac{y}{\sqrt{n}}\right)\right]^{l_p-1}\left[1 - F\left(\theta + \frac{y}{\sqrt{n}}\right)\right]^{n-l_p} f\left(\theta + \frac{y}{\sqrt{n}}\right),$$

where $l_p = np$ if np is an integer and $l_p = 1+$ the integer part of np if np is not an integer. Using Stirling's formula, we obtain that

$$\lim_n \sqrt{n}\binom{n-1}{l_p-1}p^{l_p-1}(1-p)^{n-l_p} = \frac{1}{\sqrt{2\pi p(1-p)}}.$$

From Taylor's expansion and the fact that $l_p/(np) \to 1$,

$$\left[\frac{F\left(\theta + \frac{y}{\sqrt{n}}\right)}{p}\right]^{l_p-1}\left[\frac{1 - F\left(\theta + \frac{y}{\sqrt{n}}\right)}{1-p}\right]^{n-l_p}$$

$$
\begin{aligned}
&= \exp\left\{ (l_p - 1)\log \frac{F\left(\theta + \frac{y}{\sqrt{n}}\right)}{p} + (n - l_p)\log \frac{1 - F\left(\theta + \frac{y}{\sqrt{n}}\right)}{1-p} \right\} \\
&= \exp\left\{ np\log \frac{F\left(\theta + \frac{y}{\sqrt{n}}\right)}{p} + (n - np)\log \frac{1 - F\left(\theta + \frac{y}{\sqrt{n}}\right)}{1-p} + o(1) \right\} \\
&= \exp\left\{ np\log\left(1 + \frac{f(\theta)y}{p\sqrt{n}}\right) + (n-np)\log\left(1 - \frac{f(\theta)y}{(1-p)\sqrt{n}}\right) + o(1) \right\} \\
&= \exp\left\{ \sqrt{n}f(\theta)y - \frac{[f(\theta)]^2y^2}{2p} + o(1) - \sqrt{n}f(\theta)y - \frac{[f(\theta)]^2y^2}{2(1-p)} + o(1) \right\} \\
&= \exp\left\{ -\frac{[f(\theta)]^2y^2}{2p(1-p)} \right\} + o(1)
\end{aligned}
$$

Also, $\lim_n f\left(\theta + \frac{y}{\sqrt{n}}\right) = f(\theta)$. Hence, the density of $\sqrt{n}[F_n^{-1}(p) - \theta]$ converges to

$$
\frac{f(\theta)}{\sqrt{2\pi p(1-p)}} \exp\left\{ -\frac{[f(\theta)]^2y^2}{2p(1-p)} \right\}
$$

for any y, which is the Lebesgue density of $N\left(0, p(1-p)/[f(\theta)]^2\right)$. Hence, the result follows from Scheffé's theorem. ∎

Exercise 27 (#5.55). Let $\{k_n\}$ be a sequence of integers satisfying $k_n/n = p + o(n^{-1/2})$ with $p \in (0,1)$, and let $(X_1, ..., X_n)$ be a random sample from a distribution F on $\mathcal{R}$ with $F'(\theta_p) > 0$, where $\theta_p = F^{-1}(p)$. Let $X_{(j)}$ be the jth order statistic. Show that

$$
\sqrt{n}(X_{(k_n)} - \theta_p) \to_d N(0, p(1-p)/[F'(\theta_p)]^2).
$$

Solution. Let $p_n = k_n/n = p + o(n^{-1/2})$ and F_n be the empirical distribution. Then $X_{(k_n)} = F_n^{-1}(p_n)$ for any n. Let $t \in \mathcal{R}$, $\sigma = \sqrt{p(1-p)}/F'(\theta_p)$, $p_{nt} = F(\theta_p + t\sigma n^{-1/2})$, and $c_{nt} = \sqrt{n}(p_{nt} - p_n)/\sqrt{p_{nt}(1-p_{nt})}$. Define $Z_{nt} = [B_n(p_{nt}) - np_{nt}]/\sqrt{np_{nt}(1-p_{nt})}$, where $B_n(q)$ denotes a random variable having the binomial distribution with size n and probability q. Then

$$
\begin{aligned}
P\big(\sqrt{n}(X_{(k_n)} - \theta_p) \le t\sigma\big) &= P\big(F_n^{-1}(p_n) \le \theta_p + t\sigma n^{-1/2}\big) \\
&= P\big(p_n \le F_n(\theta_p + t\sigma n^{-1/2})\big) \\
&= P\big(Z_{nt} \ge -c_{nt}\big) \\
&= \Phi(c_{nt}) + o(1)
\end{aligned}
$$

by the central limit theorem and Pólya's theorem, where Φ is the cumulative distribution function of $N(0,1)$. The result follows if we can show that

$\lim_n c_{nt} = t$. By Taylor's expansion,

$$p_{nt} = F(\theta_p + t\sigma n^{-1/2}) = p + t\sigma n^{-1/2} F'(\theta_p) + o(n^{-1/2}).$$

Then, as $n \to \infty$, $p_{nt} \to p$ and $\sqrt{n}(p_{nt} - p_n) = t\sigma F'(\theta_p) + o(1)$, since $p_n - p = o(n^{-1/2})$. Hence,

$$c_{nt} = \frac{\sqrt{n}(p_{nt} - p_n)}{\sqrt{p_{nt}(1 - p_{nt})}} \to \frac{t\sigma F'(\theta_p)}{\sqrt{p(1-p)}} = t. \quad \blacksquare$$

Exercise 28 (#5.112). Let G, G_1, G_2,..., be cumulative distributions on $\mathcal{R}$. Suppose that $\lim_n \sup_x |G_n(x) - G(x)| = 0$, G is continuous, and G^{-1} is continuous at $p \in (0, 1)$.
(i) Show that $\lim_n G_n^{-1}(p) = G^{-1}(p)$.
(ii) Show that the result in (i) holds for any $p \in (0, 1)$ if $G'(x)$ exists and is positive for any $x \in \mathcal{R}$.
Solution. (i) Let $\epsilon > 0$. Since $\lim_n \sup_x |G_n(x) - G(x)| = 0$ and G is continuous, $\lim_n G_n(G^{-1}(p - \epsilon)) = G(G^{-1}(p - \epsilon)) = p - \epsilon < p$. Hence, for sufficiently large n, $G_n(G^{-1}(p - \epsilon)) \le p$, i.e., $G^{-1}(p - \epsilon) \le G_n^{-1}(p)$. Thus,

$$G^{-1}(p - \epsilon) \le \liminf_n G_n^{-1}(p).$$

Similarly,

$$G^{-1}(p + \epsilon) \ge \limsup_n G_n^{-1}(p).$$

Letting $\epsilon \to 0$, by the continuity of G^{-1} at p, we conclude that $\lim_n G_n^{-1}(p) = G^{-1}(p)$.
(ii) If $G'(x)$ exists and is positive for any $x \in \mathcal{R}$, then G^{-1} is continuous on $(0, 1)$. The result follows from the result in (i). $\blacksquare$

Exercise 29 (#5.47). Calculate the asymptotic relative efficiency of the Hodges-Lehmann estimator with respect to the sample mean based on a random sample from F when
(i) F is the cumulative distribution of $N(\mu, \sigma^2)$;
(ii) F is the cumulative distribution of the logistic distribution with location parameter μ and scale parameter σ;
(iii) F is the cumulative distribution of the double exponential distribution with location parameter μ and scale parameter σ;
(iv) $F(x) = F_0(x - \mu)$, where $F_0(x)$ is the cumulative distribution of the t-distribution t_ν with $\nu \ge 3$.
Solution. In any case, as estimators of μ, the sample mean is asymptotically normal with asymptotic mean squared error $\text{Var}(X)/n$, where X denotes a random variable with distribution F, and the Hodges-Lehmann estimator is asymptotically normal with asymptotic mean squared error

$(12\gamma^2)^{-1}$, where $\gamma = \int [F'(x)]^2 dx$ (e.g., Example 5.8 in Shao, 2003). Hence, the asymptotic relative efficiency to be calculated is $12\gamma^2\mathrm{Var}(X)$.
(i) In this case, $\mathrm{Var}(X) = \sigma^2$ and

$$\gamma = \frac{1}{2\pi\sigma^2}\int_{-\infty}^{\infty} e^{-(x-\mu)^2/\sigma^2}dx = \frac{1}{2\pi\sigma}\int_{-\infty}^{\infty} e^{-x^2}dx = \frac{1}{2\sqrt{\pi}\sigma}.$$

Hence, the asymptotic relative efficiency is $12\gamma^2\mathrm{Var}(X) = 3/\pi$.
(ii) Note that $\mathrm{Var}(X) = \sigma^2\pi^2/3$ and

$$\gamma = \frac{1}{\sigma^2}\int_{-\infty}^{\infty} \frac{e^{-2(x-\mu)/\sigma}}{[1+e^{-(x-\mu)/\sigma}]^4}dx = \frac{1}{\sigma}\int_{-\infty}^{\infty}\frac{e^{2x}}{(1+e^x)^4}dx = \frac{1}{6\sigma}$$

(Exercise 66 in Chapter 4). Hence, the asymptotic relative efficiency is $12\gamma^2\mathrm{Var}(X) = \pi^2/9$.
(iii) In this case, $\mathrm{Var}(X) = 2\sigma^2$ and

$$\gamma = \frac{1}{4\sigma^2}\int_{-\infty}^{\infty} e^{-2|x-\mu|/\sigma}dx = \frac{1}{4\sigma}\int_{-\infty}^{\infty} e^{-2|x|}dx = \frac{1}{4\sigma}.$$

Hence, the asymptotic relative efficiency is $12\gamma^2\mathrm{Var}(X) = 3/2$.
(iv) Note that $\mathrm{Var}(X) = \nu/(\nu-2)$ and

$$\gamma = \frac{\left[\Gamma\left(\frac{\nu+1}{2}\right)\right]^2}{\nu\pi\left[\Gamma\left(\frac{\nu}{2}\right)\right]^2}\int_{-\infty}^{\infty}\frac{dx}{\left(1+\frac{x^2}{\nu}\right)^{\nu+1}} = \frac{\sqrt{\nu\pi}\left[\Gamma\left(\frac{\nu+1}{2}\right)\right]^2\Gamma\left(\frac{2\nu+1}{2}\right)}{\left[\Gamma\left(\frac{\nu}{2}\right)\right]^2\Gamma(\nu+1)}.$$

Hence, the asymptotic relative efficiency is

$$12\gamma^2\mathrm{Var}(X) = \frac{12\nu^2\pi\left[\Gamma\left(\frac{\nu+1}{2}\right)\right]^4\left[\Gamma\left(\frac{2\nu+1}{2}\right)\right]^2}{(\nu-2)\left[\Gamma\left(\frac{\nu}{2}\right)\right]^4[\Gamma(\nu+1)]^2}. \blacksquare$$

Exercise 30 (#5.61, #5.62, #5.63). Consider a random sample from a distribution F on $\mathcal{R}$. In each of the following cases, obtain the asymptotic relative efficiency of the sample median with respect to the sample mean.
(i) F is the cumulative distribution of the uniform distribution on the interval $(\theta-\frac{1}{2}, \theta+\frac{1}{2})$, $\theta \in \mathcal{R}$.
(ii) $F(x) = F_0(x-\theta)$ and F_0 is the cumulative distribution function with Lebesgue density $(1+x^2)^{-1}I_{(-c,c)}(x)/\int_{-c}^{c}(1+x^2)^{-1}dt$.
(iii) $F(x) = (1-\epsilon)\Phi\left(\frac{x-\mu}{\sigma}\right)+\epsilon D\left(\frac{x-\mu}{\sigma}\right)$, where $\epsilon \in (0,1)$ is a known constant, Φ is the cumulative distribution function of the standard normal distribution, D is the cumulative distribution function of the double exponential distribution with location parameter 0 and scale parameter 1, and $\mu \in \mathcal{R}$ and $\sigma > 0$ are unknown parameters.
Solution. In each case, the asymptotic relative efficiency of the sample

median with respect to the sample mean is $4[F'(\theta)]^2\text{Var}(X_1)$.
(i) Let θ be the mean of F. In this case, $\text{Var}(X_1) = 1/12$ and $F'(\theta) = 1$. Hence, the asymptotic relative efficiency of the sample median with respect to the sample mean is $1/3$.
(ii) The Lebesgue density of F_0 is

$$f(x) = \frac{I_{(-c,c)}(x)}{2\arctan(c)(1+x^2)}.$$

Hence, $F'(\theta) = [2\arctan(c)]^{-1}$. Note that

$$\begin{aligned}
\text{Var}(X_1) &= \int_{-c}^{c} \frac{x^2 dx}{2\arctan(c)(1+x^2)} \\
&= \int_{-c}^{c} \frac{dx}{2\arctan(c)} - \int_{-c}^{c} \frac{dx}{2\arctan(c)(1+x^2)} \\
&= \frac{c}{\arctan(c)} - 1.
\end{aligned}$$

Therefore, the asymptotic relative efficiency of the sample median with respect to the sample mean is $[c - \arctan(c)]/[\arctan(c)]^3$.
(iii) Note that

$$\text{Var}(X_1) = (1-\epsilon)\sigma^2 + 2\epsilon\sigma^2 = (1+\epsilon)\sigma^2$$

and

$$F'(\mu) = \frac{1-\epsilon}{\sqrt{2\pi}\sigma} + \frac{\epsilon}{2\sigma}.$$

Hence, the asymptotic relative efficiency of the sample median with respect to the sample mean is $4\left(\frac{1-\epsilon}{\sqrt{2\pi}} + \frac{\epsilon}{2}\right)/(1+\epsilon)$. ∎

Exercise 31 (#5.64). Let $(X_1, ..., X_n)$ be a random sample from a distribution on $\mathcal{R}$ with Lebesgue density $2^{-1}(1-\theta^2)e^{\theta x - |x|}$, where $\theta \in (-1, 1)$ is unknown.
(i) Show that the median of the distribution of X_1 is given by $m(\theta) = (1-\theta)^{-1}\log(1+\theta)$ when $\theta \geq 0$ and $m(\theta) = -m(-\theta)$ when $\theta < 0$.
(ii) Show that the mean of the distribution of X_1 is $\mu(\theta) = 2\theta/(1-\theta^2)$.
(iii) Show that the inverse functions of $m(\theta)$ and $\mu(\theta)$ exist. Obtain the asymptotic relative efficiency of $m^{-1}(\hat{m})$ with respect to $\mu^{-1}(\bar{X})$, where $\hat{m}$ is the sample median and $\bar{X}$ is the sample mean.
(iv) Is $\mu^{-1}(\bar{X})$ asymptotically efficient in estimating θ?
Solution. (i) The cumulative distribution function of X_1 is

$$F_\theta(x) = \begin{cases} \frac{1-\theta}{2} e^{(1+\theta)x} & x \leq 0 \\ 1 - \frac{1+\theta}{2} e^{-(1-\theta)x} & x > 0. \end{cases}$$

If $\theta > 0$, $F_\theta(0) = \frac{1-\theta}{2} < \frac{1}{2}$. Hence, the median is the solution to

$$\frac{1}{2} = 1 - \frac{1+\theta}{2} e^{-(1-\theta)x},$$

i.e., $(1-\theta)^{-1}\log(1+\theta) = m(\theta)$. If $\theta < 0$, then the median is the solution to

$$\frac{1}{2} = \frac{1-\theta}{2} e^{(1+\theta)x},$$

i.e., $-(1+\theta)^{-1}\log(1-\theta) = -m(-\theta)$. If $\theta = 0$, the median is clearly $0 = m(0)$.

(ii) The mean of X_1 is

$$\begin{aligned}\frac{1-\theta^2}{2}\int_{-\infty}^{\infty} xe^{\theta x - |x|}dx &= \frac{1-\theta^2}{2}\left[\int_{-\infty}^{0} xe^{(1+\theta)x}dx + \int_{0}^{\infty} xe^{-(1-\theta)x}dx\right] \\ &= \frac{1-\theta^2}{2}\left[-\frac{1}{(1+\theta)^2} + \frac{1}{(1-\theta)^2}\right] \\ &= \frac{2\theta}{1-\theta^2}.\end{aligned}$$

(iii) Since

$$\mu'(\theta) = \frac{2}{1-\theta^2} + \frac{4\theta^2}{(1-\theta^2)^2} > 0,$$

$\mu(\theta)$ is increasing in θ and, thus, the inverse function μ^{-1} exists. For $\theta \geq 0$, $m(\theta)$ is the product of $\log(1+\theta)$ and $(1-\theta)^{-1}$, both of which are increasing in θ. Hence, $m(\theta)$ is increasing in θ for $\theta \in [0,1)$. Since $m(\theta) = -m(-\theta)$ for $\theta < 0$, the median function $m(\theta)$ is increasing in θ. Hence, the inverse function m^{-1} exists.

When $\theta \geq 0$, the density of X_1 evaluated at the median $m(\theta)$ is equal to

$$\frac{1-\theta^2}{2} e^{\theta m(\theta) - |m(\theta)|} = \frac{1-\theta^2}{2} e^{-(1-\theta)m(\theta)} = \frac{1-\theta}{2}.$$

When $\theta < 0$, the density of X_1 evaluated at the median $-m(-\theta)$ is equal to

$$\frac{1-\theta^2}{2} e^{-\theta m(-\theta) - |-m(-\theta)|} = \frac{1-\theta^2}{2} e^{-(1+\theta)m(-\theta)} = \frac{1+\theta}{2}.$$

In any case, the density of X_1 evaluated at the median is $(1-|\theta|)/2$. By the asymptotic theory for sample median (e.g., Theorem 5.10 in Shao, 2003),

$$\sqrt{n}[\hat{m} - m(\theta)] \to_d N\left(0, \frac{1}{(1-|\theta|)^2}\right).$$

When $\theta \geq 0$,

$$m'(\theta) = \frac{1}{1-\theta^2} + \frac{\log(1+\theta)}{(1-\theta)^2}.$$

Hence, for $\theta \in (-1, 1)$,

$$m'(\theta) = \frac{1}{1-\theta^2} + \frac{\log(1+|\theta|)}{(1-|\theta|)^2} = \frac{1-|\theta| + (1+|\theta|)\log(1+|\theta|)}{(1+|\theta|)(1-|\theta|)^2}.$$

By the δ-method,

$$\sqrt{n}[m^{-1}(\hat{m}) - \theta] \to_d N\left(0, \frac{(1-\theta^2)^2}{[1-|\theta| + (1+|\theta|)\log(1+|\theta|)]^2}\right).$$

For $\mu^{-1}(\bar{X})$, it is shown in the next part of the solution that

$$\sqrt{n}[\mu^{-1}(\bar{X}) - \theta] \to_d N\left(0, \frac{(1-\theta^2)^2}{2(1+\theta^2)}\right).$$

Hence, the asymptotic relative efficiency of $m^{-1}(\hat{m})$ with respect to $\mu^{-1}(\bar{X})$ is

$$\frac{[1-|\theta| + (1+|\theta|)\log(1+|\theta|)]^2}{2(1+\theta^2)}.$$

(iv) The likelihood function is

$$\ell(\theta) = 2^{-n}(1-\theta^2)^n \exp\left\{n\theta\bar{X} - \sum_{i=1}^n |X_i|\right\}.$$

Then,

$$\frac{\partial \log \ell(\theta)}{\partial \theta} = \frac{n}{1+\theta} - \frac{n}{1-\theta} + n\bar{X}$$

and

$$\frac{\partial^2 \log \ell(\theta)}{\partial \theta^2} = -\frac{n}{(1+\theta)^2} - \frac{n}{(1-\theta)^2} < 0.$$

Hence, the solution to the likelihood equation $\frac{n}{1+\theta} - \frac{n}{1-\theta} + n\bar{X} = 0$ is the MLE of θ. Since $\frac{n}{1+\theta} - \frac{n}{1-\theta} = n\mu(\theta)$, we conclude that $\mu^{-1}(\bar{X})$ is the MLE of θ. Since the distribution of X_1 is from an exponential family, $\mu^{-1}(\bar{X})$ is asymptotically efficient and $\sqrt{n}[\mu^{-1}(\bar{X}) - \theta] \to_d N(0, [I_1(\theta)]^{-1})$, where

$$I_1(\theta) = \frac{1}{(1+\theta)^2} + \frac{1}{(1-\theta)^2} = \frac{2(1+\theta^2)}{(1-\theta^2)^2}. \quad \blacksquare$$

Exercise 32 (#5.70, #5.71). Obtain the asymptotic relative efficiency of the trimmed sample mean $\bar{X}_\alpha$ (Exercise 21) with respect to
(i) the sample mean, based on a random sample of size n from the double exponential distribution with location parameter $\theta \in \mathcal{R}$ and scale parameter 1;

(ii) the sample median, based on a random sample of size n from the Cauchy distribution with location parameter $\theta \in \mathcal{R}$ and scale parameter 1.

Solution. (i) Let F_0 be the double exponential distribution with location parameter 0 and scale parameter 1. The variance of F_0 is 2. Hence, the asymptotic relative efficiency of the trimmed sample mean $\bar{X}_\alpha$ with respect to the sample mean is $2/\sigma_\alpha^2$, where σ_α^2 is given in Exercise 21(ii). Note that $F_0^{-1}(1-\alpha) = -\log(2\alpha)$. Hence,

$$\begin{aligned}\sigma_\alpha^2 &= \frac{1}{(1-2\alpha)^2}\int_0^{-\log(2\alpha)} x^2 e^{-x}dx + \frac{2\alpha[\log(2\alpha)]^2}{(1-2\alpha)^2} \\ &= \frac{2\alpha[\log(2\alpha)-1]+1}{(1-2\alpha)^2}.\end{aligned}$$

Thus, the asymptotic relative efficiency is $2(1-2\alpha)^2/\{2\alpha[\log(2\alpha)-1]+1\}$.

(ii) Let F_0 be the Cauchy distribution with location parameter 0 and scale parameter 1. Note that $F_0'(0) = 1/\pi$ and $F_0^{-1}(1-\alpha) = \tan(\pi - \pi\alpha)$. Hence, the asymptotic relative efficiency of the trimmed sample mean $\bar{X}_\alpha$ with respect to the sample median is $\pi^2/(4\sigma_\alpha^2)$, where

$$\begin{aligned}\sigma_\alpha^2 &= \frac{2}{(1-2\alpha)^2}\int_0^{\tan(\pi-\pi\alpha)} \frac{x^2dx}{\pi(1+x^2)} + \frac{2\alpha[\tan(\pi-\pi\alpha)]^2}{(1-2\alpha)^2} \\ &= \frac{2}{(1-2\alpha)^2}\left[\frac{\tan(\pi-\pi\alpha)}{\pi} - \int_0^{\tan(\pi-\pi\alpha)}\frac{dx}{\pi(1+x^2)}\right] \\ &\quad + \frac{2\alpha[\tan(\pi-\pi\alpha)]^2}{(1-2\alpha)^2} \\ &= \frac{2}{(1-2\alpha)^2}\left[\frac{\tan(\pi-\pi\alpha)}{\pi} - \frac{1-2\alpha}{2}\right] + \frac{2\alpha[\tan(\pi-\pi\alpha)]^2}{(1-2\alpha)^2} \\ &= \frac{2\tan(\pi-\pi\alpha)}{\pi(1-2\alpha)^2} - \frac{1}{1-2\alpha} + \frac{2\alpha[\tan(\pi-\pi\alpha)]^2}{(1-2\alpha)^2}. \quad \blacksquare\end{aligned}$$

Exercise 33 (#5.85). Let $(X_1, ..., X_n)$ be a random sample from a distribution F on $\mathcal{R}$ that is symmetric about $\theta \in \mathcal{R}$. Huber's estimator of θ is defined as a solution of $\sum_{i=1}^n \psi(X_i, t) = 0$, where

$$\psi(x,t) = \begin{cases} C & t - x > C \\ t - x & |x - t| \le C \\ -C & t - x < -C \end{cases}$$

and $C > 0$ is a constant. Assume that F is continuous at $\theta - C$ and $\theta + C$.

(i) Show that the function

$$\Psi(\gamma) = \int_{\gamma-C}^{\gamma+C} (\gamma - x)dF(x) + CF(\gamma - C) - C[1 - F(\gamma + C)]$$

is differentiable at θ and $\Psi(\theta) = 0$.
(ii) Show that the asymptotic relative efficiency of Huber's estimator with respect to the sample mean is $\mathrm{Var}(X_1)/\sigma_F^2$, where

$$\sigma_F^2 = \frac{\int_{\theta-C}^{\theta+C}(\theta - x)^2 dF(x) + C^2 F(\theta - C) + C^2[1 - F(\theta + C)]}{[F(\theta + C) - F(\theta - C)]^2}.$$

Solution. (i) Since F is symmetric about θ, $F(\theta - C) = 1 - F(\theta + C)$, $dF(\theta - y) = dF(\theta + y)$, and

$$\int_{\theta-C}^{\theta+C}(\theta - x)dF(x) = \int_{-C}^{C} ydF(\theta + y) = -\int_{-C}^{C} ydF(\theta - y),$$

where the first equality follows by considering $x = \theta + y$ and the second equality follows by considering $x = \theta - y$. Hence, $\int_{\theta-C}^{\theta+C}(\theta - x)dF(x) = 0$ and, thus, $\Psi(\theta) = 0$.

From integration by parts,

$$\int_{\gamma-C}^{\gamma+C}(\gamma - x)dF(x) = -C[F(\gamma + C) + F(\gamma - C)] + \int_{\gamma-C}^{\gamma+C} F(x)dx.$$

Hence,

$$\Psi(\gamma) = \int_{\gamma-C}^{\gamma+C} F(x)dx - C,$$

which is differentiable at θ and $\Psi'(\theta) = F(\theta + C) - F(\theta - C)$.
(ii) The function

$$\int [\psi(x, \gamma)]^2 dF(x) = \int_{\gamma-C}^{\gamma+C}(\gamma - x)^2 dF(x) + C^2 F(\gamma - C) + C^2[1 - F(\gamma + C)]$$

is continuous at θ. Hence, by the result in (i) and Theorem 5.13(i) in Shao (2003), Huber's estimator $\hat\theta$ satisfies $\sqrt{n}(\hat\theta - \theta) \to_d N(0, \sigma_F^2)$. This proves the result. ∎

Exercise 34 (#5.86). For Huber's estimator $\hat\theta$ in the previous exercise, obtain a formula $e(F)$ for the asymptotic relative efficiency of $\hat\theta$ with respect to the sample mean, when

$$F(x) = (1 - \epsilon)\Phi\left(\tfrac{x-\theta}{\sigma}\right) + \epsilon\Phi\left(\tfrac{x-\theta}{\tau\sigma}\right),$$

where Φ is the cumulative distribution function of $N(0, 1)$, $\sigma > 0$, $\tau > 0$, and $0 < \epsilon < 1$. Show that $\lim_{\tau\to\infty} e(F) = \infty$. Find the value of $e(F)$ when $\epsilon = 0$, $\sigma = 1$, and $C = 1.5$.

Solution. The variance of F is $(1-\epsilon)\sigma^2 + \epsilon\tau^2\sigma^2$. Let σ_F^2 be given in the previous exercise. Then

$$e(F) = \frac{(1-\epsilon)\sigma^2 + \epsilon\tau^2\sigma^2}{\sigma_F^2},$$

where

$$\sigma_F^2 = \frac{\int_{-C}^{C} y^2 d\big[(1-\epsilon)\Phi\left(\frac{y}{\sigma}\right) + \epsilon\Phi\left(\frac{y}{\tau\sigma}\right)\big] + 2C^2\big[(1-\epsilon)\Phi\left(-\frac{C}{\sigma}\right) + \epsilon\Phi\left(-\frac{C}{\tau\sigma}\right)\big]}{2\left[(1-\epsilon)\Phi\left(\frac{C}{\sigma}\right) + \epsilon\Phi\left(\frac{C}{\tau\sigma}\right)\right] - 1}.$$

Since σ_F^2 is a bounded function of τ, $\lim_{\tau\to\infty} e(F) = \infty$. When $\epsilon = 0$, $\sigma = 1$, and $C = 1.5$,

$$\begin{aligned} e(F) &= \frac{\frac{1}{\sqrt{2\pi}}\int_{-C}^{C} y^2 e^{-y^2/2}dy + 2C^2\Phi(-C)}{2\Phi(C) - 1} \\ &= \frac{-\sqrt{\frac{2}{\pi}}Ce^{-C^2/2} + \Phi(C) - \Phi(-C) + 2C^2\Phi(-C)}{2\Phi(C) - 1} \\ &= \frac{-0.3886 + 0.8664 + 0.3006}{0.8664} \\ &= 0.8984. \ \blacksquare \end{aligned}$$

Exercise 35 (#5.99). Consider the L-functional T defined in Exercise 17. Let F_n be the empirical distribution based on a random sample of size n from a distribution F, $\sigma^2 = \int [\phi_F(x)]^2 dF(x)$, and $\sigma_{F_n}^2 = \int [\phi_{F_n}(x)]^2 dF_n(x)$, where ϕ_G denotes the influence function of T at distribution G. Show that $\lim_n \sigma_{F_n}^2 = \sigma_F^2$ a.s., under one of the following two conditions:
(a) J is bounded, $J(t) = 0$ when $t \in [0,\alpha] \cup [\beta, 1]$ for some constants $\alpha < \beta$, and the set $D = \{x : J \text{ is discontinuous at } F(x)\}$ has Lebesgue measure 0.
(b) J is continuous on $[0,1]$ and $\int x^2 dF(x) < \infty$.
Solution. (i) Assume condition (a). Let $C = \sup_x |J(x)|$. Note that $\lim_n \sup_y |F_n(y) - F(y)| = 0$ a.s. Hence, there are constants $a < b$ such that

$$\phi_F(x) = -\int_a^b (\delta_x - F)(y) J(F(y)) dy$$

and

$$\phi_{F_n}(x) = -\int_a^b (\delta_x - F_n)(y) J(F_n(y)) dy \quad \text{a.s.}$$

The condition that D has Lebesgue measure 0 ensures that

$$\lim_n \int_a^b |J(F(y)) - J(F_n(y))| dy = 0 \quad \text{a.s.}$$

Hence,

$$\begin{aligned}|\phi_{F_n}(x)-\phi_F(x)| &= \left|\int_a^b (F_n-F)(y)J(F_n(y))dy \right.\\ &\quad \left. + \int_a^b (\delta_x - F)(y)[J(F(y))-J(F_n(y))]dy\right| \\ &\leq C(b-a)\sup_y |F_n(y)-F(y)| \\ &\quad + \int_a^b |J(F(y))-J(F_n(y))|dy \\ &\to 0 \ \text{ a.s.}\end{aligned}$$

Since $\sup_x |\phi_{F_n}(x)| \leq C(b-a)$, by the dominated convergence theorem,

$$\lim_n \int [\phi_{F_n}(x)]^2 dF(x) = \int [\phi_F(x)]^2 dF(x) \ \text{ a.s.}$$

By the extended dominated convergence theorem (e.g., Proposition 18 in Royden, 1968, p. 232),

$$\lim_n \int [\phi_{F_n}(x)]^2 d(F_n - F)(x) = 0 \ \text{ a.s.}$$

This proves the result.
(ii) Assume condition (b). Let $C = \sup_x |J(x)|$. From the previous proof, we still have

$$\begin{aligned}\phi_{F_n}(x)-\phi_F(x) &= \int_{-\infty}^{\infty} (F_n-F)(y)J(F_n(y))dy \\ &\quad + \int_{-\infty}^{\infty} (\delta_x - F)(y)[J(F(y))-J(F_n(y))]dy.\end{aligned}$$

The first integral in the previous expression is bounded in absolute value by $C\int_{-\infty}^{\infty} |F_n - F|(y)dy$, which converges to 0 a.s. by Theorem 5.2(i) in Shao (2003). The second integral in the previous expression is bounded in absolute value by

$$\sup_y |J(F_n(y)) - J(F(y))| \left\{\int_{-\infty}^{x} F(y)dy + \int_x^{\infty} [1-F(y)]dy\right\},$$

which converges to 0 a.s. by the continuity of J and $\int x^2 dF(x) < \infty$. Hence, $\lim_n \phi_{F_n}(x) = \phi_F(x)$ a.s. for any x. The rest of the proof is the same as that in part (i) of the solution, since $[\phi_{F_n}(x)] \leq C^2[|x| + \int |x|dF(x)]^2$ (see the solution of Exercise 17) and $\int x^2 dF(x) < \infty$. ∎

Exercise 36 (#5.100). Let $(X_1, ..., X_n)$ be a random sample from a distribution F on $\mathcal{R}$ and let U_n be a U-statistic (see Exercise 25 in Chapter

3) with kernel $h(x_1, ..., x_m)$ satisfying $E[h(X_1, ..., X_m)]^2 < \infty$, where $m < n$. Assume that $\zeta_1 = \mathrm{Var}(h_1(X_1)) > 0$, where $h_1(x) = E[h(x, X_2, ..., X_m)]$. Derive a consistent variance estimator for U_n.

Solution. From Exercise 25 in Chapter 3, it suffices to derive a consistent estimator of ζ_1. Since $\zeta_1 = E[h_1(X_1)]^2 - \{E[h_1(X_1)]\}^2 = E[h_1(X_1)]^2 - \{E(U_n)\}^2$ and U_n is a consistent estimator of $E(U_n)$, it suffices to derive a consistent estimator of $\rho = E[h_1(X_1)]^2$. Note that

$$\begin{aligned}\rho &= \int \left[\int \cdots \int h(x, y_1, ..., y_{m-1}) dF(y_1) \cdots dF(y_{m-1}) \right]^2 dF(x) \\ &= \int \cdots \int h(x, y_1, ..., y_{m-1}) h(x, y_m, ..., y_{2m+1}) dF(y_1) \cdots dF(y_{2m+1}) dF(x).\end{aligned}$$

Hence, a consistent estimator of ρ is the U-statistic with kernel

$$h(x, y_1, ..., y_{m-1}) h(x, y_m, ..., y_{2m+1}). \blacksquare$$

Exercise 37 (#5.101). For Huber's estimator defined in Exercise 33, derive a consistent estimator of its asymptotic variance σ_F^2.

Solution. Let $\hat{\theta}$ be Huber's estimator of θ, F_n be the empirical distribution based on $X_1, ..., X_n$, and

$$\sigma_{F_n}^2 = \frac{\int_{\hat{\theta}-C}^{\hat{\theta}+C} (\hat{\theta} - x)^2 dF_n(x) + C^2 F_n(\hat{\theta} - C) + C^2 [1 - F_n(\hat{\theta} + C)]}{[F_n(\hat{\theta} + C) - F_n(\hat{\theta} - C)]^2}.$$

Using integration by parts, we obtain that

$$\sigma_F^2 = \frac{2 \int_{\theta-C}^{\theta+C} (\theta - x) F(x) dx + C^2}{[F(\theta + C) - F(\theta - C)]^2}$$

and

$$\sigma_{F_n}^2 = \frac{2 \int_{\hat{\theta}-C}^{\hat{\theta}+C} (\hat{\theta} - x) F_n(x) dx + C^2}{[F_n(\hat{\theta} + C) - F_n(\hat{\theta} - C)]^2}.$$

To show that $\sigma_{F_n}^2$ is a consistent estimator of σ_F^2, it suffices to show that

$$\lim_n F_n(\hat{\theta} + C) \to_p F(\theta + C)$$

and

$$\lim_n \int_{\hat{\theta}-C}^{\hat{\theta}+C} (\hat{\theta} - x) F_n(x) dx \to_p \int_{\theta-C}^{\theta+C} (\theta - x) F(x) dx.$$

The first required result follows from

$$|F_n(\hat{\theta} + C) - F(\theta + C)| \leq |F(\hat{\theta} + C) - F(\theta + C)| + \sup_x |F_n(x) - F(x)|,$$

the fact that $\lim_n \sup_x |F_n(x) - F(x)| = 0$ a.s., the consistency of $\hat{\theta}$, and the assumption that F is continuous at $F(\theta + C)$. Let

$$g(\gamma) = \int_{\gamma - C}^{\gamma + C} (\gamma - x) F(x) dx.$$

Then $g(\gamma)$ is continuous at θ and, thus, $g(\hat{\theta}) \to_p g(\theta)$. Note that

$$\int_{\hat{\theta}-C}^{\hat{\theta}+C} (\hat{\theta} - x) F_n(x) dx = \int_{\hat{\theta}-C}^{\hat{\theta}+C} (\hat{\theta} - x)[F_n(x) - F(x)] dx + g(\hat{\theta})$$

and

$$\left| \int_{\hat{\theta}-C}^{\hat{\theta}+C} (\hat{\theta} - x)[F_n(x) - F(x)] dx \right| \leq 2C^2 \sup_x |F_n(x) - F(x)|.$$

Hence, the second required result follows. ∎

Exercise 38 (#5.104). Let $X_1, ..., X_n$ be random variables. For any estimator $\hat{\theta}$, its jackknife variance estimator is defined as

$$v_{_J} = \frac{n-1}{n} \sum_{i=1}^{n} \left(\hat{\theta}_{-i} - \frac{1}{n} \sum_{j=1}^{n} \hat{\theta}_{-j} \right)^2,$$

where $\hat{\theta}_{-i}$ is the same as $\hat{\theta}$ but is based on $n-1$ observations $X_1, ..., X_{i-1}$, $X_{i+1}, ..., X_n$, $i = 1, ..., n$. Let $\bar{X}$ be the sample mean and $\hat{\theta} = \bar{X}^2$. Show that

$$v_{_J} = \frac{4\bar{X}^2 \hat{c}_2}{n-1} - \frac{4\bar{X}\hat{c}_3}{(n-1)^2} + \frac{\hat{c}_4 - \hat{c}_2^2}{(n-1)^3},$$

where $\hat{c}_k = n^{-1} \sum_{i=1}^{n} (X_i - \bar{X})^k$, $k = 2, 3, 4$.

Solution. Let $\bar{X}_{-i}$ be the sample mean based on $X_1, ..., X_{i-1}, X_{i+1}, ..., X_n$, $i = 1, ..., n$. Then $\hat{\theta}_{-i} = \bar{X}_{-i}^2$,

$$\bar{X}_{-i} = \frac{n\bar{X} - X_i}{n-1},$$

$$\bar{X}_{-i} - \bar{X} = \frac{\bar{X} - X_i}{n-1},$$

$$\bar{X}_{-i} + \bar{X} = \frac{\bar{X} - X_i}{n-1} + 2\bar{X},$$

and

$$\frac{1}{n} \sum_{i=1}^{n} \bar{X}_{-i}^2 = \frac{1}{n} \sum_{i=1}^{n} \left(\bar{X} + \frac{\bar{X} - X_i}{n-1} \right)^2 = \bar{X}^2 + \frac{\hat{c}_2}{(n-1)^2}.$$

Therefore,

$$\begin{aligned}
v_J &= \frac{n-1}{n}\sum_{i=1}^n\left(\bar{X}_{-i}^2 - \frac{1}{n}\sum_{j=1}^n \bar{X}_{-j}^2\right)^2 \\
&= \frac{n-1}{n}\sum_{i=1}^n\left(\bar{X}_{-i}^2 - \bar{X}^2 + \bar{X}^2 - \frac{1}{n}\sum_{j=1}^n \bar{X}_{-j}^2\right)^2 \\
&= \frac{n-1}{n}\sum_{i=1}^n\left(\bar{X}_{-i}^2 - \bar{X}^2\right)^2 - (n-1)\left(\bar{X}^2 - \frac{1}{n}\sum_{j=1}^n \bar{X}_{-j}^2\right)^2 \\
&= \frac{n-1}{n}\sum_{i=1}^n\left(\bar{X}_{-i} - \bar{X}\right)^2\left(\bar{X}_{-i} + \bar{X}\right)^2 - \frac{\hat{c}_2^2}{(n-1)^3} \\
&= \frac{n-1}{n}\sum_{i=1}^n\left(\frac{\bar{X} - X_i}{n-1}\right)^2\left(\frac{\bar{X} - X_i}{n-1} + 2\bar{X}\right)^2 - \frac{\hat{c}_2^2}{(n-1)^3} \\
&= \frac{1}{n(n-1)^3}\sum_{i=1}^n\left(\bar{X} - X_i\right)^4 + \frac{4\bar{X}}{n(n-1)^2}\sum_{i=1}^n(\bar{X} - X_i)^3 \\
&\quad + \frac{4\bar{X}^2}{n(n-1)}\sum_{i=1}^n(\bar{X} - X_i)^2 - \frac{\hat{c}_2^2}{(n-1)^3} \\
&= \frac{\hat{c}_4}{(n-1)^3} - \frac{4\bar{X}\hat{c}_3}{(n-1)^2} + \frac{4\bar{X}^2\hat{c}_2}{n-1} - \frac{\hat{c}_2^2}{(n-1)^3} \\
&= \frac{4\bar{X}^2\hat{c}_2}{n-1} - \frac{4\bar{X}\hat{c}_3}{(n-1)^2} + \frac{\hat{c}_4 - \hat{c}_2^2}{(n-1)^3}. \quad \blacksquare
\end{aligned}$$

Exercise 39 (#5.111). Let $X_1, ..., X_n$ be random variables and $X_1^*, ..., X_n^*$ be a random sample (i.e., a simple random sample with replacement) from $X_1, ..., X_n$. For any estimator $\hat{\theta}$, its bootstrap variance estimator is $v_B = \text{Var}_*(\hat{\theta}^*)$, where $\hat{\theta}^*$ is the same as $\hat{\theta}$ but is based on $X_1^*, ..., X_n^*$ and Var_* is the variance with respect to the distribution of $X_1^*, ..., X_n^*$, given $X_1, ..., X_n$. Let $\bar{X}$ be the sample mean and $\hat{\theta} = \bar{X}^2$. Show that

$$v_B = \frac{4\bar{X}^2\hat{c}_2}{n} + \frac{4\bar{X}\hat{c}_3}{n^2} + \frac{\hat{c}_4 - \hat{c}_2^2}{n^3},$$

where $\hat{c}_k = n^{-1}\sum_{i=1}^n (X_i - \bar{X})^k$, $k = 2, 3, 4$.

Solution. Let E_* be the expectation with respect to the distribution of $X_1^*, ..., X_n^*$, given $X_1, ..., X_n$. Note that

$$E_*(\bar{X}^*) = E_*(X_1^*) = \bar{X},$$

$$\mathrm{Var}_*(\bar{X}^*) = \frac{\mathrm{Var}_*(X_1^*)}{n} = \frac{\hat{c}_2}{n},$$

and

$$E_*[(X_i^* - \bar{X})(X_j^* - \bar{X})(X_k^* - \bar{X})(X_l^* - \bar{X})] = \begin{cases} \hat{c}_4 & \text{if } i = j = k = l \\ \hat{c}_2^2 & \text{if } i = k, j = l, i \neq j \\ 0 & \text{otherwise} \end{cases}$$

Thus,

$$\begin{aligned} E_*(\bar{X}^* - \bar{X})^4 &= \frac{1}{n^4} \sum_{1 \leq i,j,k,l \leq n} E_*[(X_i^* - \bar{X})(X_j^* - \bar{X})(X_k^* - \bar{X})(X_l^* - \bar{X})] \\ &= \frac{1}{n^4} \sum_{1 \leq i \leq n} E_*(X_i^* - \bar{X})^4 \\ &\quad + \frac{1}{n^4} \sum_{1 \leq i,j \leq n, i \neq j} E_*[(X_i^* - \bar{X})^2 (X_j^* - \bar{X})^2] \\ &= \frac{\hat{c}_4}{n^3} + \frac{(n-1)\hat{c}_2^2}{n^3} \end{aligned}$$

and, hence,

$$\begin{aligned} \mathrm{Var}_*(\bar{X}^* - \bar{X})^2 &= E_*(\bar{X}^* - \bar{X})^4 - [E_*(\bar{X}^* - \bar{X})^2]^2 \\ &= \frac{\hat{c}_4}{n^3} + \frac{(n-1)\hat{c}_2^2}{n^3} - [\mathrm{Var}_*(\bar{X}^*)]^2 \\ &= \frac{\hat{c}_4 - \hat{c}_2^2}{n^3}. \end{aligned}$$

Also,

$$E_*[(X_i^* - \bar{X})(X_j^* - \bar{X})(X_k^* - \bar{X})] = \begin{cases} \hat{c}_3 & \text{if } i = j = k \\ 0 & \text{otherwise} \end{cases}$$

and, thus,

$$\begin{aligned} E_*(\bar{X}^* - \bar{X})^3 &= \frac{1}{n^3} \sum_{1 \leq i,j,k \leq n} E_*[(X_i^* - \bar{X})(X_j^* - \bar{X})(X_k^* - \bar{X})] \\ &= \frac{\hat{c}_3}{n^2}. \end{aligned}$$

Let Cov_* be the covariance with respect to the distribution of $X_1^*, ..., X_n^*$, given $X_1, ..., X_n$. Then

$$\mathrm{Cov}_*\left((\bar{X}^* - \bar{X})^2, \bar{X}^* - \bar{X}\right) = E_*(\bar{X}^* - \bar{X})^3 = \frac{\hat{c}_3}{n^2}.$$

Combining all the results, we obtain that

$$\begin{aligned}
\mathrm{Var}_*(\hat{\theta}^*) &= \mathrm{Var}_*(\bar{X}^{*2}) \\
&= \mathrm{Var}_*(\bar{X}^{*2} - \bar{X}^2) \\
&= \mathrm{Var}_*\left((\bar{X}^* - \bar{X})(\bar{X}^* - \bar{X} + 2\bar{X})\right) \\
&= \mathrm{Var}_*\left((\bar{X}^* - \bar{X})^2 + 2\bar{X}(\bar{X}^* - \bar{X})\right) \\
&= \mathrm{Var}_*\left((\bar{X}^* - \bar{X})^2\right) + 4\bar{X}^2\mathrm{Var}_*\left(\bar{X}^* - \bar{X}\right)\right) \\
&\quad + 4\bar{X}\mathrm{Cov}_*\left((\bar{X}^* - \bar{X})^2, \bar{X}^* - \bar{X}\right) \\
&= \frac{4\bar{X}^2\hat{c}_2}{n} + \frac{4\bar{X}\hat{c}_3}{n^2} + \frac{\hat{c}_4 - \hat{c}_2^2}{n^3}. \ \blacksquare
\end{aligned}$$

Exercise 40 (#5.113). Let $X_1, ..., X_n$ be a random sample from a distribution on $\mathcal{R}^k$ with a finite $\mathrm{Var}(X_1)$. Let $X_1^*, ..., X_n^*$ be a random sample from $X_1, ..., X_n$. Show that for almost all given sequences $X_1, X_2, ...$, $\sqrt{n}(\bar{X}^* - \bar{X}) \to_d N_k(0, \mathrm{Var}(X_1))$, where $\bar{X}$ is the sample mean based on $X_1, ..., X_n$ and $\bar{X}^*$ is the sample mean based on $X_1^*, ..., X_n^*$.
Solution. Since we can take linear combinations of components of $\bar{X}^* - \bar{X}$, it is enough to establish the result for the case $k = 1$, i.e., $X_1, ..., X_n$ are random variables.

Let $Y_i = X_i^* - \bar{X}$, $i = 1, ..., n$. Given $X_1, ..., X_n$, $Y_1, ..., Y_n$ are independent and identically distributed random variables with mean 0 and variance $\hat{c}_2 = n^{-1}\sum_{i=1}^n (X_i - \bar{X})^2$. Note that

$$\bar{X}^* - \bar{X} = \frac{1}{n}\sum_{i=1}^n Y_i$$

and

$$\mathrm{Var}_*(\bar{X}^* - \bar{X}) = \frac{\hat{c}_2}{n},$$

where Var_* is the variance with respect to the distribution of $X_1^*, ..., X_n^*$, given $X_1, ..., X_n$. To apply Lindeberg's central limit theorem, we need to check whether

$$\frac{1}{n\hat{c}_2}\sum_{i=1}^n E_*(Y_i^2 I_{\{|Y_i| > \epsilon\sqrt{n\hat{c}_2}\}})$$

converges to 0 as $n \to \infty$, where $\epsilon > 0$ is fixed and E_* is the expectation with respect to P_*, the conditional distribution of $X_1^*, ..., X_n^*$ given $X_1, ..., X_n$. Since Y_i's are identically distributed,

$$\frac{1}{n}\sum_{i=1}^n E_*(Y_i^2 I_{\{|Y_i| > \epsilon\sqrt{n\hat{c}_2}\}}) = E_*(Y_1^2 I_{\{|Y_1| > \epsilon\sqrt{n\hat{c}_2}\}}).$$

Note that

$$\begin{aligned} E_*(Y_1^2 I_{\{|Y_1|>\epsilon\sqrt{n\hat{c}_2}\}}) &\leq \left(\max_{1\leq i\leq n} X_i^2\right) P_*\left(|Y_1| > \epsilon\sqrt{n\hat{c}_2}\right) \\ &\leq \left(\max_{1\leq i\leq n} X_i^2\right) \frac{E_*|Y_1|^2}{\epsilon^2 n\hat{c}_2} \\ &= \frac{\max_{1\leq i\leq n} X_i^2}{\epsilon^2 n}, \end{aligned}$$

which converges to 0 a.s. (Exercise 46 in Chapter 1). Thus, by Lindeberg's central limit theorem, for almost all given sequences $X_1, X_2, ...,$

$$\frac{1}{\sqrt{n\hat{c}_2}} \sum_{i=1}^{n} Y_i \to_d N(0,1).$$

The result follows since $\lim_n \hat{c}_2 = \text{Var}(X_1)$ a.s. ∎

Chapter 6

Hypothesis Tests

Exercise 1 (#6.2). Let X be a sample from a population P and consider testing hypotheses $H_0 : P = P_0$ versus $H_1 : P = P_1$, where P_j is a known population with probability density f_j with respect to a σ-finite measure ν, $j = 0, 1$. Let $\beta(P)$ be the power function of a UMP (uniformly most powerful) test of size $\alpha \in (0, 1)$. Show that $\alpha < \beta(P_1)$ unless $P_0 = P_1$.
Solution. Suppose that $\alpha = \beta(P_1)$. Then the test $T_0 \equiv \alpha$ is also a UMP test by definition. By the uniqueness of the UMP test (e.g., Theorem 6.1(ii) in Shao, 2003), we must have $f_1(x) = cf_0(x)$ a.e. ν, which implies $c = 1$. Therefore, $f_1(x) = f_0(x)$ a.e. ν, i.e., $P_0 = P_1$. ∎

Exercise 2 (#6.3). Let X be a sample from a population P and consider testing hypotheses $H_0 : P = P_0$ versus $H_1 : P = P_1$, where P_j is a known population with probability density f_j with respect to a σ-finite measure ν, $j = 0, 1$. For any $\alpha > 0$, define

$$T_\alpha(X) = \begin{cases} 1 & f_1(X) > c(\alpha)f_0(X) \\ \gamma(\alpha) & f_1(X) = c(\alpha)f_0(X) \\ 0 & f_1(X) < c(\alpha)f_0(X), \end{cases}$$

where $0 \leq \gamma(\alpha) \leq 1$, $c(\alpha) \geq 0$, $E_0[T_\alpha(X)] = \alpha$, and E_j denotes the expectation with respect to P_j. Show that
(i) if $\alpha_1 < \alpha_2$, then $c(\alpha_1) \geq c(\alpha_2)$;
(ii) if $\alpha_1 < \alpha_2$, then the type II error probability of T_{α_1} is larger than that of T_{α_2}, i.e., $E_1[1 - T_{\alpha_1}(X)] > E_1[1 - T_{\alpha_2}(X)]$.
Solution. (i) Assume $\alpha_1 < \alpha_2$. Suppose that $c(\alpha_1) < c(\alpha_2)$. Then $f_1(x) \geq c(\alpha_2)f_0(x)$ implies that $f_1(x) > c(\alpha_1)f_0(x)$ unless $f_1(x) = f_0(x) = 0$. Thus, $T_{\alpha_1}(x) \geq T_{\alpha_2}(x)$ a.e. ν, which implies that $E_0[T_{\alpha_1}(X)] \geq E_0[T_{\alpha_2}(X)]$. Then $\alpha_1 \geq \alpha_2$. This contradiction proves that $c(\alpha_1) \geq c(\alpha_2)$.
(ii) Assume $\alpha_1 < \alpha_2$. Since T_{α_1} is of level α_2 and T_{α_2} is UMP, $E_1[T_{\alpha_1}(X)] \leq E_1[T_{\alpha_2}(X)]$. The result follows if we can show that the equality can not

hold. If $E_1[T_{\alpha_1}(X)] = E_1[T_{\alpha_2}(X)]$, then T_{α_1} is also UMP. By the uniqueness of the UMP and the fact that $c(\alpha_1) \geq c(\alpha_2)$ (part (i)),

$$P_j\big(c(\alpha_2)f_0(X) \leq f_1(X) \leq c(\alpha_1)f_0(X)\big) = 0, \quad j = 0, 1.$$

This implies that $E_0[T_{\alpha_2}(X)] = 0 < \alpha_2$. Thus, $E_1[T_{\alpha_1}(X)] < E_1[T_{\alpha_2}(X)]$, i.e., $E_1[1 - T_{\alpha_1}(X)] > E_1[1 - T_{\alpha_2}(X)]$. ∎

Exercise 3 (#6.4). Let X be a sample from a population P and P_0 and P_1 be two known populations. Suppose that T_* is a UMP test of size $\alpha \in (0,1)$ for testing $H_0 : P = P_0$ versus $H_1 : P = P_1$ and that $\beta < 1$, where β is the power of T_* when H_1 is true. Show that $1 - T_*$ is a UMP test of size $1 - \beta$ for testing $H_0 : P = P_1$ versus $H_1 : P = P_0$.
Solution. Let f_j be a probability density for P_j, $j = 0, 1$. By the uniqueness of the UMP test,

$$T_*(X) = \begin{cases} 1 & f_1(X) > cf_0(X) \\ 0 & f_1(X) < cf_0(X). \end{cases}$$

Since $\alpha \in (0,1)$ and $\beta < 1$, c must be a positive constant. Note that

$$1 - T_*(X) = \begin{cases} 1 & f_0(X) > c^{-1}f_1(X) \\ 0 & f_0(X) < c^{-1}f_1(X). \end{cases}$$

For testing $H_0 : P = P_1$ versus $H_1 : P = P_0$, clearly $1 - T_*$ has size $1 - \beta$. The fact that it is UMP follows from the Neyman-Pearson Lemma. ∎

Exercise 4 (#6.6). Let $(X_1, ..., X_n)$ be a random sample from a population on $\mathcal{R}$ with Lebesgue density f_θ. Let θ_0 and θ_1 be two constants. Find a UMP test of size α for testing $H_0 : \theta = \theta_0$ versus $H_1 : \theta = \theta_1$ in the following cases:
(i) $f_\theta(x) = e^{-(x-\theta)}I_{(\theta,\infty)}(x)$, $\theta_0 < \theta_1$;
(ii) $f_\theta(x) = \theta x^{-2}I_{(\theta,\infty)}(x)$, $\theta_0 \neq \theta_1$.
Solution. (i) Let $X_{(1)}$ be the smallest order statistic. Since

$$\frac{f_{\theta_1}(X)}{f_{\theta_0}(X)} = \begin{cases} e^{n(\theta_1 - \theta_0)} & X_{(1)} > \theta_1 \\ 0 & \theta_0 < X_{(1)} \leq \theta_1, \end{cases}$$

the UMP test is either

$$T_1 = \begin{cases} 1 & X_{(1)} > \theta_1 \\ \gamma & \theta_0 < X_{(1)} \leq \theta_1 \end{cases}$$

or

$$T_2 = \begin{cases} \gamma & X_{(1)} > \theta_1 \\ 0 & \theta_0 < X_{(1)} \leq \theta_1. \end{cases}$$

When $\theta = \theta_0$, $P(X_{(1)} > \theta_1) = e^{n(\theta_0-\theta_1)}$. If $e^{n(\theta_0-\theta_1)} \leq \alpha$, then T_1 is the UMP test since, under $\theta = \theta_0$,

$$\begin{aligned} E(T_1) &= P(X_{(1)} > \theta_1) + \gamma P(\theta_0 < X_{(1)} \leq \theta_1) \\ &= e^{n(\theta_0-\theta_1)} + \gamma(1 - e^{n(\theta_0-\theta_1)}) \\ &= \alpha \end{aligned}$$

with $\gamma = (\alpha - e^{n(\theta_0-\theta_1)})/(1 - e^{n(\theta_0-\theta_1)})$. If $e^{n(\theta_0-\theta_1)} > \alpha$, then T_2 is the UMP test since, under $\theta = \theta_0$,

$$E(T_2) = \gamma P(X_{(1)} > \theta_1) = \gamma e^{n(\theta_0-\theta_1)} = \alpha$$

with $\gamma = \alpha/e^{n(\theta_0-\theta_1)}$.
(ii) Suppose $\theta_1 > \theta_0$. Then

$$\frac{f_{\theta_1}(X)}{f_{\theta_0}(X)} = \begin{cases} \frac{\theta_1^n}{\theta_0^n} & X_{(1)} > \theta_1 \\ 0 & \theta_0 < X_{(1)} \leq \theta_1. \end{cases}$$

The UMP test is either

$$T_1 = \begin{cases} 1 & X_{(1)} > \theta_1 \\ \gamma & \theta_0 < X_{(1)} \leq \theta_1 \end{cases}$$

or

$$T_2 = \begin{cases} \gamma & X_{(1)} > \theta_1 \\ 0 & \theta_0 < X_{(1)} \leq \theta_1. \end{cases}$$

When $\theta = \theta_0$, $P(X_{(1)} > \theta_1) = \theta_0^n/\theta_1^n$. If $\theta_0^n/\theta_1^n \leq \alpha$, then T_1 is the UMP test since, under $\theta = \theta_0$,

$$E(T_1) = \frac{\theta_0^n}{\theta_1^n} + \gamma\left(1 - \frac{\theta_0^n}{\theta_1^n}\right) = \alpha$$

with $\gamma = (\alpha - \frac{\theta_0^n}{\theta_1^n})/(1 - \frac{\theta_0^n}{\theta_1^n})$. If $\theta_0^n/\theta_1^n > \alpha$, then T_2 is the UMP test since, under $\theta = \theta_0$,

$$E(T_2) = \gamma\frac{\theta_0^n}{\theta_1^n} = \alpha$$

with $\gamma = \alpha\theta_1^n/\theta_0^n$.

Suppose now that $\theta_1 < \theta_0$. Then

$$\frac{f_{\theta_1}(X)}{f_{\theta_0}(X)} = \begin{cases} \frac{\theta_1^n}{\theta_0^n} & X_{(1)} > \theta_0 \\ \infty & \theta_1 < X_{(1)} \leq \theta_0. \end{cases}$$

The UMP test is either

$$T_1 = \begin{cases} 0 & X_{(1)} > \theta_0 \\ \gamma & \theta_1 < X_{(1)} \leq \theta_0 \end{cases}$$

or

$$T_2 = \begin{cases} \gamma & X_{(1)} > \theta_0 \\ 1 & \theta_1 < X_{(1)} \le \theta_0. \end{cases}$$

When $\theta = \theta_0$, $E(T_1) = 0$ and $E(T_2) = \gamma$. Hence, the UMP test is T_2 with $\gamma = \alpha$. ∎

Exercise 5 (#6.7). Let $f_1, ..., f_{m+1}$ be Borel functions on $\mathcal{R}^p$ that are integrable with respect to a σ-finite measure ν. For given constants $t_1, ..., t_m$, let $\mathcal{T}$ be the class of Borel functions ϕ (from $\mathcal{R}^p$ to $[0, 1]$) satisfying

$$\int \phi f_i d\nu \le t_i, \quad i = 1, ..., m,$$

and $\mathcal{T}_0$ be the set of ϕ's in $\mathcal{T}$ satisfying

$$\int \phi f_i d\nu = t_i, \quad i = 1, ..., m.$$

Show that if there are constants $c_1, ..., c_m$ such that

$$\phi_*(x) = \begin{cases} 1 & f_{m+1}(x) > c_1 f_1(x) + \cdots + c_m f_m(x) \\ 0 & f_{m+1}(x) < c_1 f_1(x) + \cdots + c_m f_m(x) \end{cases}$$

is a member of $\mathcal{T}_0$, then ϕ_* maximizes $\int \phi f_{m+1} d\nu$ over $\phi \in \mathcal{T}_0$. Show that if $c_i \ge 0$ for all i, then ϕ_* maximizes $\int \phi f_{m+1} d\nu$ over $\phi \in \mathcal{T}$.

Solution. Suppose that $\phi_* \in \mathcal{T}_0$. By the definition of ϕ_*, for any other $\phi \in \mathcal{T}_0$,

$$(\phi_* - \phi)(f_{m+1} - c_1 f_1 - \cdots - c_m f_m) \ge 0.$$

Therefore

$$\int (\phi_* - \phi)(f_{m+1} - c_1 f_1 - \cdots - c_m f_m) d\nu \ge 0,$$

i.e.,

$$\int (\phi_* - \phi) f_{m+1} d\nu \ge \sum_{i=1}^m c_i \int (\phi_* - \phi) f_i d\nu = 0.$$

Hence ϕ_* maximizes $\int \phi f_{m+1} d\nu$ over $\phi \in \mathcal{T}_0$. If $c_i \ge 0$, for $\phi \in \mathcal{T}$, we still have

$$(\phi_* - \phi)(f_{m+1} - c_1 f_1 - \cdots - c_m f_m) \ge 0$$

and, thus,

$$\int (\phi_* - \phi) f_{m+1} d\nu \ge \sum_{i=1}^n c_i \int (\phi_* - \phi) f_i d\nu \ge 0,$$

because $c_i \int (\phi_* - \phi) f_i d\nu \ge 0$ for each i. Therefore ϕ_* maximizes $\int \phi f_{m+1} d\nu$ over $\phi \in \mathcal{T}$. ∎

Exercise 6 (#6.9). Let f_0 and f_1 be Lebesgue integrable functions on $\mathcal{R}$ and

$$\phi_*(x) = \begin{cases} 1 & f_0(x)<0 \text{ or } f_0(x)=0, f_1(x)\geq 0 \\ 0 & \text{otherwise.} \end{cases}$$

Show that ϕ_* maximizes $\int \phi(x)f_1(x)dx$ over all Borel functions ϕ on $\mathcal{R}$ satisfying $0 \leq \phi(x) \leq 1$ and $\int \phi(x)f_0(x)dx = \int \phi_*(x)f_0(x)dx$.

Solution. From the definition of ϕ_*, $\int \phi_*(x)f_0(x)dx = \int_{\{f_0(x)<0\}} f_0(x)dx$. Since $0 \leq \phi(x) \leq 1$ and $\int \phi(x)f_0(x)dx = \int \phi_*(x)f_0(x)dx$,

$$\begin{aligned} 0 &\leq \int_{\{f_0(x)>0\}} \phi(x)f_0(x)dx \\ &= \int \phi(x)f_0(x)dx - \int_{\{f_0(x)<0\}} \phi(x)f_0(x)dx \\ &= \int \phi_*(x)f_0(x)dx - \int_{\{f_0(x)<0\}} \phi(x)f_0(x)dx \\ &= \int_{\{f_0(x)<0\}} f_0(x)dx - \int_{\{f_0(x)<0\}} \phi(x)f_0(x)dx \\ &= \int_{\{f_0(x)<0\}} [1-\phi(x)]f_0(x)dx \\ &\leq 0. \end{aligned}$$

That is,

$$\int_{\{f_0(x)>0\}} \phi(x)f_0(x)dx = \int_{\{f_0(x)<0\}} [1-\phi(x)]f_0(x)dx = 0.$$

Hence, $\phi(x) = 0$ a.e. on the set $\{f_0(x) > 0\}$ and $\phi(x) = 1$ a.e. on the set $\{f_0(x) < 0\}$. Then, the result follows from

$$\begin{aligned} \int [\phi_*(x) - \phi(x)]f_1(x)dx &= \int_{\{f_0(x)<0\}} [1-\phi(x)]f_1(x)dx \\ &\quad - \int_{\{f_0(x)>0\}} \phi(x)f_1(x)dx \\ &\quad + \int_{\{f_0(x)=0, f_1(x)\geq 0\}} [1-\phi(x)]f_1(x)dx \\ &\quad - \int_{\{f_0(x)=0, f_1(x)<0\}} \phi(x)f_1(x)dx \\ &= \int_{\{f_0(x)=0, f_1(x)\geq 0\}} [1-\phi(x)]f_1(x)dx \\ &\quad - \int_{\{f_0(x)=0, f_1(x)<0\}} \phi(x)f_1(x)dx \\ &\geq 0. \quad \blacksquare \end{aligned}$$

Exercise 7 (#6.10). Let F_1 and F_2 be two cumulative distribution functions on $\mathcal{R}$. Show that $F_1(x) \leq F_2(x)$ for all x if and only if $\int g(x)dF_2(x) \leq \int g(x)dF_1(x)$ for any nondecreasing function g.

Solution. If $\int g(x)dF_2(x) \leq \int g(x)dF_1(x)$ for any nondecreasing function g, then

$$1 - F_2(y) = \int I_{(y,\infty)}(x)dF_2(x) \leq \int I_{(y,\infty)}(x)dF_1(x) = 1 - F_1(y)$$

for any y, since $I_{(y,\infty)}(x)$ is nondecreasing. Assume now that $F_1(x) \leq F_2(x)$ for all x. Then, for any $t \in \mathcal{R}$, $\{x : F_1(x) \geq t\} \subset \{x : F_2(x) \geq t\}$ and, hence,

$$F_1^{-1}(t) = \inf\{x : F_1(x) \geq t\} \geq \inf\{x : F_2(x) \geq t\} = F_2^{-1}(t)$$

for any t. Let U be a random variable having the uniform distribution on $(0,1)$. Then $F_j^{-1}(U)$ has distribution F_j, $j = 1, 2$. If g is nondecreasing, then $g(F_1^{-1}(U)) \geq g(F_2^{-1}(U))$ and, therefore,

$$\int g(x)dF_1(x) = E[g(F_1^{-1}(U))] \geq E[g(F_2^{-1}(U))] = \int g(x)dF_2(x). \blacksquare$$

Exercise 8 (#6.11). Let X be an observation with a probability density in the family $\mathcal{P} = \{f_\theta : \theta \in \Theta\}$, where $\Theta \subset \mathcal{R}$ is the possible values of the parameter θ.

(i) Show that $\mathcal{P}$ has monotone likelihood ratio in X when $\Theta = \mathcal{R}$ and f_θ is the Lebesgue density of the double exponential distribution with location parameter θ and a known scale parameter c.

(ii) Show that $\mathcal{P}$ has monotone likelihood ratio in X when $\Theta = \mathcal{R}$ and f_θ is the Lebesgue density of the exponential distribution on the interval (θ, ∞) with a known scale parameter c.

(iii) Show that $\mathcal{P}$ has monotone likelihood ratio in X when $\Theta = \mathcal{R}$ and f_θ is the Lebesgue density of the logistic distribution with location parameter θ and a known scale parameter c.

(iv) Show that $\mathcal{P}$ has monotone likelihood ratio in X when $\Theta = \mathcal{R}$ and f_θ is the Lebesgue density of the uniform distribution on $(\theta, \theta + 1)$.

(v) Show that $\mathcal{P}$ has monotone likelihood ratio in X when $\Theta = \{1, 2, ...\}$ and $f_\theta(x) = \binom{\theta}{x}\binom{N-\theta}{r-x}/\binom{N}{r}$ when x is an integer between $r - \theta$ and $\min\{r, \theta\}$, where r and N are known integers.

(vi) Show that $\mathcal{P}$ does not have monotone likelihood ratio in X when $\Theta = \mathcal{R}$ and f_θ is the Lebesgue density of the Cauchy distribution with location parameter θ and a known scale parameter c.

Solution. (i) We need to show $f_{\theta_2}(x)/f_{\theta_1}(x)$ is nondecreasing in x for any

$\theta_1 < \theta_2$ with at least one of $f_{\theta_i}(x)$ is positive. For $\theta_1 < \theta_2$,

$$\frac{f_{\theta_2}(x)}{f_{\theta_1}(x)} = e^{-(|\theta_2 - x| - |\theta_1 - x|)/c} = \begin{cases} e^{-(\theta_2-\theta_1)/c} & x \le \theta_1 \\ e^{-(\theta_2+\theta_1-2x)/c} & \theta_1 < x \le \theta_2 \\ e^{(\theta_2-\theta_1)/c} & x > \theta_2, \end{cases}$$

which is a nondecreasing function of x.
(ii) For $\theta_1 < \theta_2$,

$$\frac{f_{\theta_2}(x)}{f_{\theta_1}(x)} = \begin{cases} 0 & \theta_1 < x \le \theta_2 \\ e^{(\theta_2-\theta_1)/c} & x > \theta_2, \end{cases}$$

which is a nondecreasing function of x.
(iii) For $\theta_1 < \theta_2$,

$$\frac{f_{\theta_2}(x)}{f_{\theta_1}(x)} = e^{(\theta_1-\theta_2)/c} \left(\frac{1 + e^{(x-\theta_1)/c}}{1 + e^{(x-\theta_2)/c}} \right)^2 .$$

Since

$$\frac{d}{dx} \left(\frac{1 + e^{(x-\theta_1)/c}}{1 + e^{(x-\theta_2)/c}} \right) = \frac{e^{(x-\theta_1)/c} - e^{(x-\theta_2)/c}}{c(1 + e^{(x-\theta_2)/c})^2} > 0$$

when $\theta_1 < \theta_2$, the ratio $f_{\theta_2}(x)/f_{\theta_1}(x)$ is increasing in x.
(iv) For $\theta_1 < \theta_2 < \theta_1 + 1$,

$$\frac{f_{\theta_2}(x)}{f_{\theta_1}(x)} = \begin{cases} 0 & \theta_1 < x \le \theta_2 \\ 1 & \theta_2 < x < \theta_1 + 1 \\ \infty & \theta_1 + 1 \le x < \theta_2 + 1. \end{cases}$$

For $\theta_1 + 1 \le \theta_2$,

$$\frac{f_{\theta_2}(x)}{f_{\theta_1}(x)} = \begin{cases} 0 & \theta_1 < x < \theta_1 + 1 \\ \infty & \theta_2 \le x < \theta_2 + 1. \end{cases}$$

In any case, the ratio $f_{\theta_2}(x)/f_{\theta_1}(x)$ is nondecreasing in x.
(v) Note that

$$\frac{f_\theta(x)}{f_{\theta-1}(x)} = \frac{\binom{\theta}{x}\binom{N-\theta}{r-x}/\binom{N}{r}}{\binom{\theta-1}{x}\binom{N-\theta+1}{r-x}/\binom{N}{r}} = \frac{\theta(N-\theta-r+x+1)}{(\theta-x)(N-\theta+1)}$$

is an increasing function of x. Hence, for $\theta_1 < \theta_2$,

$$\frac{f_{\theta_2}(x)}{f_{\theta_1}(x)} = \frac{f_{\theta_1+1}(x)}{f_{\theta_1}(x)} \frac{f_{\theta_1+2}(x)}{f_{\theta_1+1}(x)} \cdots \frac{f_{\theta_2}(x)}{f_{\theta_2-1}(x)}$$

is a product of increasing functions in x and, hence, it is increasing in x.
(vi) For $\theta_1 < \theta_2$,

$$\frac{f_{\theta_2}(x)}{f_{\theta_1}(x)} = \frac{c^2 + (x-\theta_1)^2}{c^2 + (x-\theta_2)^2},$$

which converges to 1 when $x \to \pm\infty$, is smaller than 1 when $x = \theta_1$, and is larger than 1 when $x = \theta_2$. Hence, the ratio $f_{\theta_2}(x)/f_{\theta_1}(x)$ is not monotone in x. ∎

Exercise 9. Let $\Theta \subset \mathcal{R}$ and $\mathcal{P} = \{f_\theta(x) : \theta \in \Theta\}$ be a family of functions on $\mathcal{X} \subset \mathcal{R}$ satisfying $f_\theta(x) > 0$ for all $\theta \in \Theta$ and $x \in \mathcal{X}$. Assume that $\frac{\partial^2}{\partial\theta\partial x} \log f_\theta(x)$ exists.
(i) Show that $\mathcal{P}$ has monotone likelihood ratio in x is equivalent to one of the following conditions:
(a) $\frac{\partial^2}{\partial\theta\partial x} \log f_\theta(x) \geq 0$ for all x and θ;
(b) $f_\theta(x)\frac{\partial^2}{\partial\theta\partial x} f_\theta(x) \geq \frac{\partial}{\partial\theta} f_\theta(x) \frac{\partial}{\partial x} f_\theta(x)$ for all x and θ.
(ii) Let $f_\theta(x)$ be the Lebesgue density of the noncentral chi-square distribution $\chi_1^2(\theta)$ with the noncentrality parameter $\theta \geq 0$. Show that the family $\mathcal{P} = \{f_\theta(x) : \theta \geq 0\}$ has monotone likelihood ratio in x.
Solution. (i) Note that

$$\frac{\partial^2}{\partial\theta\partial x} \log f_\theta(x) = \frac{\partial}{\partial\theta} \frac{\frac{\partial}{\partial x} f_\theta(x)}{f_\theta(x)} = \frac{\frac{\partial^2}{\partial\theta\partial x} f_\theta(x)}{f_\theta(x)} - \frac{\frac{\partial}{\partial x} f_\theta(x)\frac{\partial}{\partial\theta} f_\theta(x)}{[f_\theta(x)]^2}.$$

Since $f_\theta(x) > 0$, conditions (a) and (b) are equivalent.

Condition (a) is equivalent to

$$\frac{\partial}{\partial x} \log f_\theta(x) = \frac{\frac{\partial}{\partial x} f_\theta(x)}{f_\theta(x)}$$

is nondecreasing in θ for any fixed x. Hence, it is equivalent to, for $\theta_1 < \theta_2$ and any x,

$$\frac{\frac{\partial}{\partial x} f_{\theta_1}(x)}{f_{\theta_1}(x)} \leq \frac{\frac{\partial}{\partial x} f_{\theta_2}(x)}{f_{\theta_2}(x)},$$

which is equivalent to, for $\theta_1 < \theta_2$ and any x,

$$\frac{\partial}{\partial x} \frac{f_{\theta_2}(x)}{f_{\theta_1}(x)} = \frac{f_{\theta_1}(x)\frac{\partial}{\partial x} f_{\theta_2}(x) - f_{\theta_2}(x)\frac{\partial}{\partial x} f_{\theta_1}(x)}{[f_{\theta_1}(x)]^2} \geq 0,$$

i.e., $\mathcal{P}$ has monotone likelihood ratio in x.
(ii) Let Z be a random variable having distribution $N(\sqrt{\theta}, 1)$. By definition, Z^2 has the noncentral chi-square distribution $\chi_1^2(\theta)$. Hence,

$$f_\theta(x) = \frac{1}{2\sqrt{2\pi x}} \left[e^{-(\sqrt{x}-\theta)^2/2} + e^{-(\sqrt{x}+\theta)^2/2} \right].$$

For $0 \leq \theta_1 < \theta_2$,

$$\frac{f_{\theta_2}(x)}{f_{\theta_1}(x)} = \frac{e^{-(\sqrt{x}-\theta_2)^2/2} + e^{-(\sqrt{x}+\theta_2)^2/2}}{e^{-(\sqrt{x}-\theta_1)^2/2} + e^{-(\sqrt{x}+\theta_1)^2/2}} = \frac{e^{-\theta_2^2/2}(e^{\theta_2\sqrt{x}} + e^{-\theta_2\sqrt{x}})}{e^{-\theta_1^2/2}(e^{\theta_1\sqrt{x}} + e^{-\theta_1\sqrt{x}})}.$$

Hence, we may apply the result in (i) to functions $g_\theta(y) = e^{\theta y} + e^{-\theta y}$. Note that

$$\frac{\partial}{\partial y} g_\theta(y) = \theta(e^{\theta y} - e^{-\theta y}),$$

$$\frac{\partial}{\partial \theta} g_\theta(y) = y(e^{\theta y} - e^{-\theta y}),$$

and

$$\frac{\partial^2}{\partial \theta \partial y} g_\theta(y) = \theta y(e^{\theta y} + e^{-\theta y}).$$

Hence,

$$\begin{aligned} g_\theta(y)\frac{\partial^2}{\partial \theta \partial y} g_\theta(y) &= \theta y(e^{\theta y} + e^{-\theta y})^2 \\ &\geq \theta y(e^{\theta y} - e^{-\theta y})^2 \\ &= \frac{\partial}{\partial y} g_\theta(y) \frac{\partial}{\partial \theta} g_\theta(y), \end{aligned}$$

i.e., condition (b) in (i) holds. Hence $\mathcal{P}$ has monotone likelihood ratio in y. Since $y = \sqrt{x}$ is an increasing function of x, $\mathcal{P}$ also has monotone likelihood ratio in x. ∎

Exercise 10 (#6.14). Let $X = (X_1, ..., X_n)$ be a random sample from a distribution on $\mathcal{R}$ with Lebesgue density f_θ, $\theta \in \Theta = (0, \infty)$. Let θ_0 be a positive constant. Find a UMP test of size α for testing $H_0 : \theta \leq \theta_0$ versus $H_1 : \theta > \theta_0$ when
(i) $f_\theta(x) = \theta^{-1} e^{-x/\theta} I_{(0,\infty)}(x)$;
(ii) $f_\theta(x) = \theta^{-1} x^{\theta-1} I_{(0,1)}(x)$;
(iii) $f_\theta(x)$ is the density of $N(1, \theta)$;
(iv) $f_\theta(x) = \theta^{-c} c x^{c-1} e^{-(x/\theta)^c} I_{(0,\infty)}(x)$, where $c > 0$ is known.
Solution. (i) The family of densities has monotone likelihood ratio in $T(X) = \sum_{i=1}^n X_i$, which has the gamma distribution with shape parameter n and scale parameter θ. Under H_0, $2T/\theta_0$ has the chi-square distribution χ^2_{2n}. Hence, the UMP test is

$$T_*(X) = \begin{cases} 1 & T(X) > \theta_0 \chi^2_{2n,\alpha}/2 \\ 0 & T(X) \leq \theta_0 \chi^2_{2n,\alpha}/2, \end{cases}$$

where $\chi^2_{r,\alpha}$ is the $(1-\alpha)$th quantile of the chi-square distribution χ^2_r.
(ii) The family of densities has monotone likelihood ratio in $T(X) = \sum_{i=1}^n \log X_i$, which has the gamma distribution with shape parameter n and scale parameter θ^{-1}. Therefore, the UMP test is the same as T_* in part (i) of the solution but with θ_0 replaced by θ_0^{-1}.
(iii) The family of densities has monotone likelihood ratio in $T(X) = \sum_{i=1}^n (X_i - 1)^2$ and $T(X)/\theta$ has the chi-square distribution χ^2_n. Therefore, the UMP test is

$$T_*(X) = \begin{cases} 1 & T(X) > \theta_0 \chi^2_{n,\alpha} \\ 0 & T(X) \le \theta_0 \chi^2_{n,\alpha}. \end{cases}$$

(iv) The family of densities has monotone likelihood ratio in $T(X) = \sum_{i=1}^n X_i^c$, which has the Gamma distribution with shape parameter n and scale parameter θ^c. Therefore, the UMP test is the same as T_* in part (i) of the solution but with θ_0 replaced by θ_0^c. ∎

Exercise 11 (#6.15). Suppose that the distribution of X is in a family $\{f_\theta : \theta \in \Theta\}$ with monotone likelihood ratio in $Y(X)$, where $Y(X)$ has a continuous distribution. Consider the hypotheses $H_0 : \theta \le \theta_0$ versus $H_1 : \theta > \theta_0$, where $\theta_0 \in \Theta$ is known. Show that the p-value of the UMP test is given by $P_{\theta_0}(Y \ge y)$, where y is the observed value of Y and P_θ is the probability corresponding to f_θ.
Solution. The UMP test of size α is

$$T_\alpha = \begin{cases} 1 & Y \ge c_\alpha \\ 0 & Y < c_\alpha, \end{cases}$$

where c_α satisfies $P_{\theta_0}(Y \ge c_\alpha) = \alpha$. When y is the observed value of Y, the rejection region of the UMP test is $\{y \ge c_\alpha\}$. By the definition of p-value, it is equal to

$$\begin{aligned} \hat{\alpha} &= \inf\{\alpha : 0 < \alpha < 1, T_\alpha = 1\} \\ &= \inf\{\alpha : 0 < \alpha < 1, y \ge c_\alpha\} \\ &= \inf_{y \ge c_\alpha} P_{\theta_0}(Y \ge c_\alpha) \\ &\ge P_{\theta_0}(Y \ge y), \end{aligned}$$

where the inequality follows from $P_{\theta_0}(Y \ge y) \le P_{\theta_0}(Y \ge c_\alpha)$ for any α such that $y \ge c_\alpha$. Let F_θ be the cumulative distribution function of P_θ. Since F_{θ_0} is continuous, $c_\alpha = F_{\theta_0}^{-1}(1-\alpha)$. Let $\alpha^* = P_{\theta_0}(Y \ge y) = 1 - F_{\theta_0}(y)$. Since F_{θ_0} is continuous,

$$c_{\alpha^*} = F_{\theta_0}^{-1}(1-\alpha^*) = F_{\theta_0}^{-1}(F_{\theta_0}(y)) \le y.$$

This means that $\alpha^* \in \{\alpha : 0 < \alpha < 1, y \ge c_\alpha\}$ and, thus, the p-value $\hat{\alpha} \le \alpha^*$. Therefore, the p-value is equal to $\alpha^* = P_{\theta_0}(Y \ge y)$. ∎

Exercise 12 (#6.17). Let F and G be two known cumulative distribution functions on $\mathcal{R}$ and X be a single observation from the cumulative distribution function $\theta F(x) + (1-\theta)G(x)$, where $\theta \in [0,1]$ is unknown.
(i) Find a UMP test of size α for testing $H_0 : \theta \leq \theta_0$ versus $H_1 : \theta > \theta_0$, where $\theta_0 \in [0,1]$ is known.
(ii) Show that the test $T_*(X) \equiv \alpha$ is a UMP test of size α for testing $H_0 : \theta \leq \theta_1$ or $\theta \geq \theta_2$ versus $H_1 : \theta_1 < \theta < \theta_2$, where $\theta_j \in [0,1]$ is known, $j = 1,2$, and $\theta_1 < \theta_2$.
Solution. (i) Let $f(x)$ and $g(x)$ be the Randon-Nikodym derivatives of $F(x)$ and $G(x)$ with respect to the measure ν induced by $F(x) + G(x)$, respectively. The probability density of X is $\theta f(x) + (1-\theta)g(x)$. For $0 \leq \theta_1 < \theta_2 \leq 1$,

$$\frac{\theta_2 f(x) + (1-\theta_2)g(x)}{\theta_1 f(x) + (1-\theta_1)g(x)} = \frac{\theta_2 \frac{f(x)}{g(x)} + (1-\theta_2)}{\theta_1 \frac{f(x)}{g(x)} + (1-\theta_1)}$$

is nondecreasing in $Y(x) = f(x)/g(x)$. Hence, the family of densities of X has monotone likelihood ratio in $Y(X) = f(X)/g(X)$ and a UMP test is given as

$$T = \begin{cases} 1 & Y(X) > c \\ \gamma & Y(X) = c \\ 0 & Y(X) < c, \end{cases}$$

where c and γ are uniquely determined by $E[T(X)] = \alpha$ when $\theta = \theta_0$.
(ii) For any test T, its power is

$$\begin{aligned} \beta_T(\theta) &= \int T(x)[\theta f(x) + (1-\theta)g(x)]d\nu \\ &= \theta \int T(x)[f(x) - g(x)]d\nu + \int T(x)g(x)du, \end{aligned}$$

which is a linear function of θ on $[0,1]$. If T has level α, then $\beta_T(\theta) \leq \alpha$ for any $\theta \in [0,1]$. Since the power of T_* is equal to the constant α, we conclude that T_* is a UMP test of size α. ∎

Exercise 13 (#6.18). Let $(X_1, ..., X_n)$ be a random sample from the uniform distribution on $(\theta, \theta+1)$, $\theta \in \mathcal{R}$. Suppose that $n \geq 2$.
(i) Show that a UMP test of size $\alpha \in (0,1)$ for testing $H_0 : \theta \leq 0$ versus $H_1 : \theta > 0$ is of the form

$$T_*(X_{(1)}, X_{(n)}) = \begin{cases} 0 & X_{(1)} < 1 - \alpha^{1/n},\ X_{(n)} < 1 \\ 1 & \text{otherwise,} \end{cases}$$

where $X_{(j)}$ is the jth order statistic.
(ii) Does the family of all densities of $(X_{(1)}, X_{(n)})$ have monotone likelihood

ratio?

Solution A. (i) The Lebesgue density of $(X_{(1)}, X_{(n)})$ is

$$f_\theta(x, y) = n(n-1)(y-x)^{n-2} I_{(\theta,y)}(x) I_{(x,\theta+1)}(y).$$

A direct calculation of $\beta_{T_*}(\theta) = \int T_*(x,y) f_\theta(x,y)dxdy$, the power function of T_*, leads to

$$\beta_{T_*}(\theta) = \begin{cases} 0 & \theta < -\alpha^{1/n} \\ (\theta + \alpha^{1/n})^n & -\alpha^{1/n} \le \theta \le 0 \\ 1 + \alpha - (1-\theta)^n & 0 < \theta \le 1 - \alpha^{1/n} \\ 1 & \theta > 1 - \alpha^{1/n}. \end{cases}$$

For any $\theta_1 \in (0, 1-\alpha^{1/n}]$, by the Neyman-Pearson Lemma, the UMP test T of size α for testing $H_0 : \theta = 0$ versus $H_1 : \theta = \theta_1$ is

$$T = \begin{cases} 1 & X_{(n)} > 1 \\ \alpha/(1-\theta_1)^n & \theta_1 < X_{(1)} < X_{(n)} < 1 \\ 0 & \text{otherwise.} \end{cases}$$

The power of T at θ_1 is computed as

$$\beta_T(\theta_1) = 1 - (1-\theta_1)^n + \alpha,$$

which agrees with the power of T_* at θ_1. When $\theta > 1 - \alpha^{1/n}$, T_* has power 1. Therefore T_* is a UMP test of size α for testing $H_0 : \theta \le 0$ versus $H_1 : \theta > 0$.

(ii) The answer is no. Suppose that the family of densities of $(X_{(1)}, X_{(n)})$ has monotone likelihood ratio. By the theory of UMP test (e.g., Theorem 6.2 in Shao, 2003), there exists a UMP test T_0 of size $\alpha \in (0, \frac{1}{2})$ for testing $H_0 : \theta \le 0$ versus $H_1 : \theta > 0$ and T_0 has the property that, for $\theta_1 \in (0, 1-\alpha^{1/n})$, T_0 is UMP of size $\alpha_0 = 1 + \alpha - (1-\theta_1)^n$ for testing $H_0 : \theta \le \theta_1$ versus $H_1 : \theta > \theta_1$. Using the transformation $X_i - \theta_1$ and the result in (i), the test

$$T_{\theta_1}(X_{(1)}, X_{(n)}) = \begin{cases} 0 & X_{(1)} < 1 + \theta_1 - \alpha_0^{1/n}, X_{(n)} < 1 + \theta_1 \\ 1 & \text{otherwise} \end{cases}$$

is a UMP test of size α_0 for testing $H_0 : \theta \le \theta_1$ versus $H_1 : \theta > \theta_1$. At $\theta = \theta_2 \in (\theta_1, 1-\alpha^{1/n}]$, it follows from part (i) of the solution that the power of T_0 is $1+\alpha-(1-\theta_2)^n$ and the power of T_{θ_1} is $1+\alpha_0-[1-(\theta_2-\theta_1)]^n$. Since both T_0 and T_{θ_1} are UMP tests, $1+\alpha-(1-\theta_2)^n = 1+\alpha_0-[1-(\theta_2-\theta_1)]^n$. Because $\alpha_0 = 1 + \alpha - (1-\theta_1)^n$, this means that

$$1 = (1-\theta_1)^n - (1-\theta_2)^n + [1 - (\theta_2 - \theta_1)]^n$$

holds for all $0 < \theta_1 < \theta_2 \le 1-\alpha^{1/n}$, which is impossible. This contradiction proves that the family of all densities of $(X_{(1)}, X_{(n)})$ does not have monotone likelihood ratio.

Solution B. This is an alternative solution to part (i) provided by Mr. Jialiang Li in 2002 as a student at the University of Wisconsin-Madison. Let $\beta_T(\theta)$ be the power function of a test T. Since $\beta_{T_*}(\theta) = 1$ when $\theta > 1-\alpha^{1/n}$, it suffices to show that $\beta_{T_*}(\theta) \ge \beta_T(\theta)$ for $\theta \in (0, 1-\alpha^{1/n})$ and any other test T. Define $A = \{0 < X_{(1)} \le X_{(n)} < 1\}$, $B = \{\theta < X_{(1)} \le X_{(n)} < 1\}$, and $C = \{1 < X_{(1)} \le X_{(n)} < \theta + 1\}$. Then

$$\begin{aligned}
\beta_{T_*}(\theta) - \beta_T(\theta) &= E[(T_* - T)I_B] + E[(T_* - T)I_C] \\
&= E[(T_* - T)I_B] + E[(1 - T)I_C] \\
&\ge E(T_* I_B) - E(T I_B) \\
&= E(T_* I_A) - E(T I_B) \\
&\ge E(T_* I_A) - E(T I_A) \\
&= \beta_{T_*}(0) - \beta_T(0) \\
&= \alpha - \beta_T(0),
\end{aligned}$$

where the second equality follows from $T_* = 1$ when $(X_{(1)}, X_{(n)}) \in C$, the third equality follows from $T_* = 0$ when $(X_{(1)}, X_{(n)}) \in A$ but $(X_{(1)}, X_{(n)}) \notin B$ (since $0 < X_{(1)} \le \theta \le 1 - \alpha^{1/n}$), and the second inequality follows from $I_A \ge I_B$. Therefore, if T has level α, then $\beta_{T_*}(\theta) \ge \beta_T(\theta)$ for all $\theta > 0$. ∎

Exercise 14 (#6.19). Let $X = (X_1, ..., X_n)$ be a random sample from the discrete uniform distribution on points $1, ..., \theta$, where $\theta = 1, 2, ...$

(i) Consider $H_0 : \theta \le \theta_0$ versus $H_1 : \theta > \theta_0$, where $\theta_0 > 0$ is known. Show that

$$T_*(X) = \begin{cases} 1 & X_{(n)} > \theta_0 \\ \alpha & X_{(n)} \le \theta_0 \end{cases}$$

is a UMP test of size α.

(ii) Consider $H_0 : \theta = \theta_0$ versus $H_1 : \theta \ne \theta_0$. Show that

$$T_*(X) = \begin{cases} 1 & X_{(n)} > \theta_0 \text{ or } X_{(n)} \le \theta_0 \alpha^{1/n} \\ 0 & \text{otherwise} \end{cases}$$

is a UMP test of size α.

(iii) Show that the results in (i) and (ii) still hold if the discrete uniform distribution is replaced by the uniform distribution on the interval $(0, \theta)$, $\theta > 0$.

Solution A. In (i)-(ii), without loss of generality we may assume that θ_0 is an integer.

(i) Let P_θ be the probability distribution of the largest order statistic $X_{(n)}$ and E_θ be the expectation with respect to P_θ. The family $\{P_\theta : \theta = 1, 2, ...\}$

is dominated by the counting measure and has monotone likelihood ratio in $X_{(n)}$. Therefore, a UMP test of size α is

$$T_1(X) = \begin{cases} 1 & X_{(n)} > c \\ \gamma & X_{(n)} = c \\ 0 & X_{(n)} < c, \end{cases}$$

where c is an integer and $\gamma \in [0,1]$ satisfying

$$E_{\theta_0}(T_1) = 1 - \left(\frac{c}{\theta_0}\right)^n + \gamma \frac{c^n - (c-1)^n}{\theta_0^n} = \alpha.$$

For any $\theta > \theta_0$, the power of T_1 is

$$\begin{aligned} E_\theta(T_1) &= P_\theta(X_{(n)} > c) + \gamma P_\theta(X_{(n)} = c) \\ &= 1 - \frac{c^n}{\theta^n} + \gamma \frac{c^n - (c-1)^n}{\theta^n} \\ &= 1 - (1-\alpha)\frac{\theta_0^n}{\theta^n}. \end{aligned}$$

On the other hand, for $\theta \geq \theta_0$, the power of T_* is

$$E_\theta(T_*) = P_\theta(X_{(n)} > \theta_0) + \alpha P_\theta(X_{(n)} \leq \theta_0) = 1 - \frac{\theta_0^n}{\theta^n} + \alpha \frac{\theta_0^n}{\theta^n}.$$

Hence, T_* has the same power as T_1. Since

$$\sup_{\theta \leq \theta_0} E_\theta(T_*) = \sup_{\theta \leq \theta_0} \alpha P_\theta(X_{(n)} \leq \theta_0) = \alpha P_{\theta_0}(X_{(n)} \leq \theta_0) = \alpha,$$

T_* is a UMP test of size α.
(ii) Consider $H_0 : \theta = \theta_0$ versus $H_1 : \theta > \theta_0$. The test T_1 in (i) is UMP. For $\theta > \theta_0$,

$$E_\theta(T_*) = P_\theta(X_{(n)} > \theta_0) + P_\theta(X_{(n)} \leq \theta_0 \alpha^{1/n}) = 1 - \frac{\theta_0^n}{\theta^n} + \frac{\alpha \theta_0^n}{\theta^n},$$

which is the same as the power of T_1. Now, consider hypotheses $H_0 : \theta = \theta_0$ versus $H_1 : \theta < \theta_0$. The UMP test is

$$T_2(X) = \begin{cases} 1 & X_{(n)} < d \\ \eta & X_{(n)} = d \\ 0 & X_{(n)} > d \end{cases}$$

with

$$E_{\theta_0}(T_2) = \frac{(d-1)^n}{\theta_0^n} + \eta \frac{c^n - (c-1)^n}{\theta_0^n} = \alpha.$$

For $\theta \le \theta_0$,

$$
\begin{aligned}
E_\theta(T_*) &= P_\theta(X_{(n)} > \theta_0) + P_\theta(X_{(n)} \le \theta_0 \alpha^{1/n}) \\
&= P_\theta(X_{(n)} \le \theta_0 \alpha^{1/n}) \\
&= \min\left\{1, \frac{\alpha\theta_0^n}{\theta^n}\right\}.
\end{aligned}
$$

On the other hand, the power of T_2 when $\theta \ge \theta_0\alpha^{1/n}$ is

$$
\begin{aligned}
E_\theta(T_2) &= P_\theta(X_{(n)} < d) + \eta P_\theta(X_{(n)} = d) \\
&= \frac{(d-1)^n}{\theta^n} + \eta\frac{d^n - (d-1)^n}{\theta^n} \\
&= \alpha\frac{\theta_0^n}{\theta^n}.
\end{aligned}
$$

Thus, we conclude that T_* has size α and its power is the same as the power of T_1 when $\theta > \theta_0$ and is no smaller than the power of T_2 when $\theta < \theta_0$. Thus, T_* is UMP.
(iii) The results for the uniform distribution on $(0,\theta)$ can be established similarly. Instead of providing details, we consider an alternative solution for (i)-(iii).
Solution B. (i) Let T be a test of level α. For $\theta > \theta_0$,

$$
\begin{aligned}
E_\theta(T_*) - E_\theta(T) &= E_\theta[(T_* - T)I_{\{X_{(n)} > \theta_0\}}] + E_\theta[(T_* - T)I_{\{X_{(n)} \le \theta_0\}}] \\
&= E_\theta[(1 - T)I_{\{X_{(n)} > \theta_0\}}] + E_\theta[(\alpha - T)I_{\{X_{(n)} \le \theta_0\}}] \\
&\ge E_\theta[(\alpha - T)I_{\{X_{(n)} \le \theta_0\}}] \\
&= E_{\theta_0}[(\alpha - T)I_{\{X_{(n)} \le \theta_0\}}](\theta_0/\theta)^n \\
&= E_{\theta_0}(\alpha - T)(\theta_0/\theta)^n \\
&\ge 0,
\end{aligned}
$$

where the second equality follows from the definition of T_* and the third equality follows from a scale transformation. Hence, T_* is UMP. It remains to show that the size of T_* is α, which has been shown in part (i) of Solution A.
(ii) Let T be a test of level α. For $\theta > \theta_0$,

$$
\begin{aligned}
E_\theta(T_*) - E_\theta(T) &= E_\theta[(1 - T)I_{\{X_{(n)} > \theta_0\}}] + E_\theta[(T_* - T)I_{\{X_{(n)} \le \theta_0\}}] \\
&\ge E_\theta[(T_* - T)I_{\{X_{(n)} \le \theta_0\}}] \\
&= P_\theta(X_{(n)} \le \theta_0\alpha^{1/n}) - E_\theta(TI_{\{X_{(n)} \le \theta_0\}}) \\
&= \alpha(\theta_0/\theta)^n - E_{\theta_0}(TI_{\{X_{(n)} \le \theta_0\}})(\theta_0/\theta)^n \\
&\ge 0.
\end{aligned}
$$

Similarly, for $\theta < \theta_0$,

$$E_\theta(T_*) - E_\theta(T) \geq P_\theta(X_{(n)} \leq \theta_0 \alpha^{1/n}) - E_\theta(T),$$

which is equal to $1 - E_\theta(T) \geq 0$, if $\theta \leq \theta_0 \alpha^{1/n}$, and is equal to

$$\alpha(\theta_0/\theta)^n - E_\theta(T) = \alpha(\theta_0/\theta)^n - E_{\theta_0}(T)(\theta_0/\theta)^n \geq 0$$

if $\theta_0 > \theta > \theta_0 \alpha^{1/n}$. Hence, T_* is UMP. It remains to show that the size of T_* is α, which has been shown in part (ii) of Solution A.
(iii) Note that the results for the power hold for both discrete uniform distribution and uniform distribution on $(0, \theta)$. Hence, it remains to show that T_* has size α. For T_* in (i),

$$\sup_{\theta \leq \theta_0} E_\theta(T_*) = \sup_{\theta \leq \theta_0} \alpha P_\theta(X_{(n)} \leq \theta_0) = \alpha P_{\theta_0}(X_{(n)} \leq \theta_0) = \alpha.$$

For T_* in (ii),

$$E_{\theta_0}(T_*) = P_{\theta_0}(X_{(n)} \leq \theta_0 \alpha^{1/n}) = \alpha. \quad \blacksquare$$

Exercise 15 (#6.20). Let $(X_1, ..., X_n)$ be a random sample from the exponential distribution on the interval (a, ∞) with scale parameter θ, where $a \in \mathcal{R}$ and $\theta > 0$.
(i) Derive a UMP test of size α for testing $H_0 : a = a_0$ versus $H_1 : a \neq a_0$, when θ is known.
(ii) For testing $H_0 : a = a_0$ versus $H_1 : a = a_1 < a_0$, show that any UMP test T_* of size α has power $\beta_{T_*}(a_1) = 1 - (1 - \alpha)e^{-n(a_0 - a_1)/\theta}$.
(iii) For testing $H_0 : a = a_0$ versus $H_1 : a = a_1 < a_0$, show that the power of any size α test that rejects H_0 when $Y \leq c_1$ or $Y \geq c_2$ is the same as that in part (ii), where $Y = (X_{(1)} - a_0)/\sum_{i=1}^n (X_i - X_{(1)})$ and $X_{(1)}$ is the smallest order statistic and $0 \leq c_1 < c_2$ are constants.
(iv) Derive a UMP test of size α for testing $H_0 : a = a_0$ versus $H_1 : a \neq a_0$.
(v) Derive a UMP test of size α for testing $H_0 : \theta = \theta_0, a = a_0$ versus $H_1 : \theta < \theta_0, a < a_0$.
Solution. (i) Let $Y_i = e^{-X_i/\theta}$, $i = 1, ..., n$. Then $(Y_1, ..., Y_n)$ is a random sample from the uniform distribution on $(0, e^{-a/\theta})$. Note that the hypotheses $H_0 : a = a_0$ versus $H_1 : a \neq a_0$ are the same as $H_0 : e^{-a/\theta} = e^{-a_0/\theta}$ versus $H_1 : e^{-a/\theta} \neq e^{-a_0/\theta}$. Also, the largest order statistic of $Y_1, ..., Y_n$ is equal to $e^{-X_{(1)}/\theta}$. Hence, it follows from the previous exercise that a UMP test of size α is

$$T = \begin{cases} 1 & X_{(1)} < a_0 \text{ or } X_{(1)} \geq a_0 - \frac{\theta}{n} \log \alpha \\ 0 & \text{otherwise.} \end{cases}$$

(ii) A direct calculation shows that, at $a_1 < a_0$, the power of the UMP test in part (i) of the solution is $1 - (1 - \alpha)e^{-n(a_0 - a_1)/\theta}$. Hence, for each fixed

θ, the power of T_* at a_1 can not be larger than $1-(1-\alpha)e^{-n(a_0-a_1)/\theta}$. On the other hand, in part (iii) it is shown that there are tests for testing $H_0: a = a_0$ versus $H_1: a = a_1 < a_0$ that have power $1-(1-\alpha)e^{-n(a_0-a_1)/\theta}$ at $a_1 < a_0$. Therefore, the power of T_* at a_1 can not be smaller than $1-(1-\alpha)e^{-n(a_0-a_1)/\theta}$.

(iii) Let

$$T = \begin{cases} 1 & Y \le c_1 \text{ or } Y \ge c_2 \\ 0 & \text{otherwise} \end{cases}$$

be a test of size α for testing $H_0: a = a_0$ versus $H_1: a = a_1 < a_0$. Let $Z = \sum_{i=1}^n (X_i - X_{(1)})$. By Exercise 27 in Chapter 2, Z and $X_{(1)}$ are independent. Then, the power of T at a_1 is

$$\begin{aligned} E(T) &= 1 - P(c_1 < Y < c_2) \\ &= 1 - P(a_0 + c_1 Z < X_{(1)} < a_0 + c_2 Z) \\ &= 1 - \frac{n}{\theta} E\left(\int_{a_0+c_1 Z}^{a_0+c_2 Z} e^{-n(x-a_1)/\theta} dx \right) \\ &= 1 - E\left(e^{-n(a_0-a_1+c_1 Z)/\theta} - e^{-n(a_0-a_1+c_2 Z)/\theta} \right) \\ &= 1 - e^{-n(a_0-a_1)/\theta} E\left(e^{-nc_1 Z/\theta} - e^{-nc_2 Z/\theta} \right). \end{aligned}$$

Since $2Z/\theta$ has the chi-square distribution $\chi^2_{2(n-1)}$ (Exercise 7 in Chapter 2), $b = E\left(e^{-nc_1 Z/\theta} - e^{-nc_2 Z/\theta}\right)$ does not depend on θ. Since T has size α, $E(T)$ at $a = a_0$, which is $1-b$, is equal to α. Thus, $b = 1-\alpha$ and

$$E(T) = 1 - (1-\alpha)e^{-n(a_0-a_1)/\theta}.$$

(iv) Consider the test T in (iii) with $c_1 = 0$ and $c_2 = c > 0$. From the result in (iii), T has size α and is UMP for testing $H_0: a = a_0$ versus $H_1: a < a_0$. Hence, it remains to show that T is UMP for testing $H_0: a = a_0$ versus $H_1: a > a_0$. Let $a_1 > a_0$ be fixed and θ be fixed. From the Neyman-Pearson lemma, a UMP test for $H_0: a = a_0$ versus $H_1: a = a_1$ has the rejection region

$$\left\{ \frac{e^{a_1/\theta} I_{(a_1,\infty)}(X_{(1)})}{e^{a_0/\theta} I_{(a_0,\infty)}(X_{(1)})} > c_0 \right\}$$

for some constant c_0. Since $a_1 > a_0$, this rejection region is the same as $\{Y > c\}$ for some constant c. Since the region $\{Y > c\}$ does not depend on (a, θ), T is UMP for testing $H_0: a = a_0$ versus $H_1: a > a_0$.

(v) For fixed $\theta_1 < \theta_0$ and $a_1 < a_0$, by the Neyman-Pearson lemma, the UMP test of size α for $H_0: a = a_0, \theta = \theta_0$ versus $H_1: a = a_1, \theta = \theta_1$ has the rejection region

$$R = \left\{ \frac{\theta_0^n e^{-\sum_{i=1}^n (X_i - a_1)/\theta_1} I_{(a_1,\infty)}(X_{(1)})}{\theta_1^n e^{-\sum_{i=1}^n (X_i - a_0)/\theta_0} I_{(a_0,\infty)}(X_{(1)})} > c_0 \right\}$$

for some c_0. The ratio in the previous expression is equal to ∞ when $a < X_{(1)} \leq a_0$ and

$$\frac{\theta_0^n}{\theta_1^n} e^{na_1/\theta_1} e^{-na_0/\theta_0} e^{(\theta_0^{-1}-\theta_1^{-1})\sum_{i=1}^n X_i}$$

when $X_{(1)} > a_0$. Since $\theta_0^{-1} - \theta_1^{-1} < 0$,

$$R = \left\{X_{(1)} \leq a_0\right\} \cup \left\{\sum_{i=1}^n X_i < c\right\}$$

for some constant c satisfying $P(R) = \alpha$ when $a = a_0$ and $\theta = \theta_0$. Hence, c depends on a_0 and θ_0. Since this test does not depend on (a_1, θ_1), it is UMP for testing $H_0 : \theta = \theta_0, a = a_0$ versus $H_1 : \theta < \theta_0, a < a_0$. ∎

Exercise 16 (#6.22). In Exercise 11(i) in Chapter 3, derive a UMP test of size $\alpha \in (0, 1)$ for testing $H_0 : \theta \leq \theta_0$ versus $H_1 : \theta > \theta_0$, where $\theta_0 > 1$ is known.

Solution. From Exercise 11(i) in Chapter 3, the probability density (with respect to the sum of Lebesgue measure and point mass at 1) of the sufficient and complete statistic $X_{(n)}$, the largest order statistic, is

$$f_\theta(x) = \theta^{-n} I_{\{1\}}(x) + n\theta^{-n} x^{n-1} I_{(1,\theta)}(x).$$

The family $\{f_\theta : \theta > 1\}$ has monotone likelihood ratio in $X_{(n)}$. Hence, a UMP test of size α is

$$T = \begin{cases} 1 & X_{(n)} > c \\ \gamma & X_{(n)} = c \\ 0 & X_{(n)} < c, \end{cases}$$

where c and γ are determined by the size of T. When $\theta = \theta_0$ and $1 < c \leq \theta_0$,

$$E(T) = P(X_{(n)} > c) = \frac{n}{\theta_0^n} \int_c^{\theta_0} x^{n-1} dx = 1 - \frac{c^n}{\theta_0^n}.$$

If $\theta_0 > (1 - \alpha)^{-1/n}$, then T has size α with $c = \theta_0 (1 - \alpha)^{1/n}$ and $\gamma = 0$. If $\theta_0 > (1 - \alpha)^{-1/n}$, then the size of T is

$$P(X_{(n)} > 1) + \gamma P(X_{(n)} = 1) = 1 - \frac{1}{\theta_0^n} + \frac{\gamma}{\theta_0^n}.$$

Hence, T has size α with $c = 1$ and $\gamma = 1 - (1 - \alpha)\theta_0^n$. ∎

Exercise 17 (#6.25). Let $(X_1, ..., X_n)$ be a random sample from $N(\theta, 1)$. Show that $T = I_{(-c,c)}(\bar{X})$ is a UMP test of size $\alpha \in (0, \frac{1}{2})$ for testing

$H_0 : |\theta| \geq \theta_1$ versus $H_1 : |\theta| < \theta_1$, where $\bar{X}$ is the sample mean and $\theta_1 > 0$ is a constant. Provide a formula for determining c.

Solution. From Theorem 6.3 in Shao (2003), the UMP test is of the form $T = I_{(c_1,c_2)}(\bar{X})$, where c_1 and c_2 satisfy

$$\Phi\big(\sqrt{n}(c_2 + \theta_1)\big) - \Phi\big(\sqrt{n}(c_1 + \theta_1)\big) = \alpha,$$

$$\Phi\big(\sqrt{n}(c_2 - \theta_1)\big) - \Phi\big(\sqrt{n}(c_1 - \theta_1)\big) = \alpha,$$

and Φ is the cumulative distribution function of $N(0, 1)$. Let $Y_i = -X_i$, $i = 1, ..., n$. Then $(Y_1, ..., Y_n)$ is a random sample from $N(-\theta, 1)$. Since the hypotheses are not changed with θ replaced by $-\theta$, the UMP test for testing the same hypotheses but based on Y_i's is $T_1 = I_{(c_1,c_2)}(-\bar{X})$ with the same c_1 and c_2. By the uniqueness of the UMP test, $T = T_1$ and, thus, $c_1 = -c$ and $c_2 = c > 0$. The constraints on c_i's reduce to

$$\Phi\big(\sqrt{n}(\theta_1 + c)\big) - \Phi\big(\sqrt{n}(\theta_1 - c)\big) = \alpha. \ \blacksquare$$

Exercise 18 (#6.29). Consider Exercise 12 with $H_0 : \theta \in [\theta_1, \theta_2]$ versus $H_1 : \theta \notin [\theta_1, \theta_2]$, where $0 < \theta_1 \leq \theta_2 < 1$ are constants.
(i) Show that a UMP test does not exist.
(ii) Obtain a UMPU (uniformly most power unbiased) test of size α.

Solution. (i) Let $\beta_T(\theta)$ be the power function of a test T. For any test T of level α such that $\beta_T(\theta)$ is not constant, either $\beta_T(0)$ or $\beta_T(1)$ is strictly less than α. Without loss of generality, assume that $\beta_T(0) < \alpha$. This means that at $\theta = 0$, which is one of parameter values under H_1, the power of T is smaller than $T_* \equiv \alpha$. Hence, any T with nonconstant power function can not be UMP. From Exercise 12, the UMP test of size α for testing $H_0 : \theta \leq \theta_1$ versus $H_1 : \theta > \theta_1$ clearly has power larger than α at $\theta = 1$. Hence, $T_* \equiv \alpha$ is not UMP. Therefore, a UMP test does not exists.
(ii) If a test T of level α has a nonconstant power function, then either $\beta_T(0)$ or $\beta_T(1)$ is strictly less than α and, hence, T is not unbiased. Therefore, only tests with constant power functions may be unbiased. This implies that $T_* \equiv \alpha$ is a UMPU test of size α. ∎

Exercise 19. Let X be a random variable with probability density f_θ. Assume that $\{f_\theta : \theta \in \Theta\}$ has monotone likelihood ratio in X, where $\Theta \subset \mathcal{R}$. Suppose that for each $\theta_0 \in \Theta$, a UMPU test of size α for testing $H_0 : \theta = \theta_0$ has the acceptance region $\{c_1(\theta_0) \leq X \leq c_2(\theta_0)\}$ and is strictly unbiased (i.e., its power is larger than α when $\theta \neq \theta_0$). Show that the functions $c_1(\theta)$ and $c_2(\theta)$ are increasing in θ.

Solution. Let $\theta_0 < \theta_1$ be two values in Θ and T_0 and T_1 be the UMPU tests with acceptance regions $\{c_1(\theta_0) \leq X \leq c_2(\theta_0)\}$ and $\{c_1(\theta_1) \leq X \leq c_2(\theta_1)\}$, respectively. Let $\psi(X) = T_1(X) - T_0(X)$ and E_θ be the expectation with respect to f_θ. It follows from the strict unbiasedness of the tests that

$$E_{\theta_0}\psi(X) = E_{\theta_0}(T_1) - \alpha > 0 > \alpha - E_{\theta_1} = E_{\theta_1}\psi(X).$$

If $[c_1(\theta_0), c_2(\theta_0)] \subset [c_1(\theta_1), c_2(\theta_1)]$, then $\psi(X) \leq 0$ and $E_{\theta_0}\psi(X) \leq 0$, which is impossible. If $[c_1(\theta_1), c_2(\theta_1)] \subset [c_1(\theta_0), c_2(\theta_0)]$, then $\psi(X) \geq 0$ and $E_{\theta_1}\psi(X) \geq 0$, which is impossible. Hence, neither of the two intervals contain the other. If $c_1(\theta_1) \leq c_1(\theta_0) \leq c_2(\theta_1) \leq c_2(\theta_0)$, then there is a $x_0 \in [c_1(\theta_0), c_2(\theta_1)]$ such that $\psi(X) \geq 0$ if $X < x_0$ and $\psi(X) \leq 0$ if $X \geq x_0$, i.e., the function ψ has a single change of sign. Since the family has monotone likelihood ratio in X, it follows from Lemma 6.4(i) in Shao (2003) that there is a θ_* such that $E_\theta\psi(X) \leq 0$ for $\theta < \theta_*$ and $E_\theta\psi(X) \geq 0$ for $\theta > \theta_*$. But this contradicts to the fact that $E_{\theta_0}\psi(X) > 0 > E_{\theta_1}\psi(X)$ and $\theta_0 < \theta_1$. Therefore, we must have $c_1(\theta_1) > c_1(\theta_0)$ and $c_2(\theta_1) > c_2(\theta_0)$, i.e., both $c_1(\theta)$ and $c_2(\theta)$ are increasing in θ. ∎

Exercise 20 (#6.34). Let X be a random variable from the geometric distribution with mean p^{-1}. Find a UMPU test of size α for $H_0 : p = p_0$ versus $H_1 : p \neq p_0$, where $p_0 \in (0, 1)$ is known.

Solution. The probability density of X with respect to the counting measure is

$$f(x) = \exp\left\{x \log(1-p) + \log\frac{p}{1-p}\right\} I_{\{1,2,...\}}(X),$$

which is in an exponential family. Applying Theorem 6.4 in Shao (2003), we conclude that the UMPU test of size α is

$$T_* = \begin{cases} 1 & X < c_1 \text{ or } X > c_2 \\ \gamma_i & X = c_i \quad i = 1, 2 \\ 0 & \text{otherwise,} \end{cases}$$

where c_i's are positive integers and c_i's and γ_i's are uniquely determined by

$$\frac{\alpha}{p_0} = \sum_{k=1}^{c_1-1}(1-p_0)^{k-1} + \sum_{k=c_2+1}^{\infty}(1-p_0)^{k-1} + \sum_{i=1,2}\gamma_i(1-p_0)^{c_i-1}$$

and

$$\frac{\alpha}{p_0^2} = \sum_{k=1}^{c_1-1}k(1-p_0)^{k-1} + \sum_{k=c_2+1}^{\infty}k(1-p_0)^{k-1} + \sum_{i=1,2}\gamma_i c_i(1-p_0)^{c_i-1}. \quad ∎$$

Exercise 21 (#6.36). Let $X = (X_1, ..., X_n)$ be a random sample from $N(\mu, \sigma^2)$ with unknown μ and σ^2.

(i) Show that the power of the one-sample t-test depends on a noncentral t-distribution.
(ii) Show that the power of the one-sample t-test is an increasing function of $(\mu - \mu_0)/\sigma$ for testing $H_0 : \mu \leq \mu_0$ versus $H_1 : \mu > \mu_0$ and of $|\mu - \mu_0|/\sigma$ for testing $H_0 : \mu = \mu_0$ versus $H_1 : \mu \neq \mu_0$, where μ_0 is a known constant.
Note. For testing $H_0 : \mu \leq \mu_0$ versus $H_1 : \mu > \mu_0$, the one-sample t-test of size α rejects H_0 if and only if $t(X) > t_{n-1,\alpha}$, where $t(X) = \sqrt{n}(\bar{X} - \mu_0)/S$, $\bar{X}$ is the sample mean, S^2 is the sample variance, and $t_{r,\alpha}$ is the $(1-\alpha)$th quantile of the t-distribution t_r. For testing $H_0 : \mu = \mu_0$ versus $H_1 : \mu \neq \mu_0$, the one-sample t-test of size α rejects H_0 if and only if $|t(X)| > t_{n-1,\alpha/2}$.
Solution. (i) Let $Z = \sqrt{n}(\bar{X} - \mu_0)/\sigma$, $U = S/\sigma$, and $\delta = \sqrt{n}(\mu - \mu_0)/\sigma$. Then Z is distributed as $N(\delta, 1)$, $(n-1)U^2$ has the chi-square distribution χ^2_{n-1}, and Z and U are independent. By definition, $t(X) = Z/U$ has the noncentral t-distribution $t_{n-1}(\delta)$ with the noncentrality parameter δ.
(ii) For testing $H_0 : \mu \leq \mu_0$ versus $H_1 : \mu > \mu_0$, the power of the one-sample t-test is

$$P\big(t(X) > t_{n-1,\alpha}\big) = P\big(Z > t_{n-1,\alpha}U\big) = E\left[\Phi(\delta - t_{n-1,\alpha}U)\right],$$

where Φ is the cumulative distribution function of $N(0,1)$. Since Φ is an increasing function, the power is an increasing function of δ.

For testing $H_0 : \mu = \mu_0$ versus $H_1 : \mu \neq \mu_0$, the power of the one-sample t-test is

$$\begin{aligned} P\big(|t(X)| > t_{n-1,\alpha/2}\big) &= P\big(Z > t_{n-1,\alpha/2}U\big) + P\big(Z < -t_{n-1,\alpha/2}U\big) \\ &= E\left[\Phi(\delta - t_{n-1,\alpha/2}U) + \Phi(-\delta - t_{n-1,\alpha/2}U)\right] \\ &= E\left[\Phi(|\delta| - t_{n-1,\alpha/2}U) + \Phi(-|\delta| - t_{n-1,\alpha/2}U)\right]. \end{aligned}$$

To show that the power is an increasing function of $|\delta|$, it suffices to show that $\Phi(x-a) + \Phi(-x-a)$ is increasing in $x > 0$ for any fixed $a > 0$. The result follows from

$$\begin{aligned} \frac{d}{dx}[\Phi(x-a) + \Phi(-x-a)] &= \frac{e^{-(x-a)^2/2} - e^{-(x+a)^2/2}}{2\pi} \\ &= \frac{e^{-(x^2+a^2)/2}(e^{ax} - e^{-ax})}{2\pi} \\ &> 0. \quad \blacksquare \end{aligned}$$

Exercise 22. Let $X = (X_1, ..., X_n)$ be a random sample from $N(\mu, \sigma^2)$ with unknown μ and σ^2 and $t(X) = \sqrt{n}\bar{X}/S$, where $\bar{X}$ is the sample mean and S^2 is the sample variance. For testing $H_0 : \mu/\sigma \leq \theta_0$ versus $H_1 : \mu/\sigma > \theta_0$, find a test of size α that is UMP among all tests based on $t(X)$, where θ_0 is a known constant.

Solution. From the previous exercise, we know that $t(X)$ has the noncentral t-distribution $t_{n-1}(\delta)$, where $\delta = \sqrt{n}\mu/\sigma$ is the noncentrality parameter. The Lebesgue density of $t(X)$ is (e.g., Shao, 2003, p. 26)

$$f_\delta(t) = \int_0^\infty g_\delta(t,y)dy,$$

where

$$g_\delta(t,y) = \frac{y^{(n-2)/2}e^{-\{[t/\sqrt{y/(n-1)}-\delta]^2+y\}/2}}{2^{n/2}\Gamma(\frac{n-1}{2})\sqrt{\pi(n-1)}}.$$

We now show that the family $\{f_\delta(t) : \delta \in \mathcal{R}\}$ has monotone likelihood ratio in t. For $\delta_1 < \delta_2$, it suffices to show that

$$\frac{d}{dt}\frac{f_{\delta_2}(t)}{f_{\delta_1}(t)} = \frac{f'_{\delta_2}(t)f_{\delta_1}(t) - f_{\delta_2}(t)f'_{\delta_1}(t)}{[f_{\delta_1}(t)]^2} \geq 0, \quad t \in \mathcal{R}.$$

Since

$$f'_\delta(t) = \int_0^\infty [\delta - t/\sqrt{y/(n-1)}]g_\delta(t,y)dy = \delta f_\delta(t) - t\tilde{f}_\delta(t),$$

where

$$\tilde{f}_\delta(t) = \int_0^\infty \sqrt{\frac{y}{n-1}}g_\delta(t,y)dy,$$

we obtain that

$$f'_{\delta_2}(t)f_{\delta_1}(t) - f_{\delta_2}(t)f'_{\delta_1}(t) = f_{\delta_1}(t)f_{\delta_2}(t)\left[\delta_2 - \delta_1 + t\left(\frac{\tilde{f}_{\delta_1}(t)}{f_{\delta_1}(t)} - \frac{\tilde{f}_{\delta_2}(t)}{f_{\delta_2}(t)}\right)\right].$$

For any fixed $t \in \mathcal{R}$, the family of densities $\left\{\frac{g_\delta(t,y)}{f_\delta(t)} : \delta \in \mathcal{R}\right\}$ is an exponential family having monotone likelihood ratio in $t/\sqrt{y}$. Hence, by Lemma 6.3 in Shao (2003), the integral $\tilde{f}_\delta(t)$ is nonincreasing in δ when $t > 0$ and is nondecreasing in δ when $t < 0$. Hence, for $t > 0$ and $t < 0$,

$$t\left(\frac{\tilde{f}_{\delta_1}(t)}{f_{\delta_1}(t)} - \frac{\tilde{f}_{\delta_2}(t)}{f_{\delta_2}(t)}\right) \geq 0$$

and, therefore, $f'_{\delta_2}(t)f_{\delta_1}(t) - f_{\delta_2}(t)f'_{\delta_1}(t) \geq 0$. Consequently, for testing $H_0 : \mu/\sigma \leq \theta_0$ versus $H_1 : \mu/\sigma > \theta_0$, a test of size α that is UMP among all tests based on $t(X)$ rejects H_0 when $t(X) > c$, where c is the $(1-\alpha)$th quantile of the noncentral t-distribution $t_{n-1}(\sqrt{n}\theta_0)$. ∎

Exercise 23 (#6.37). Let $(X_1, ..., X_n)$ be a random sample from the gamma distribution with unknown shape parameter θ and unknown scale

parameter γ. Let $\theta_0 > 0$ and $\gamma_0 > 0$ be known constants.
(i) For testing $H_0 : \theta \leq \theta_0$ versus $H_1 : \theta > \theta_0$ and $H_0 : \theta = \theta_0$ versus $H_1 : \theta \neq \theta_0$, show that there exist UMPU tests whose rejection regions are based on $V = \prod_{i=1}^n (X_i/\bar{X})$, where $\bar{X}$ is the sample mean.
(ii) For testing $H_0 : \gamma \leq \gamma_0$ versus $H_1 : \gamma > \gamma_0$, show that a UMPU test rejects H_0 when $\sum_{i=1}^n X_i > C(\prod_{i=1}^n X_i)$ for some function C.
Solution. (i) Let $Y = \log(\prod_{i=1}^n X_i)$ and $U = n\bar{X}$. The joint density of $(X_1, ..., X_n)$ can be written as $\left[\Gamma(\theta)\gamma^\theta\right]^{-n} e^{\theta Y - U/\gamma - Y}$, which belongs to an exponential family. From Theorem 6.4 in Shao (2003), UMPU tests are functions of Y and U. By Basu's theorem, $V_1 = Y - n\log(U/n) = \log\left(\prod_{i=1}^n (X_i/\bar{X})\right)$ satisfies the conditions in Lemma 6.7 of Shao (2003). Hence, the rejection regions of the UMPU tests can be determined by using V_1. Since $V = e^{V_1}$, the rejection regions of the UMPU tests can also be determined by using V.
(ii) Let $U = \log(\prod_{i=1}^n X_i)$ and $Y = n\bar{X}$. The joint density of $(X_1, ..., X_n)$ can be written as $\left[\Gamma(\theta)\gamma^\theta\right]^{-n} e^{-\gamma^{-1}Y + \theta U - U}$, From Theorem 6.4 in Shao (2003), for testing $H_0 : \gamma \leq \gamma_0$ versus $H_1 : \gamma > \gamma_0$, the UMPU test is

$$T^* = \begin{cases} 1 & Y > C_1(U) \\ 0 & Y \leq C_1(U), \end{cases}$$

where C_1 is a function such that $E(T^*|U) = \alpha$ when $\gamma = \gamma_0$. The result follows by letting $C(x) = C_1(\log x)$. ∎

Exercise 24 (#6.39). Let X_1 and X_2 be independent observations from the binomial distributions with sizes n_1 and n_2 and probabilities p_1 and p_2, respectively, where n_i's are known and p_i's are unknown.
(i) Let $Y = X_2$ and $U = X_1 + X_2$. Show that

$$P(Y = y|U = u) = K_u(\theta)\binom{n_1}{u-y}\binom{n_2}{y} e^{\theta y} I_A(y), \qquad u = 0, 1, ..., n_1 + n_2,$$

where $A = \{y : y = 0, 1, ..., \min\{u, n_2\}, u - y \leq n_1\}$, $\theta = \log \frac{p_2(1-p_1)}{p_1(1-p_2)}$, and

$$K_u(\theta) = \left[\sum_{y \in A} \binom{n_1}{u-y}\binom{n_2}{y} e^{\theta y}\right]^{-1}.$$

(ii) Find a UMPU test of size α for testing $H_0 : p_1 \geq p_2$ versus $H_1 : p_1 < p_2$.
(iii) Repeat (ii) for $H_0 : p_1 = p_2$ versus $H_1 : p_1 \neq p_2$.
Solution. (i) When $u = 0, 1, ..., n_1 + n_2$ and $y \in A$,

$$P(Y = y, U = u) = \binom{n_1}{u-y}\binom{n_2}{y} p_1^{u-y}(1-p_1)^{n_1-u+y} p_2^y (1-p_2)^{n_2-y}$$

and

$$P(U = u) = \sum_{y \in A} \binom{n_1}{u-y} \binom{n_2}{y} p_1^{u-y} (1-p_1)^{n_1-u+y} p_2^y (1-p_2)^{n_2-y}.$$

Then, when $y \in A$,

$$P(Y = y|U = u) = \frac{P(Y = y, U = u)}{P(U = u)} = \binom{n_1}{u-y} \binom{n_2}{y} e^{\theta y} K_u(\theta).$$

(ii) Since $\theta = \log \frac{p_2(1-p_1)}{p_1(1-p_2)}$, the testing problem is equivalent to testing $H_0 : \theta \leq 0$ versus $H_1 : \theta > 0$. By Theorem 6.4 in Shao (2003), the UMPU test is

$$T_*(Y, U) = \begin{cases} 1 & Y > C(U) \\ \gamma(U) & Y = C(U) \\ 0 & Y < C(U), \end{cases}$$

where C and γ are functions of U such that $E(T_*|U) = \alpha$ when $\theta = 0$ ($p_1 = p_2$), which can be determined using the conditional distribution of Y given U. When $\theta = 0$, this conditional distribution is, by the result in (i),

$$P(Y = y|U = u) = \binom{n_1+n_2}{u}^{-1} \binom{n_1}{u-y} \binom{n_2}{y} I_A(y), \quad u = 0, 1, ..., n_1+n_2.$$

(iii) The testing problem is equivalent to testing $H_0 : \theta = 0$ versus $H_1 : \theta \neq 0$. Thus, the UMPU test is

$$T_* = \begin{cases} 1 & Y > C_1(U) \text{ or } Y < C_2(U) \\ \gamma_i(U) & Y = C_i(U) \quad i = 1, 2 \\ 0 & C_1(U) < Y < C_2(U), \end{cases}$$

where C_i's and γ_i's are functions such that $E(T_*|U) = \alpha$ and $E(T_* Y|U) = \alpha E(Y|U)$ when $\theta = 0$, which can be determined using the conditional distribution of Y given U in part (ii) of the solution. ∎

Exercise 25 (#6.40). Let X_1 and X_2 be independently distributed as the negative binomial distributions with sizes n_1 and n_2 and probabilities p_1 and p_2, respectively, where n_i's are known and p_i's are unknown.
(i) Show that there exists a UMPU test of size α for testing $H_0 : p_1 \leq p_2$ versus $H_1 : p_1 > p_2$.
(ii) Let $Y = X_1$ and $U = X_1 + X_2$. Determine the conditional distribution of Y given U when $n_1 = n_2 = 1$.
Solution. (i) The joint probability density of X_1 and X_2 is

$$\frac{\binom{x_1-1}{n_1-1} \binom{x_2-1}{n_2-1} p_1^{n_1} p_2^{n_2}}{(1-p_1)^{n_1} (1-p_2)^{n_2}} e^{\theta Y + U \log(1-p_2)}$$

where $\theta = \log(\frac{1-p_1}{1-p_2})$, $Y = X_1$, and $U = X_1 + X_2$. The testing problem is equivalent to testing $H_0 : \theta \geq 0$ versus $H_1 : \theta < 0$. By Theorem 6.4 in Shao (2003), the UMPU test is

$$T_*(Y, U) = \begin{cases} 1 & Y < C(U) \\ \gamma(U) & Y = C(U) \\ 0 & Y > C(U), \end{cases}$$

where $C(U)$ and $\gamma(U)$ satisfy $E(T_*|U) = \alpha$ when $\theta = 0$.
(ii) When $n_1 = n_2 = 1$,

$$\begin{aligned} P(U = u) &= \sum_{k=1}^{u-1} P(X_1 = k, X_2 = u - k) \\ &= \frac{p_1 p_2 (1-p_2)^{u-1}}{1-p_1} \sum_{k=1}^{u-1} \left(\frac{1-p_1}{1-p_2}\right)^k \\ &= \frac{p_1 p_2 (1-p_2)^{u-1}}{1-p_1} \sum_{k=1}^{u-1} e^{\theta k} \end{aligned}$$

for $u = 2, 3, ...$, and

$$P(Y = y, U = u) = (1-p_1)^{y-1} p_1 (1-p_2)^{u-y-1} p_2$$

for $y = 1, ..., u-1$, $u = 2, 3, ...$ Hence

$$P(Y = y|U = u) = \frac{(1-p_1)^{y-1} p_1 (1-p_2)^{u-y-1} p_2}{\frac{p_1 p_2 (1-p_2)^{u-1}}{1-p_1} \sum_{k=1}^{u-1} e^{\theta k}} = \frac{e^{\theta y}}{\sum_{k=1}^{u-1} e^{\theta k}}$$

for $y = 1, ..., u-1$, $u = 2, 3, ...$ When $\theta = 0$, this conditional distribution is the discrete uniform distribution on $\{1, ..., u-1\}$. ∎

Exercise 26 (#6.44). Let X_j, $j = 1, 2, 3$, be independent from the Poisson distributions with means λ_j, $j = 1, 2, 3$, respectively. Show that there exists a UMPU test of size α for testing H_0: $\lambda_1 \lambda_2 \leq \lambda_3^2$ versus H_1: $\lambda_1 \lambda_2 > \lambda_3^2$.
Solution. The joint probability density for (X_1, X_2, X_3) is

$$\frac{e^{-(\lambda_1+\lambda_2+\lambda_3)}}{X_1! X_2! X_3!} e^{X_1 \log \lambda_1 + X_2 \log \lambda_2 + X_3 \log \lambda_3},$$

which is the same as

$$\frac{e^{-(\lambda_1+\lambda_2+\lambda_3)}}{X_1! X_2! X_3!} e^{\theta Y + U_1 \log \lambda_2 + U_2 \log \lambda_3},$$

where $\theta = \log \lambda_1 + \log \lambda_2 - 2 \log \lambda_3$, $Y = X_1$, $U_1 = X_2 - X_1$, and $U_2 = X_3 + 2X_1$. By Theorem 6.4 in Shao (2003), there exists a UMPU test of size

α for testing $H_0 : \theta \leq 0$ versus $H_1 : \theta > 0$, which is equivalent to testing $H_0 : \lambda_1\lambda_2 \leq \lambda_3^2$ versus $H_1 : \lambda_1\lambda_2 > \lambda_3^2$. ∎

Exercise 27 (#6.49). Let $(X_{i1}, ..., X_{in_i})$, $i = 1, 2$, be two independent random samples from $N(\mu_i, \sigma^2)$, respectively, where $n_i \geq 2$ and μ_i's and σ are unknown. Show that a UMPU test of size α for $H_0 : \mu_1 = \mu_2$ versus $H_1 : \mu_1 \neq \mu_2$ rejects H_0 when $|t(X)| > t_{n_1+n_2-1,\alpha/2}$, where

$$t(X) = \frac{(\bar{X}_2 - \bar{X}_1)/\sqrt{n_1^{-1} + n_2^{-1}}}{\sqrt{[(n_1-1)S_1^2 + (n_2-1)S_2^2]/(n_1+n_2-2)}},$$

$\bar{X}_i$ and S_i^2 are the sample mean and variance based on $X_{i1}, ..., X_{in_i}$, $i = 1, 2$, and $t_{n_1+n_2-1,\alpha}$ is the $(1-\alpha)$th quantile of the t-distribution $t_{n_1+n_2-1}$. Derive the power function of this test.

Solution. Let $Y = \bar{X}_2 - \bar{X}_1$, $U_1 = n_1\bar{X}_1 + n_2\bar{X}_2$, $U_2 = \sum_{i=1}^2 \sum_{j=1}^n X_{ij}^2$, $\theta = (\mu_1 - \mu_2)/[(n_1^{-1} + n_2^{-1})\sigma^2]$, $\varphi_1 = (n_1\mu_1 + n_2\mu_2)/[(n_1+n_2)\sigma^2]$, and $\varphi_2 = -(2\sigma^2)^{-1}$. Then, the joint density of $X_{i1}, ..., X_{in_i}$, $i = 1, 2$, can be written as

$$(\sqrt{2\pi}\sigma)^{n_1+n_2} e^{\theta Y + \varphi_1 U_1 + \varphi_2 U_2}.$$

The statistic $V = Y/\sqrt{U_2 - U_1^2/(n_1+n_2)}$ satisfies the conditions in Lemma 6.7(ii) in Shao (2003). Hence, the UMPU test has the rejection region $V < c_1$ or $V > c_2$. Under H_0, V is symmetrically distributed around 0, i.e., V and $-V$ have the same distribution. Thus, a UMPU test rejects H_0 when $-V < c_1$ or $-V > c_2$, which is the same as rejecting H_0 when $V < -c_2$ or $V > -c_1$. By the uniqueness of the UMPU test, we conclude that $c_1 = -c_2$, i.e., the UMPU test rejects when $|V| > c$. Since

$$U_2 - \frac{U_1^2}{n_1+n_2} = (n_1-1)S_1^2 + (n_2-1)S_2^2 + \frac{n_1 n_2 Y^2}{n_1+n_2},$$

we obtain that

$$\frac{1}{V^2} = \frac{n_1 n_2}{(n_1+n_2)(n_1+n_2-2)} \frac{1}{[t(X)]^2} + \frac{n_1 n_2}{n_1+n_2}.$$

Hence, $|V|$ is an increasing function of $|t(X)|$. Also, $t(X)$ has the t-distribution with $t_{n_1+n_2-2}$ under H_0. Thus, the UMPU test rejects H_0 when $|t(X)| > t_{n_1+n_2-2,\alpha/2}$. Under H_1, $t(X)$ is distributed as the noncentral t-distribution $t_{n_1+n_2-2}(\delta)$ with noncentrality parameter

$$\delta = \frac{\mu_2 - \mu_1}{\sigma\sqrt{n_1^{-1} + n_2^{-1}}}.$$

Thus the power function of the UMPU test is

$$1 - G_\delta(t_{n_1+n_2-2,\alpha/2}) + G_\delta(-t_{n_1+n_2-2,\alpha/2}),$$

where G_δ denotes the cumulative distribution function of the noncentral t-distribution $t_{n_1+n_2-2}(\delta)$. ∎

Exercise 28 (#6.50). Let $(X_{i1}, ..., X_{in})$, $i = 1, 2$, be two independent random samples from $N(\mu_i, \sigma_i^2)$, respectively, where $n > 1$ and μ_i's and σ_i's are unknown. Show that a UMPU test of size α for testing $H_0 : \sigma_2^2 = \Delta_0\sigma_1^2$ versus $H_1 : \sigma_2^2 \neq \Delta_0\sigma_1^2$ rejects H_0 when

$$\max\left\{\frac{S_2^2}{\Delta_0 S_1^2}, \frac{\Delta_0 S_1^2}{S_2^2}\right\} > \frac{1-c}{c},$$

where $\Delta_0 > 0$ is a constant, $\int_0^c f_{(n-1)/2,(n-1)/2}(v)dv = \alpha/2$ and $f_{a,b}$ is the Lebesgue density of the beta distribution with parameter (a, b).

Solution. From Shao (2003, p. 413), the UMPU test rejects H_0 if $V < c_1$ or $V > c_2$, where

$$V = \frac{S_2^2/\Delta_0}{S_1^2 + S_2^2/\Delta_0}$$

and S_i^2 is the sample variance based on $X_{i1}, ..., X_{in}$. Under H_0 ($\sigma_1^2 = \sigma_2^2$), V has the beta distribution with parameter $(\frac{n-1}{2}, \frac{n-1}{2})$, which is symmetric about $\frac{1}{2}$, i.e., V has the same distribution as $1 - V$. Thus, a UMPU test rejects H_0 when $1 - V < c_1$ or $1 - V > c_2$, which is the same as rejecting H_0 when $V < 1 - c_2$ or $V > 1 - c_1$. By the uniqueness of the UMPU test, we conclude that $c_1 + c_2 = 1$. Let $c_1 = c$. Then the UMPU test rejects H_0 when $V < c$ or $V > 1-c$ and c satisfies $\int_0^c f_{(n-1)/2,(n-1)/2}(v)dv = \alpha/2$. Let $F = S_2^2/(\Delta_0 S_1^2)$. Then $V = F/(1+F)$, $V < c$ if and only if $F^{-1} > (1-c)/c$, and $V > 1 - c$ if and only if $F > (1 - c)/c$. Hence, the UMPU test rejects when $\max\{F, F^{-1}\} > (1 - c)/c$, which is the desired result. ∎

Exercise 29 (#6.51). Suppose that $X_i = \beta_0 + \beta_1 t_i + \varepsilon_i$, $i = 1, ..., n$, where t_i's are fixed constants that are not all the same, ε_i's are independent and identically distributed as $N(0, \sigma^2)$, and β_0, β_1, and σ^2 are unknown parameters. Derive a UMPU test of size α for testing

(i) $H_0 : \beta_0 \leq \theta_0$ versus $H_1 : \beta_0 > \theta_0$;
(ii) $H_0 : \beta_0 = \theta_0$ versus $H_1 : \beta_0 \neq \theta_0$;
(iii) $H_0 : \beta_1 \leq \theta_0$ versus $H_1 : \beta_1 > \theta_0$;
(iv) $H_0 : \beta_1 = \theta_0$ versus $H_1 : \beta_1 \neq \theta_0$.

Solution: Note that $(X_1, ..., X_n)$ follows a simple linear regression model. Let $D = n\sum_{i=1}^n t_i^2 - \left(\sum_{i=1}^n t_i\right)^2$,

$$\hat{\beta}_0 = \frac{1}{D}\left(\sum_{i=1}^n t_i^2 \sum_{i=1}^n X_i - \sum_{i=1}^n t_i \sum_{i=1}^n t_i X_i\right)$$

and

$$\hat{\beta}_1 = \frac{1}{D}\left(n\sum_{i=1}^n t_i X_i - \sum_{i=1}^n t_i \sum_{i=1}^n X_i\right)$$

be the least squares estimators, and let

$$\hat{\sigma}^2 = \frac{1}{n-2}\sum_{i=1}^{n}(y_i - \hat{\beta}_0\hat{\beta}_1 t_i)^2.$$

From the theory of linear models, $\hat{\beta}_0$ has the normal distribution with mean β_0 and variance $\sigma^2\sum_{i=1}^n t_i^2/D$, $\hat{\beta}_1$ has the normal distribution with mean β_1 and variance $\sigma^2 n/D$, $(n-2)\hat{\sigma}^2/\sigma^2$ has the chi-square distribution χ^2_{n-2}, and $(\hat{\beta}_0, \hat{\beta}_1)$ and $\hat{\sigma}^2$ are independent. Thus, the following results follow from the example in Shao (2003, p. 416).
(i) The UMPU test of size α for testing $H_0 : \beta_0 \leq \theta_0$ versus $H_1 : \beta_0 > \theta_0$ rejects H_0 when $t_0 > t_{n-2,\alpha}$, where

$$t_0 = \frac{\sqrt{D}(\hat{\beta}_0 - \theta_0)}{\hat{\sigma}\sqrt{\sum_{i=1}^n t_i^2}}$$

and $t_{r,\alpha}$ is the $(1-\alpha)$th quantile of the t-distribution t_r.
(ii) The UMPU test of size α for testing $H_0 : \beta_0 = \theta_0$ versus $H_1 : \beta_0 \neq \theta_0$ rejects H_0 when $|t_0| > t_{n-2,\alpha/2}$.
(iii) The UMPU test of size α for testing $H_0 : \beta_1 \leq \theta_0$ versus $H_1 : \beta_1 > \theta_0$ rejects H_0 when $t_1 > t_{n-2,\alpha}$, where

$$t_1 = \frac{\sqrt{D}(\hat{\beta}_1 - \theta_0)}{\sqrt{n}\hat{\sigma}}.$$

(iv) The UMPU test of size α for testing $H_0 : \beta_1 = \theta_0$ versus $H_1 : \beta_1 \neq \theta_0$ rejects H_0 when $|t_1| > t_{n-2,\alpha/2}$. ∎

Exercise 30 (#6.53). Let X be a sample from $N_n(Z\beta, \sigma^2 I_n)$, where $\beta \in \mathcal{R}^p$ and $\sigma^2 > 0$ are unknown and Z is an $n \times p$ known matrix of rank $r \leq p < n$. For testing $H_0 : \sigma^2 \leq \sigma_0^2$ versus $H_1 : \sigma^2 > \sigma_0^2$ and $H_0 : \sigma^2 = \sigma_0^2$ versus $H_1 : \sigma^2 \neq \sigma_0^2$, show that UMPU tests of size α are functions of $\text{SSR} = \|X - Z\hat{\beta}\|^2$, where $\hat{\beta}$ is the least squares estimator of β, and their rejection regions can be determined using chi-square distributions.
Solution. Since $H = Z(Z^\tau Z)^- Z^\tau$ is a projection matrix of rank r, there exists an $n \times n$ orthogonal matrix Γ such that

$$\Gamma = (\ \Gamma_1 \ \ \Gamma_2\ ) \qquad \text{and} \qquad H\Gamma = (\ \Gamma_1 \ \ 0\ ),$$

where Γ_1 is $n\times r$ and Γ_2 is $n \times (n-r)$. Let $Y_j = \Gamma_j^\tau X$, $j = 1, 2$. Consider the transformation $(Y_1, Y_2) = \Gamma^\tau X$. Since $\Gamma^\tau\Gamma = I_n$, (Y_1, Y_2) has distribution $N_n(\Gamma^\tau Z\beta, \sigma^2 I_n)$. Note that

$$E(Y_2) = E(\Gamma_2^\tau X) = \Gamma_2^\tau Z\beta = \Gamma_2^\tau HZ\beta = 0.$$

Let $\eta = \Gamma_1^\tau Z\beta = E(Y_1)$. Then the density of (Y_1, Y_2) is

$$\frac{1}{(2\pi\sigma^2)^{n/2}} \exp\left\{ \frac{\|Y_1\|^2 + \|Y_2\|^2}{2\sigma^2} + \frac{\eta^\tau Y_1}{\sigma^2} - \frac{\|\eta\|^2}{2\sigma^2} \right\}.$$

From Theorem 6.4 in Shao (2003), the UMPU tests are based on $Y = \|Y_1\|^2 + \|Y_2\|^2$ and $U = Y_1$. Let $V = \|Y_2\|^2$. Then $V = Y - \|U\|^2$ satisfies the conditions in Lemma 6.7 of Shao (2003). Hence, the UMPU test for $H_0 : \sigma^2 \leq \sigma_0^2$ versus $H_1 : \sigma^2 > \sigma_0^2$ rejects when $V > c$ and the UMPU test for $H_0 : \sigma^2 = \sigma_0^2$ versus $H_1 : \sigma^2 \neq \sigma_0^2$, rejects when $V < c_1$ or $V > c_2$. Since

$$\|Y_1 - \eta\|^2 + \|Y_2\|^2 = \|X - Z\beta\|^2,$$

$$\min_\eta \|Y_1 - \eta\|^2 + \|Y_2\|^2 = \min_\beta \|X - Z\beta\|^2$$

and, therefore,

$$V = \|Y_2\|^2 = \|X - Z\hat{\beta}\|^2 = \text{SSR}.$$

Finally, by Theorem 3.8 in Shao (2003), SSR/σ^2 has the chi-square distribution χ^2_{n-r}. ∎

Exercise 31 (#6.54). Let $(X_1, ..., X_n)$ be a random sample from a bivariate normal distribution with unknown means μ_1 and μ_2, variances σ_1^2 and σ_2^2, and correlation coefficient ρ. Let X_{ij} be the jth component of X_i, $j = 1, 2$, $\bar{X}_j$ and S_j^2 be the sample mean and variance based on $X_{1j}, ..., X_{nj}$, and $V = \sqrt{n-2}R/\sqrt{1-R^2}$, where

$$R = \frac{1}{S_1 S_2 (n-1)} \sum_{i=1}^n (X_{i1} - \bar{X}_1)(X_{i2} - \bar{X}_2)$$

is the sample correlation coefficient. Show that the UMPU test of size α for $H_0 : \rho \leq 0$ versus $H_1 : \rho > 0$ rejects H_0 when $V > t_{n-2,\alpha}$ and the UMPU test of size α for $H_0 : \rho = 0$ versus $H_1 : \rho \neq 0$ rejects H_0 when $|V| > t_{n-2,\alpha/2}$, where $t_{n-2,\alpha}$ is the $(1-\alpha)$th quantile of the t-distribution t_{n-2}.

Solution. The Lebesgue density of $(X_1, ..., X_n)$ can be written as

$$C(\mu_1, \mu_2, \sigma_1, \sigma_2, \rho) \exp\left\{\theta Y + \varphi^\tau U\right\},$$

where $C(\cdot)$ is a function of $(\mu_1, \mu_2, \sigma_1, \sigma_2, \rho)$,

$$Y = \sum_{i=1}^n X_{i1} X_{i2}, \qquad \theta = \frac{\rho}{\sigma_1 \sigma_2 (1 - \rho^2)},$$

$$U = \left(\sum_{i=1}^n X_{i1}^2, \sum_{i=1}^n X_{i2}^2, \sum_{i=1}^n X_{i1}, \sum_{i=1}^n X_{i2} \right),$$

and

$$\varphi = \left(-\frac{1}{2\sigma_1^2(1-\rho^2)}, -\frac{1}{2\sigma_2^2(1-\rho^2)}, \frac{\mu_1}{\sigma_1^2(1-\rho^2)} - \theta\mu_1, \frac{\mu_2}{\sigma_2^2(1-\rho^2)} - \theta\mu_2\right).$$

By Basu's theorem, R is independent of U when $\rho = 0$. Also,

$$R = \frac{Y - U_3U_4/n}{\sqrt{(U_1 - U_3^2/n)(U_2 - U_4^2/n)}},$$

which is linear in Y, where U_j is the jth component of U. Hence, we may apply Theorem 6.4 and Lemma 6.7 in Shao (2003). It remains to show that V has the t-distribution t_{n-2} when $\rho = 0$, which is a consequence of the result in the note of Exercise 17 in Chapter 1 and the result in Exercise 22(ii) in Chapter 2. ∎

Exercise 32 (#6.55). Let $(X_1, ..., X_n)$ be a random sample from a bivariate normal distribution with unknown means μ_1 and μ_2, variances σ_1^2 and σ_2^2, and correlation coefficient ρ. Let X_{ij} be the jth component of X_i, $j = 1, 2$, $\bar{X}_j$ and S_j^2 be the sample mean and variance based on $X_{1j}, ..., X_{nj}$, and $S_{12} = (n-1)^{-1}\sum_{i=1}^n (X_{i1} - \bar{X}_1)(X_{i2} - \bar{X}_2)$.
(i) Let $\Delta_0 > 0$ be a known constant. Show that a UMPU test for testing $H_0 : \sigma_2/\sigma_1 = \Delta_0$ versus $H_1 : \sigma_2/\sigma_1 \neq \Delta_0$ rejects H_0 when

$$R = |\Delta_0^2 S_1^2 - S_2^2| \Big/ \sqrt{(\Delta_0^2 S_1^2 + S_2^2)^2 - 4\Delta_0^2 S_{12}^2} > c.$$

(ii) Find the Lebesgue density of R in (i) when $\sigma_2/\sigma_1 = \Delta_0$.
(iii) Assume that $\sigma_1 = \sigma_2$. Show that a UMPU test for $H_0 : \mu_1 = \mu_2$ versus $H_1 : \mu_1 \neq \mu_2$ rejects H_0 when

$$V = |\bar{X}_2 - \bar{X}_1| \Big/ \sqrt{(n-1)(S_1^2 + S_2^2 - 2S_{12})} > c.$$

(iv) Find the Lebesgue density of V in (iii) when $\mu_1 = \mu_2$.
Solution. (i) Let

$$\begin{pmatrix} Y_{i1} \\ Y_{i2} \end{pmatrix} = \begin{pmatrix} \Delta_0 & 1 \\ \Delta_0 & -1 \end{pmatrix} \begin{pmatrix} X_{i1} \\ X_{i2} \end{pmatrix}.$$

Then $\text{Cov}(Y_{i1}, Y_{i2}) = \Delta_0^2\sigma_1^2 - \sigma_2^2$. If we let ρ_Y be the correlation between Y_{i1} and Y_{i2}, then testing $H_0 : \sigma_2/\sigma_1 = \Delta_0$ versus $H_1 : \sigma_2/\sigma_1 \neq \Delta_0$ is equivalent to testing $H_0 : \rho_Y = 0$ versus $H_1 : \rho_Y \neq 0$. By the result in the previous exercise, the UMPU test rejects when $|V_Y| > c$, where V_Y is the sample correlation coefficient based on the Y-sample. The result follows from the fact that $|V_Y| = R$.

(ii) From Exercise 22(ii) in Chapter 2, the Lebesgue density of R in (i) when $\sigma_2/\sigma_1 = \Delta_0$ is

$$\frac{2\Gamma(\frac{n-1}{2})}{\sqrt{\pi}\Gamma(\frac{n-2}{2})}(1-r^2)^{\frac{n-4}{2}} I_{(0,1)}(r).$$

(iii) Let

$$\begin{pmatrix} Y_{i1} \\ Y_{i2} \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} \begin{pmatrix} X_{i1} \\ X_{i2} \end{pmatrix}.$$

Then (Y_{i1}, Y_{i2}) has the bivariate normal distribution

$$N_2\left(\begin{pmatrix} \mu_1 + \mu_2 \\ \mu_1 - \mu_2 \end{pmatrix}, \begin{pmatrix} 2(1+\rho)\sigma^2 & 0 \\ 0 & 2(1-\rho)\sigma^2 \end{pmatrix} \right).$$

Since Y_{i1} and Y_{i2} are independent, the UMPU test of size α for testing $H_0 : \mu_1 = \mu_2$ versus $H_1 : \mu_1 \neq \mu_2$ rejects H_0 when $|t(Y)| > t_{n-1,\alpha/2}$, where $t(Y) = \sqrt{n}\bar{Y}_2/S_{Y_2}$, $\bar{Y}_2$ and $S^2_{Y_2}$ are the sample mean and variance based on $Y_{12}, ..., Y_{n2}$, and $t_{n-1,\alpha}$ is the $(1-\alpha)$th quantile of the t-distribution t_{n-1}. A direct calculation shows that

$$|t(Y)| = \frac{\sqrt{n}|\bar{X}_1 - \bar{X}_2|}{\sqrt{(S_1^2 + S_2^2 - 2S_{12})}} = \sqrt{n(n-1)}V.$$

(iv) Since $t(Y)$ has the t-distribution t_{n-1} under H_0, the Lebesgue density of V when $\mu_1 = \mu_2$ is

$$\frac{\sqrt{n}\Gamma(\frac{n}{2})}{\sqrt{\pi}\Gamma(\frac{n-1}{2})}(1+nv^2)^{-n/2}. \blacksquare$$

Exercise 33 (#6.57). Let $(X_1, ..., X_n)$ be a random sample from the exponential distribution on the interval (a, ∞) with scale parameter θ, where a and θ are unknown. Let $V = 2\sum_{i=1}^n (X_i - X_{(1)})$, where $X_{(1)}$ is the smallest order statistic.
(i) For testing $H_0 : \theta = 1$ versus $H_1 : \theta \neq 1$, show that a UMPU test of size α rejects H_0 when $V < c_1$ or $V > c_2$, where c_i's are determined by

$$\int_{c_1}^{c_2} f_{2n-2}(v)dv = \int_{c_1}^{c_2} f_{2n}(v)dv = 1 - \alpha$$

and $f_m(v)$ is the Lebesgue density of the chi-square distribution χ^2_m.
(ii) For testing $H_0 : a = 0$ versus $H_1 : a \neq 0$, show that a UMP test of size α rejects H_0 when $X_{(1)} < 0$ or $2nX_{(1)}/V > c$, where c is determined by

$$(n-1)\int_0^c (1+v)^{-n} dv = 1 - \alpha.$$

Solution. Since $(X_{(1)}, V)$ is sufficient and complete for (a, θ), we may consider tests that are functions of $(X_{(1)}, V)$.
(i) When $\theta = 1$, $X_{(1)}$ is complete and sufficient for a, V is independent of $X_{(1)}$, and $V/2$ has the Gamma distribution with shape parameter $n-1$ and scale parameter θ. By Lemma 6.5 and Lemma 6.6 in Shao (2003), the UMPU test is the UMPU test in the problem where $V/2$ is an observation from the Gamma distribution with shape parameter $n-1$ and scale parameter θ, which has monotone likelihood ratio in V. Hence, the UMPU test of size α rejects H_0 when $V < c_1$ or $V > c_2$, where V has the chi-square distribution $\chi^2_{2(n-1)}$ when $\theta = 1$ and, hence, c_i's are determined by

$$\int_{c_1}^{c_2} f_{2n-2}(v)dv = 1 - \alpha$$

and

$$\int_{c_1}^{c_2} v f_{2n-2}(v)dv = (1-\alpha)\int_0^\infty v f_{2n-2}(v)dv = (1-\alpha)(2n-2).$$

Since $(2n-2)^{-1} v f_{2n-2}(v) = f_{2n}(v)$, c_i's are determined by

$$\int_{c_1}^{c_2} f_{2n-2}(v)dv = \int_{c_1}^{c_2} f_{2n}(v)dv = 1 - \alpha.$$

(ii) From Exercise 15, for testing $H_0 : a = 0$ versus $H_1 : a \neq 0$, a UMP test of size α rejects H_0 when $X_{(1)} < 0$ or $2nX_{(1)}/V > c$. It remains to determine c. When $a = 0$, $X_{(1)}/\theta$ has the chi-square distribution χ^2_2, V/θ has the chi-square distribution χ^2_{2n-2}, and they are independent. Hence, $2nX_{(1)}/[V(n-1)]$ has the F-distribution $F_{2,2(n-1)}$. Hence, $2nX_{(1)}/V$ has Lebesgue density $f(y) = (n-1)(1+y)^{-n}$. Therefore,

$$(n-1)\int_0^c (1+y)^{-n} dy = 1 - \alpha. \quad \blacksquare$$

Exercise 34 (#6.58). Let $(X_1, ..., X_n)$ be a random sample from the uniform distribution on the interval (θ, ϑ), where $-\infty < \theta < \vartheta < \infty$.
(i) Show that the conditional distribution of the smallest order statistic $X_{(1)}$ given the largest order statistic $X_{(n)} = x$ is the distribution of the minimum of a random sample of size $n-1$ from the uniform distribution on the interval (θ, x).
(ii) Find a UMPU test of size α for testing $H_0 : \theta \leq 0$ versus $H_1 : \theta > 0$.
Solution. (i) The joint Lebesgue density of $(X_{(1)}, X_{(n)})$ is

$$f(x,y) = \frac{n(n-1)(x-y)^{n-2}}{(\vartheta - \theta)^n} I_{(\theta,x)}(y) I_{(y,\vartheta)}(x)$$

and the Lebesgue density of $X_{(n)}$ is

$$g(x) = \frac{n(x-\theta)^{n-1}}{(\vartheta-\theta)^n} I_{(\theta,\vartheta)}(x).$$

Hence, the conditional density of $X_{(1)}$ given $X_{(n)} = x$ is

$$\frac{f(x,y)}{g(x)} = \frac{(n-1)(x-y)^{n-2}}{(x-\theta)^{n-1}} I_{(\theta,x)}(y),$$

which is the Lebesgue density of the smallest order statistic based on a random sample of size $n-1$ from the uniform distribution on the interval (θ, x).
(ii) Note that $(X_{(1)}, X_{(n)})$ is complete and sufficient for (θ, ϑ) and when $\theta = 0$, $X_{(n)}$ is complete for ϑ. Thus, by Lemmas 6.5 and 6.6 in Shao (2003), the UMPU test is the same as the UMPU test in the problem where $X_{(1)}$ is the smallest order statistic of a random sample of size $n-1$ from the uniform distribution on the interval (θ, x). Let $Y = x - X_{(1)}$. Then Y is the largest order statistic of a random sample of size $n-1$ from the uniform distribution on the interval $(0, \eta)$, where $\eta = x - \theta$. Thus, by the result in Exercise 14(i), a UMPU test of size α is

$$T = \begin{cases} \alpha & Y < x \\ 0 & Y > x \end{cases}$$

conditional on $X_{(n)} = x$. Since $x - X_{(1)} = Y$,

$$T = \begin{cases} \alpha & X_{(1)} > 0 \\ 0 & X_{(1)} < 0. \end{cases} \quad \blacksquare$$

Exercise 35 (#6.82). Let X be a random variable having probability density $f_\theta(x) = \exp\{\eta(\theta)Y(x) - \xi(\theta)\}h(x)$ with respect to a σ-finite measure ν, where η is an increasing and differentiable function of $\theta \in \Theta \subset \mathcal{R}$.
(i) Show that $\log \ell(\hat{\theta}) - \log \ell(\theta_0)$ is increasing (or decreasing) in Y when $\hat{\theta} > \theta_0$ (or $\hat{\theta} < \theta_0$), where $\ell(\theta) = f_\theta(x)$, $\hat{\theta}$ is an MLE of θ, and $\theta_0 \in \Theta$.
(ii) For testing $H_0 : \theta_1 \leq \theta \leq \theta_2$ versus $H_1 : \theta < \theta_1$ or $\theta > \theta_2$ or for testing $H_0 : \theta = \theta_0$ versus $H_1 : \theta \neq \theta_0$, show that there is a likelihood ratio (LR) test whose rejection region is equivalent to $Y(X) < c_1$ or $Y(X) > c_2$ for some constants c_1 and c_2.
Solution. (i) From the property of exponential families, $\hat{\theta}$ is a solution of the likelihood equation

$$\frac{\partial \log \ell(\theta)}{\partial \theta} = \eta'(\theta)Y(X) - \xi'(\theta) = 0$$

and $\psi(\theta) = \xi'(\theta)/\eta'(\theta)$ has a positive derivative $\psi'(\theta)$. Since $\eta'(\hat{\theta})Y - \xi'(\hat{\theta}) = 0$, $\hat{\theta}$ is an increasing function of Y and $\frac{d\hat{\theta}}{dY} > 0$. Consequently, for any $\theta_0 \in \Theta$,

$$\begin{aligned}
\frac{d}{dY}\left[\log \ell(\hat{\theta}) - \log \ell(\theta_0)\right] &= \frac{d}{dY}\left[\eta(\hat{\theta})Y - \xi(\hat{\theta}) - \eta(\theta_0)Y + \xi(\theta_0)\right] \\
&= \frac{d\hat{\theta}}{dY}\eta'(\hat{\theta})Y + \eta(\hat{\theta}) - \frac{d\hat{\theta}}{dY}\xi'(\hat{\theta}) - \eta(\theta_0) \\
&= \frac{d\hat{\theta}}{dY}[\eta'(\hat{\theta})Y - \xi'(\hat{\theta})] + \eta(\hat{\theta}) - \eta(\theta_0) \\
&= \eta(\hat{\theta}) - \eta(\theta_0),
\end{aligned}$$

which is positive (or negative) if $\hat{\theta} > \theta_0$ (or $\hat{\theta} < \theta_0$).
(ii) Since $\ell(\theta)$ is increasing when $\theta \le \hat{\theta}$ and decreasing when $\theta > \hat{\theta}$,

$$\lambda(X) = \frac{\sup_{\theta_1 \le \theta \le \theta_2} \ell(\theta)}{\sup_{\theta \in \Theta} \ell(\theta)} = \begin{cases} \frac{\ell(\theta_1)}{\ell(\hat{\theta})} & \hat{\theta} < \theta_1 \\ 1 & \theta_1 \le \hat{\theta} \le \theta_2 \\ \frac{\ell(\theta_2)}{\ell(\hat{\theta})} & \hat{\theta} > \theta_2 \end{cases}$$

for $\theta_1 \le \theta_2$. Hence, $\lambda(X) < c$ if and only if $\hat{\theta} < d_1$ or $\hat{\theta} > d_2$ for some constants d_1 and d_2. From the result in (i), this means that $\lambda(X) < c$ if and only if $Y < c_1$ or $Y > c_2$ for some constants c_1 and c_2. ∎

Exercise 36 (#6.83). In Exercises 55 and 56 of Chapter 4, consider $H_0 : j = 1$ versus $H_1 : j = 2$.
(i) Derive the likelihood ratio $\lambda(X)$.
(ii) Obtain an LR test of size α in Exercise 55 of Chapter 4.
Solution. Following the notation in Exercise 55 of Chapter 4, we obtain that

$$\begin{aligned}
\lambda(X) &= \begin{cases} 1 & \hat{j} = 1 \\ \frac{\ell(\hat{\theta}_1, j=1)}{\ell(\hat{\theta}_2, j=2)} & \hat{j} = 2 \end{cases} \\
&= \begin{cases} 1 & \frac{\sqrt{T_1/n}}{T_2/n} \le \sqrt{\frac{2e}{\pi}} \\ \left(\sqrt{\frac{2e}{\pi}}\frac{T_2/n}{\sqrt{T_1/n}}\right)^n & \frac{\sqrt{T_1/n}}{T_2/n} > \sqrt{\frac{2e}{\pi}}, \end{cases}
\end{aligned}$$

where $T_1 = \sum_{i=1}^n X_i^2$ and $T_2 = \sum_{i=1}^n |X_i|$. Similarly, for Exercise 56 of Chapter 4,

$$\lambda(X) = \begin{cases} 1 & h(X) = 0 \\ \frac{e^{-n\bar{X}}}{(1-\bar{X})^{n(1-\bar{X})}} & h(X) = 1, \end{cases}$$

where $\bar{X}$ is the sample mean, $h(X) = 1$ if all X_i's are not larger than 1 and $h(X) = 0$ otherwise.
(ii) Let $c \in [0, 1]$. Then the LR test of size α rejects H_0 when

$$\sqrt{\frac{2e}{\pi}} \frac{T_2/n}{\sqrt{T_1/n}} < c^{1/n},$$

where $c^{1/n}$ is the αth quantile of the distribution of $\sqrt{\frac{2e}{\pi}} \frac{T_2/n}{\sqrt{T_1/n}}$ when $X_1, ..., X_n$ are independent and identically distributed as $N(0, 1)$. ∎

Exercise 37 (#6.84). In Exercise 12, derive the likelihood ratio $\lambda(X)$ when (a) $H_0 : \theta \leq \theta_0$; (b) $H_0 : \theta_1 \leq \theta \leq \theta_2$; and (c) $H_0 : \theta \leq \theta_1$ or $\theta \geq \theta_2$.
Solution. Let f and g be the probability densities of F and G, respectively, with respect to the measure corresponding to $F + G$. Then, the likelihood function is

$$\ell(\theta) = \theta[f(X) - g(X)] + g(X)$$

and

$$\sup_{0 \leq \theta \leq 1} \ell(\theta) = \begin{cases} f(X) & f(X) \geq g(X) \\ g(X) & f(X) < g(X). \end{cases}$$

For $\theta_0 \in [0, 1]$,

$$\sup_{0 \leq \theta \leq \theta_0} \ell(\theta) = \begin{cases} \theta_0[f(X) - g(X)] + g(X) & f(X) \geq g(X) \\ g(X) & f(X) < g(X). \end{cases}$$

Hence, for $H_0 : \theta \leq \theta_0$,

$$\lambda(X) = \begin{cases} \frac{\theta_0[f(X) - g(X)] + g(X)}{f(X)} & f(X) \geq g(X) \\ 1 & f(X) < g(X). \end{cases}$$

For $0 \leq \theta_1 \leq \theta_2 \leq 1$,

$$\sup_{0 \leq \theta_1 \leq \theta_2 \leq 1} \ell(\theta) = \begin{cases} \theta_2[f(X) - g(X)] + g(X) & f(X) \geq g(X) \\ \theta_1[f(X) - g(X)] + g(X) & f(X) < g(X) \end{cases}$$

and, thus, for $H_0 : \theta_1 \leq \theta \leq \theta_2$,

$$\lambda(X) = \begin{cases} \frac{\theta_2[f(X) - g(X)] + g(X)}{f(X)} & f(X) \geq g(X) \\ \frac{\theta_1[f(X) - g(X)] + g(X)}{g(X)} & f(X) < g(X). \end{cases}$$

Finally,

$$\sup_{0 \leq \theta \leq \theta_1, \theta_2 \leq \theta \leq 1} \ell(\theta) = \sup_{0 \leq \theta \leq 1} \ell(\theta).$$

Hence, for $H_0 : \theta \leq \theta_1$ or $\theta \geq \theta_2$, $\lambda(X) = 1$. ∎

Exercise 38 (#6.85). Let $(X_1, ..., X_n)$ be a random sample from the discrete uniform distribution on $\{1, ..., \theta\}$, where θ is an integer ≥ 2. Find a level α LR test for
(i) $H_0 : \theta \leq \theta_0$ versus $H_1 : \theta > \theta_0$, where θ_0 is a known integer ≥ 2;
(ii) $H_0 : \theta = \theta_0$ versus $H_1 : \theta \neq \theta_0$.
Solution. (i) The likelihood function is

$$\ell(\theta) = \theta^{-n} I_{\{X_{(n)}, X_{(n)}+1, ...\}}(\theta),$$

where $X_{(n)}$ is the largest order statistic. Then,

$$\sup_{\theta=2,3,...} \ell(\theta) = X_{(n)}^{-n},$$

$$\sup_{\theta=2,...,\theta_0} \ell(\theta) = \begin{cases} X_{(n)}^{-n} & X_{(n)} \leq \theta_0 \\ 0 & X_{(n)} > \theta_0, \end{cases}$$

and

$$\lambda(X) = \begin{cases} 1 & X_{(n)} \leq \theta_0 \\ 0 & X_{(n)} > \theta_0. \end{cases}$$

Hence, a level α test rejects H_0 when $X_{(n)} > \theta_0$, which has size 0.
(ii) From part (i) of the solution, we obtain that

$$\lambda(X) = \begin{cases} \left(\frac{X_{(n)}}{\theta_0}\right)^n & X_{(n)} \leq \theta_0 \\ 0 & X_{(n)} > \theta_0. \end{cases}$$

Then, $\lambda(X) < c$ is equivalent to $X_{(n)} > \theta_0$ or $X_{(n)} < \theta_0 c^{1/n}$. Let $c = \alpha$. When $\theta = \theta_0$, the type I error rate is

$$\begin{aligned} P\left(X_{(n)} < \theta_0 \alpha^{1/n}\right) &= \left[P(X_1 < \theta_0 \alpha^{1/n})\right]^n \\ &= \left(\frac{\text{the integer part of } \theta_0 \alpha^{1/n}}{\theta_0}\right)^n \\ &\leq \left(\frac{\theta_0 \alpha^{1/n}}{\theta_0}\right)^n \\ &= \alpha. \end{aligned}$$

Hence, an LR test of level α rejects H_0 if $X_{(n)} > \theta_0$ or $X_{(n)} < \theta_0 \alpha^{1/n}$. ∎

Exercise 39 (#6.87). Let $X = (X_1, ..., X_n)$ be a random sample from the exponential distribution on the interval (a, ∞) with scale parameter θ.
(i) Suppose that θ is known. Find an LR test of size α for testing $H_0 : a \leq a_0$ versus $H_1 : a > a_0$, where a_0 is a known constant.
(ii) Suppose that θ is known. Find an LR test of size α for testing $H_0 : a =$

a_0 versus $H_1 : a \neq a_0$.
(iii) Repeat part (i) for the case where θ is also unknown.
(iv) When both θ and a are unknown, find an LR test of size α for testing $H_0 : \theta = \theta_0$ versus $H_1 : \theta \neq \theta_0$.
(v) When $a > 0$ and $\theta > 0$ are unknown, find an LR test of size α for testing $H_0 : a = \theta$ versus $H_1 : a \neq \theta$.
Solution. (i) The likelihood function is

$$\ell(a, \theta) = \theta^{-n} e^{na/\theta} e^{-n\bar{X}/\theta} I_{(a,\infty)}(X_{(1)}),$$

where $\bar{X}$ is the sample mean and $X_{(1)}$ is the smallest order statistic. When θ is known, the MLE of a is $X_{(1)}$. When $a \leq a_0$, the MLE of a is $\min\{a_0, X_{(1)}\}$. Hence, the likelihood ratio is

$$\lambda(X) = \begin{cases} 1 & X_{(1)} < a_0 \\ e^{-n(X_{(1)} - a_0)/\theta} & X_{(1)} \geq a_0. \end{cases}$$

Then $\lambda(X) < c$ is equivalent to $X_{(1)} > d$ for some $d \geq a_0$. To determine d, note that

$$\begin{aligned} \sup_{a \leq a_0} P(X_{(1)} > d) &= \sup_{a \leq a_0} \frac{n e^{na/\theta}}{\theta} \int_d^\infty e^{-nx/\theta} dx \\ &= \frac{n e^{na_0/\theta}}{\theta} \int_d^\infty e^{-nx/\theta} dx \\ &= e^{n(a_0 - d)/\theta}. \end{aligned}$$

Setting this probability to α yields $d = a_0 - n^{-1}\theta \log \alpha$.
(ii) Note that $\ell(a_0, \theta) = 0$ when $X_{(1)} < a_0$. Hence the likelihood ratio is

$$\lambda(X) = \begin{cases} 0 & X_{(1)} < a_0 \\ e^{-n(X_{(1)} - a_0)/\theta} & X_{(1)} \geq a_0. \end{cases}$$

Therefore, $\lambda(X) < c$ is equivalent to $X_{(1)} \leq a_0$ or $X_{(1)} > d$ for some $d \geq a_0$. From part (i) of the solution, $d = a_0 - n^{-1}\theta \log \alpha$ leads to an LR test of size α.
(iii) The MLE of (a, θ) is $(X_{(1)}, \bar{X} - X_{(1)})$. When $a \leq a_0$, the MLE of a_0 is $\min\{a_0, X_{(1)}\}$ and the MLE of θ is

$$\hat{\theta}_0 = \begin{cases} \bar{X} - X_{(1)} & X_{(1)} < a_0 \\ \bar{X} - a_0 & X_{(1)} \geq a_0. \end{cases}$$

Therefore, the likelihood ratio is

$$\lambda(X) = \begin{cases} [T(X)]^n & X_{(1)} \geq a_0 \\ 1 & X_{(1)} < a_0, \end{cases}$$

where $T(X) = (\bar{X} - X_{(1)})/(\bar{X} - a_0)$. Hence the LR test rejects H_0 if and only if $T(X) < c^{1/n}$. From the solution to Exercise 33, $Y = (\bar{X} - X_{(1)})/(\bar{X} - a)$ has the beta distribution with parameter $(n-1, 1)$. Then,

$$\begin{aligned}
\sup_{a \le a_0} P\left(T(X) < c^{1/n}\right) &= \sup_{a \le a_0} P\left(\frac{\bar{X} - X_{(1)}}{\bar{X} - a + a - a_0} < c^{1/n}\right) \\
&= P\left(Y < c^{1/n}\right) \\
&= (n-1)\int_0^{c^{1/n}} x^{n-2} dx \\
&= c^{(n-1)/n}.
\end{aligned}$$

Setting this probability to α yields $c = \alpha^{n/(n-1)}$.
(iv) Under H_0, the MLE is $(X_{(1)}, \theta_0)$. Let $Y = \theta_0^{-1} n(\bar{X} - X_{(1)})$. Then the likelihood ratio is

$$\lambda(X) = e^n Y^n e^{-Y}.$$

Thus, $\lambda(X) < c$ is equivalent to $Y < c_1$ or $Y > c_2$. Under H_0, $2Y$ has the chi-square distribution χ^2_{2n-2}. Hence, an LR test of size α rejects H_0 when $2Y < \chi^2_{2(n-1),1-\alpha/2}$ or $2Y > \chi^2_{2(n-1),\alpha/2}$, where $\chi^2_{r,\alpha}$ is the $(1-\alpha)$th quantile of the chi-square distribution χ^2_r.
(v) Under H_0, the MLE is $(\bar{X}, \bar{X})$. Let $Y = X_{(1)}/(\bar{X} - X_{(1)})$. Then the likelihood ratio is

$$\lambda(X) = e^n \bar{X}^{-n} (\bar{X} - X_{(1)})^n = e^n (1 + Y)^{-n},$$

where $Y = X_{(1)}/(\bar{X} - X_{(1)})$. Then $\lambda(X) < c$ is equivalent to $Y > b$ for some b. Under H_0, the distribution Y is the same as that of the ratio Y_1/Y_2, where Y_1 has the exponential distribution on the interval (n, ∞) and scale parameter 1, Y_2 has the gamma distribution with shape parameter $n-1$ and scale parameter 1, and Y_1 and Y_2 are independent. Hence, b satisfies

$$\begin{aligned}
\alpha &= \frac{e^n}{\Gamma(n-1)} \int_0^\infty y_2^{n-2} e^{-y_2} \int_{\max\{n, by_2\}}^\infty e^{-y_1} dy_1 dy_2 \\
&= \frac{e^n}{\Gamma(n-1)} \int_0^\infty y_2^{n-2} e^{-y_2} e^{-\max\{n, by_2\}} dy_2. \quad \blacksquare
\end{aligned}$$

Exercise 40. Let $(X_1, ..., X_n)$ and $(Y_1, ..., Y_n)$ be independent random samples from $N(\mu_1, 1)$ and $N(\mu_2, 1)$, respectively, where $-\infty < \mu_2 \le \mu_1 < \infty$.
(i) Derive the likelihood ratio λ and an LR test of size α for $H_0 : \mu_1 = \mu_2$ versus $H_1 : \mu_1 > \mu_2$.
(ii) Derive the distribution of $-2\log\lambda$ and the power of the LR test in (ii).

(iii) Verify that the LR test in (ii) is a UMPU test.

Solution. (i) The likelihood function is

$$\ell(\mu_1, \mu_2) = \frac{1}{(2\pi)^n} \exp\left\{-\frac{1}{2}\sum_{i=1}^n (X_i - \mu_1)^2 - \frac{1}{2}\sum_{i=1}^n (Y_i - \mu_2)^2\right\}.$$

Let $\bar{X}$ be the sample mean based on X_i's and $\bar{Y}$ be the sample mean based on Y_i's. When $\mu_1 = \mu_2$, $\ell(\mu_1, \mu_2)$ is maximized at $\bar{\mu} = (\bar{X} + \bar{Y})/2$. The MLE of (μ_1, μ_2) is equal to $(\bar{X}, \bar{Y})$ when $\bar{X} \geq \bar{Y}$. Consider the case $\bar{X} < \bar{Y}$. For any fixed $\mu_1 \leq \bar{Y}$, $\ell(\mu_1, \mu_2)$ increases in μ_2 and, hence,

$$\sup_{\mu_1 \leq \bar{Y},\ \mu_2 \leq \mu_1} \ell(\mu_1, \mu_2) = \sup_{\mu_1 \leq \bar{Y}} \ell(\mu_1, \mu_1).$$

Also,

$$\sup_{\mu_1 > \bar{Y},\ \mu_2 \leq \mu_1} \ell(\mu_1, \mu_2) = \sup_{\mu_1 > \bar{Y}} \ell(\mu_1, \bar{Y}) \leq \sup_{\mu_1 \leq \bar{Y}} \ell(\mu_1, \mu_1),$$

since $\bar{X} < \bar{Y}$. Hence,

$$\sup_{\mu_2 \leq \mu_1} \ell(\mu_1, \mu_1) = \sup_{\mu_1 \leq \bar{Y}} \ell(\mu_1, \mu_1) = \ell(\bar{\mu}, \bar{\mu}),$$

since $\bar{\mu} < \bar{Y}$ when $\bar{X} < \bar{Y}$. This shows that the MLE is $(\bar{\mu}, \bar{\mu})$ when $\bar{X} < \bar{Y}$. Therefore, the likelihood ratio $\lambda = 1$ when $\bar{X} < \bar{Y}$ and

$$\begin{aligned}\lambda &= \frac{\exp\left\{-\frac{1}{2}\sum_{i=1}^n (X_i - \bar{\mu})^2 - \frac{1}{2}\sum_{i=1}^n (Y_i - \bar{\mu})^2\right\}}{\exp\left\{-\frac{1}{2}\sum_{i=1}^n (X_i - \bar{X})^2 - \frac{1}{2}\sum_{i=1}^n (Y_i - \bar{Y})^2\right\}} \\ &= \exp\left\{-\tfrac{n}{4}(\bar{X} - \bar{Y})^2\right\}\end{aligned}$$

when $\bar{X} \geq \bar{Y}$. Hence,

$$-2\log\lambda = \begin{cases} \frac{n}{2}(\bar{X} - \bar{Y})^2 & \bar{X} \geq \bar{Y} \\ 0 & \bar{X} < \bar{Y}. \end{cases}$$

Note that $\lambda < c$ for some $c \in (0, 1)$ is equivalent to $\sqrt{\frac{n}{2}}(\bar{X} - \bar{Y}) > d$ for some $d > 0$. Under H_0, $\sqrt{\frac{n}{2}}(\bar{X} - \bar{Y})$ has the standard normal distribution. Hence, setting $d = \Phi^{-1}(1 - \alpha)$ yields an LR test of size α, where Φ is the cumulative distribution function of the standard normal distribution.

(ii) Note that $-2\log\lambda \geq 0$. Let $\delta = \sqrt{\frac{n}{2}}(\mu_1 - \mu_2)$. Then $Z = \sqrt{\frac{n}{2}}(\bar{X} - \bar{Y})$ has distribution $N(\delta, 1)$. For $t > 0$,

$$\begin{aligned}P(-2\log\lambda \leq t) &= P(-2\log\lambda \leq t, Z > 0) + P(-2\log\lambda \leq t, Z \leq 0) \\ &= P(Z \leq \sqrt{t}, Z > 0) + P(Z \leq \sqrt{t}, Z \leq 0) \\ &= P(0 < Z \leq \sqrt{t}) + P(Z \leq 0) \\ &= P(Z \leq \sqrt{t}) \\ &= \Phi(\sqrt{t} - \delta).\end{aligned}$$

The power of the LR test in (ii) is then

$$1 - \Phi(d - \delta) = 1 - \Phi(\Phi^{-1}(1-\alpha) - \delta).$$

(iii) The likelihood can be written as

$$C_{X,Y} \exp\left\{\theta \bar{X} + n\mu_2 U - n(\mu_1 + \mu_2)/2\right\},$$

where $\theta = n(\mu_1 - \mu_2)$, $U = \bar{X} + \bar{Y}$, and $C_{X,Y}$ is a quantity does not depend on any parameters. When $\theta = 0$ ($\mu_1 = \mu_2$), Z is independent of U. Also, $Z = \sqrt{\frac{n}{2}}(2\bar{X} - U)$. Hence, by Theorem 6.4 and Lemma 6.7 in Shao (2003), the UMPU test for $H_0 : \theta = 0$ versus $H_1 : \theta > 0$ rejects H_0 when $Z > c$ for some c, which is the same as the LR test in (ii). ∎

Exercise 41 (#6.89). Let $X_{i1}, ..., X_{in_i}$, $i = 1, 2$, be two independent random samples from the uniform distributions on $(0, \theta_i)$, $i = 1, 2$, respectively, where $\theta_1 > 0$ and $\theta_2 > 0$ are unknown.
(i) Find an LR test of size α for testing $H_0 : \theta_1 = \theta_2$ versus $H_1 : \theta_1 \neq \theta_2$.
(ii) Derive the limiting distribution of $-2 \log \lambda$ when $n_1/n_2 \to \kappa \in (0, \infty)$, where λ is the likelihood ratio in part (i).
Solution. (i) Let $Y_i = \max_{j=1,\ldots,n_i} X_{ij}$, $i = 1, 2$, and $Y = \max\{Y_1, Y_2\}$. The likelihood function is

$$\ell(\theta_1, \theta_2) = \theta_1^{-n_1} \theta_2^{-n_2} I_{(0,\theta_1)}(Y_1) I_{(0,\theta_2)}(Y_2)$$

and the MLE of θ_i is Y_i, $i = 1, 2$. When $\theta_1 = \theta_2$,

$$\ell(\theta_1, \theta_1) = \theta_1^{-n_1 - n_2} I_{(0,\theta_1)}(Y)$$

and the MLE of θ_1 is Y. Hence, the likelihood ratio is

$$\lambda = \frac{Y_1^{n_1} Y_2^{n_2}}{Y^{n_1+n_2}}.$$

Assume that $\theta_1 = \theta_2$. For any $t \in (0, 1)$,

$$\begin{aligned}
P(\lambda < t) &= P(\lambda < t, Y_1 \geq Y_2) + P(\lambda < t, Y_1 < Y_2) \\
&= P\left(Y_2 < t^{1/n_2} Y_1, Y_1 \geq Y_2\right) + P\left(Y_1 < t^{1/n_1} Y_2, Y_1 \geq Y_2\right) \\
&= P\left(Y_2 < t^{1/n_2} Y_1\right) + P\left(Y_1 < t^{1/n_1} Y_2\right) \\
&= n_1 n_2 \int_0^1 \int_0^{t^{1/n_2} y_1} y_2^{n_2 - 1} y_1^{n_1 - 1} dy_2 dy_1 \\
&\quad + n_1 n_2 \int_0^1 \int_0^{t^{1/n_1} y_2} y_1^{n_1 - 1} y_2^{n_2 - 1} dy_1 dy_2
\end{aligned}$$

$$= n_1 t \int_0^1 y_1^{n_1+n_2-1} dy_1 + n_2 t \int_0^1 y_2^{n_1+n_2-1} dy_2$$
$$= (n_1+n_2)t \int_0^1 y^{n_1+n_2-1} dy$$
$$= t.$$

Hence, the LR test of size α rejects H_0 when $\lambda < \alpha$.
(ii) Under H_0, by part (i) of the solution, λ has the uniform distribution on $(0,1)$, which does not depend on (n_1, n_2). The distribution of $-2\log\lambda$ is then the exponential distribution on the interval $(0,\infty)$ with scale parameter 2^{-1}, which is also the chi-square distribution χ_1^2. Consider now the limiting distribution of $-2\log\lambda$ when $n_1/n_2 \to \kappa \in (0,\infty)$. Assume that $\theta_1 < \theta_2$. Then

$$P(Y_1 > Y_2) = P(Y_2 - Y_1 - (\theta_2 - \theta_1) < -(\theta_2 - \theta_1)) \to 0$$

since $Y_2 - Y_1 \to_p \theta_2 - \theta_1$. Thus, for the limiting distribution of $-2\log\lambda$, we may assume that $Y_1 \le Y_2$ and, consequently, $-2\log\lambda = 2n_1(\log Y_2 - \log Y_1)$. Note that

$$n_i(\theta_i - Y_i) \to_d \theta_i Z_i, \quad i = 1, 2,$$

where Z_i has the exponential distribution on the interval $(0,\infty)$ with scale parameter 1. By the δ-method,

$$n_i(\log\theta_i - \log Y_i) \to_d Z_i, \quad i = 1, 2.$$

Because Y_1 and Y_2 are independent,

$$\begin{aligned} -2\log\lambda + 2\log(\theta_1/\theta_2)^{n_1} &= 2[n_1(\log\theta_1 - \log Y_1) - \tfrac{n_1}{n_2} n_2(\log\theta_2 - \log Y_2)] \\ &\to_d 2(Z_1 - \kappa Z_2), \end{aligned}$$

where Z_1 and Z_2 are independent. The limiting distribution of $-2\log\lambda$ for the case of $\theta_1 > \theta_2$ can be similarly obtained. ∎

Exercise 42 (#6.90). Let $(X_{i1}, ..., X_{in_i})$, $i = 1, 2$, be two independent random samples from $N(\mu_i, \sigma_i^2)$, $i = 1, 2$, respectively, where μ_i's and σ_i^2's are unknown. For testing $H_0 : \sigma_2^2/\sigma_1^2 = \Delta_0$ versus $H_1 : \sigma_2^2/\sigma_1^2 \neq \Delta_0$ with a known $\Delta_0 > 0$, derive an LR test of size α and compare it with the UMPU test.

Solution. The MLE of $(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2)$ is $(\bar{X}_1, \bar{X}_2, \hat{\sigma}_1^2, \hat{\sigma}_2^2)$, where $\bar{X}_i$ is the sample mean based on $X_{i1}, ..., X_{in_i}$ and

$$\hat{\sigma}_i^2 = \frac{1}{n_i}\sum_{j=1}^{n_i}(X_{ij} - \bar{X}_j)^2,$$

$i = 1, 2$. Under H_0, the MLE of $(\mu_1, \mu_2, \sigma_1^2)$ is $(\bar{X}_1, \bar{X}_2, \tilde{\sigma}_1^2)$, where

$$\tilde{\sigma}_1^2 = \frac{\Delta_0 \sum_{j=1}^{n_1}(X_{1j} - \bar{X}_1)^2 + \sum_{j=1}^{n_2}(X_{2j} - \bar{X}_2)^2}{\Delta_0(n_1 + n_2)}.$$

Then the likelihood ratio is proportional to

$$\left(\frac{1}{1+F}\right)^{n_1/2} \left(\frac{F}{1+F}\right)^{n_2/2},$$

where $F = S_2^2/(\Delta_0 S_1^2)$ and S_j^2 is the sample variance based on $X_{i1}, ..., X_{in_i}$, $i = 1, 2$. Under H_0, F has the F-distribution F_{n_2-1,n_1-1}. Since the likelihood ratio is a unimodal function in F, an LR test is equivalent to the one that rejects the null hypothesis when $F < c_1$ or $F > c_2$ for some positive constants $c_1 < c_2$ chosen so that $P(F < c_1) + P(F > c_2) = \alpha$ under H_0. Note that the UMPU test is one of these tests with an additional requirement being unbiased, i.e., the c_i's must satisfy $P(B < c_1)+P(B > c_2) = \alpha$, where B has the beta distribution with parameter $(\frac{n_2+1}{2}, \frac{n_1-1}{2})$ (e.g., Shao, 2003, p. 414). ∎

Exercise 43 (#6.91). Let (X_{i1}, X_{i2}), $i = 1, ..., n$, be a random sample from the bivariate normal distribution with unknown mean and covariance matrix. For testing $H_0 : \rho = 0$ versus $H_1 : \rho \neq 0$, where ρ is the correlation coefficient, show that the test rejecting H_0 when $|R| > c$ is an LR test, where

$$R = \sum_{i=1}^{n}(X_{i1} - \bar{X}_1)(X_{i2} - \bar{X}_2)\Bigg/\left[\sum_{i=1}^{n}(X_{i1} - \bar{X}_1)^2 + \sum_{i=1}^{n}(X_{i2} - \bar{X}_2)^2\right]$$

is the sample correlation coefficient and $\bar{X}_j$ is the sample mean based on $X_{1j}, ..., X_{nj}$. Discuss the form of the limiting distribution of $-2\log\lambda$, where λ is the likelihood ratio.

Solution. From the normal distribution theory, the MLE of the means are $\bar{X}_1$ and $\bar{X}_2$ and the MLE of the variances are $n^{-1}\sum_{i=1}^{n}(X_{i1} - \bar{X}_1)^2$ and $n^{-1}\sum_{i=1}^{n}(X_{i2} - \bar{X}_2)^2$, regardless of whether H_0 holds or not. The MLE of ρ is the sample correlation coefficient R. Under H_0, $\rho = 0$. Using these results, the likelihood ratio is $\lambda = (1 - R^2)^{n/2}$. Hence, an LR test rejects H_0 when $|R| > c$ for some $c > 0$.

The distribution of R under H_0 is given in Exercise 9(ii) in Chapter 2. Hence, the Lebesgue density of $-2\log\lambda$ is

$$\frac{\Gamma(\frac{n-1}{2})}{\sqrt{\pi}\Gamma(\frac{n-2}{2})}(1 - e^{-x})^{-1/2}e^{-(n-2)x/2}I_{(0,\infty)}(x).$$

When $\rho \neq 0$, it follows from the result in Exercise 9(i) of Chapter 2 that $\sqrt{n}(R - \rho) \to_d N(0, (1 - \rho^2)^2/(1 + \rho^2))$. By the δ-method,

$$\sqrt{n}\left[\frac{-2\log\lambda}{n} - \log(1 - \rho^2)\right] \to_d N\left(0, \tfrac{4\rho^2}{1+\rho^2}\right). \quad ∎$$

Exercise 44. Consider the problem in Exercise 63 of Chapter 4. Find an LR test of size α for testing $H_0 : \theta_1 = \theta_2$ versus $H_1 : \theta_1 \neq \theta_2$. Discuss the limiting distribution of $-2\log\lambda$, where λ is the likelihood ratio.

Solution. From the solution of Exercise 44, the MLE of (θ_1, θ_2) is equal to $n^{-1}(\sqrt{T_1T_2} + T_1, \sqrt{T_1T_2} + T_2)$, where $T_1 = \sum_{i=1}^n X_i I_{(0,\infty)}(X_i)$ and $T_2 = -\sum_{i=1}^n X_i I_{(-\infty,0]}(X_i)$. When $\theta_1 = \theta_2$, the distribution of X_i is the double exponential distribution with mean 0 and scale parameter $\theta_1 = \theta_2$. Hence, the MLE of $\theta_1 = \theta_2$ is $(T_1 + T_2)/n$. Since the likelihood function is

$$\ell(\theta_1, \theta_2) = (\theta_1 + \theta_2)^{-n} \exp\left\{-\frac{T_1}{\theta_1} - \frac{T_2}{\theta_2}\right\},$$

the likelihood ratio is

$$\lambda = \frac{(\sqrt{T_1} + \sqrt{T_2})^{2n}}{2^n(T_1 + T_2)^n}.$$

Under H_0, the distribution of λ does not depend on any unknown parameter. Hence, an LR test of size α rejects H_0 when $\lambda < c$, where c is the $(1-\alpha)$th quantile of λ under H_0.

Note that

$$E(T_i) = \frac{n\theta_i^2}{\theta_1 + \theta_2}, \quad \mathrm{Var}(T_i) = \frac{n[2\theta_i^3(\theta_1 + \theta_2) - \theta_i^4]}{(\theta_1 + \theta_2)^2}, \quad i = 1, 2,$$

and

$$\mathrm{Cov}(T_1, T_2) = -E(T_1)E(T_2) = -\frac{n^2\theta_1^2\theta_2^2}{(\theta_1 + \theta_2)^2}.$$

Hence, by the central limit theorem,

$$\sqrt{n}\left[\begin{pmatrix} T_1/n \\ T_2/n \end{pmatrix} - \begin{pmatrix} \frac{\theta_1^2}{\theta_1+\theta_2} \\ \frac{\theta_2^2}{\theta_1+\theta_2} \end{pmatrix}\right] \to_d N_2\left(0, \begin{pmatrix} \frac{\theta_1^3(\theta_1+2\theta_2)}{(\theta_1+\theta_2)^2} & -\frac{\theta_1^2\theta_2^2}{(\theta_1+\theta_2)^2} \\ -\frac{\theta_1^2\theta_2^2}{(\theta_1+\theta_2)^2} & \frac{\theta_2^3(\theta_2+2\theta_1)}{(\theta_1+\theta_2)^2} \end{pmatrix}\right).$$

Let $g(x, y) = 2\log(\sqrt{x} + \sqrt{y}) - \log(x + y) - \log 2$. Then $n^{-1}\log\lambda = g(T_1/n, T_2/n)$. The derivatives

$$\frac{\partial g(x,y)}{\partial x} = \frac{1}{x + \sqrt{xy}} - \frac{1}{x + y} \quad \text{and} \quad \frac{\partial g(x,y)}{\partial y} = \frac{1}{y + \sqrt{xy}} - \frac{1}{x + y}$$

at $x = E(T_1)$ and $x = E(T_2)$ are equal to $\frac{\theta_2(\theta_2-\theta_1)}{\theta_1(\theta_1^2+\theta_2^2)}$ and $\frac{\theta_1(\theta_1-\theta_2)}{\theta_2(\theta_1^2+\theta_2^2)}$, respectively. Hence, by the δ-method,

$$\sqrt{n}\left[n^{-1}\log\lambda - \log\frac{(\theta_1+\theta_2)^2}{2(\theta_1^2+\theta_2^2)}\right] \to_d N(0, \tau^2),$$

where

$$\tau^2 = \frac{[\theta_1\theta_2^2(\theta_1 + 2\theta_2) + \theta_2\theta_1^2(\theta_2 + 2\theta_1) + 2\theta_1^2\theta_2^2](\theta_1 - \theta_2)^2}{(\theta_1 + \theta_2)^2(\theta_1^2 + \theta_2^2)^2}. \quad \blacksquare$$

Exercise 45 (#6.93). Let X_1 and X_2 be independent observations from the binomial distributions with sizes n_1 and n_2 and probabilities p_1 and p_2, respectively, where n_i's are known and p_i's are unknown.
(i) Find an LR test of level α for testing $H_0 : p_1 = p_2$ versus $H_1 : p_1 \neq p_2$.
(ii) Find an LR test of level α for testing $H_0 : p_1 \geq p_2$ versus $H_1 : p_1 < p_2$. Is this test a UMPU test?
Solution. (i) The likelihood function is

$$\ell(p_1, p_2) = C_X p_1^{X_1}(1-p_1)^{n_1-X_1} p_2^{X_2}(1-p_2)^{n_2-X_2},$$

where C_X is a quantity not depending on (p_1, p_2). The MLE of (p_1, p_2) is $(X_1/n_1, X_2/n_2)$. Under H_0, the MLE of $p_1 = p_2$ is U/n, where $U = X_1+X_2$ and $n = n_1 + n_2$. Then, the likelihood ratio is

$$\lambda_1 = \frac{\left(\frac{U}{n}\right)^U \left(1-\frac{U}{n}\right)^{n-U}}{\left(\frac{X_1}{n_1}\right)^{X_1}\left(1-\frac{X_1}{n_1}\right)^{n_1-X_1}\left(\frac{X_2}{n_2}\right)^{X_2}\left(1-\frac{X_2}{n_2}\right)^{n_2-X_2}}.$$

An LR test rejects H_0 when $\lambda_1 < c$, which is equivalent to $\psi(X_1, X_2) > g(U)$ for some function g, where

$$\psi(X_1, X_2) = \left(\frac{X_1}{n_1}\right)^{X_1}\left(1-\frac{X_1}{n_1}\right)^{n_1-X_1}\left(\frac{X_2}{n_2}\right)^{X_2}\left(1-\frac{X_2}{n_2}\right)^{n_2-X_2}$$

is the denominator of λ_1. To determine $g(U)$, we note that, under H_0, the conditional distribution of $\psi(X_1, X_2)$ given U does not depend on any unknown parameter (which follows from the sufficiency of U under H_0). Hence, if we choose $g(U)$ such that $P(\psi(X_1, X_2) > g(U)|U) \leq \alpha$, then $P(\psi(X_1, X_2) > g(U)) \leq \alpha$.
(ii) Using the same argument used in the solution of Exercise 40, we can show that the MLE of (p_1, p_2) under H_0 $(p_1 \geq p_2)$ is equal to $(X_1/n_1, X_2/n_2)$ if $X_1/n_1 \geq X_2/n_2$ and is equal to $(U/n, U/n)$ if $X_1/n_1 < X_2/n_2$. Hence, the likelihood ratio is

$$\lambda = \begin{cases} \lambda_1 & X_1/n_1 < X_2/n_2 \\ 1 & X_1/n_1 \geq X_2/n_2, \end{cases}$$

where λ_1 is given in part (i) of the solution. Hence, an LR test rejects H_0 when $\lambda_1 < c$ and $X_1/n_1 < X_2/n_2$, which is equivalent to the test that rejects H_0 when $\lambda_1^{-1} > c^{-1}$ and $U/(1 + n_1/n_2) < X_2$. Note that $\lambda_1^{-1} = h(X_2, U)/q(U)$, where

$$h(X_2, U) = \left(\frac{U-X_2}{n_1}\right)^{U-X_2}\left(1-\frac{U-X_2}{n_1}\right)^{n_1-U+X_2}\left(\frac{X_2}{n_2}\right)^{X_2}\left(1-\frac{X_2}{n_2}\right)^{n_2-X_2}$$

and

$$q(U) = \left(\tfrac{U}{n}\right)^U \left(1 - \tfrac{U}{n}\right)^{n-U}.$$

Since $\frac{X_2(n_1-U+X_2)}{(n_2-X_2)(U-X_2)} > 1$ when $U/(1+n_1/n_2) < X_2$, the derivative

$$\frac{\partial \log h(X_2, U)}{\partial X_2} = \log\left(\frac{X_2(n_1 - U + X_2)}{(n_2 - X_2)(U - X_2)}\right) > 0,$$

i.e., $h(X_2, U)$ is increasing when $U/(1 + n_1/n_2) < X_2$. Thus, the LR test is equivalent to the test that rejects H_0 when $X_2 > c(U)$ for some function $c(U)$. The difference between this test and the UMPU test derived in Exercise 24 is that the UMPU test is of size α and possibly randomized, whereas the LR test is of level α and nonrandomized. ∎

Exercise 46 (#6.95). Let X_1 and X_2 be independently distributed as the exponential distributions on the interval $(0, \infty)$ with unknown scale parameters θ_i, $i = 1, 2$, respectively. Define $\theta = \theta_1/\theta_2$. Find an LR test of size α for testing
(i) $H_0 : \theta = 1$ versus $H_1 : \theta \neq 1$;
(ii) $H_0 : \theta \leq 1$ versus $H_1 : \theta > 1$.
Solution. (i) Since the MLE of (θ_1, θ_2) is (X_1, X_2) and, under H_0, the MLE of $\theta_1 = \theta_2$ is $(X_1 + X_2)/2$, the likelihood ratio is

$$\lambda = \frac{X_1 X_2}{\left(\frac{X_1+X_2}{2}\right)^2} = \frac{4F}{(1+F)^2},$$

where $F = X_2/X_1$. Note that $\lambda < c$ is equivalent to $F < c_1$ or $F > c_2$ for $0 < c_1 < c_2$. Under H_0, F is has the F-distribution $F_{2,2}$. Hence, an LR test of size α rejects H_0 when $F < c_1$ or $F > c_2$ with c_i's determined by $P(F < c_1) + P(F > c_2) = \alpha$ under H_0.
(ii) Using the same argument used in Exercises 40 and 45, we obtain the likelihood ratio

$$\lambda = \begin{cases} 1 & X_1 < X_2 \\ \frac{4F}{(1+F)^2} & X_1 \geq X_2. \end{cases}$$

Note that $\lambda < c$ if and only if $4F/(1 + F)^2 < c$ and $X_1 \geq X_2$, which is equivalent to $F < b$ for some b. Let $F_{2,2}$ be a random variable having the F-distribution $F_{2,2}$. Then

$$\sup_{\theta_1 \leq \theta_2} P(F < b) = \sup_{\theta_1 \leq \theta_2} P\left(F_{2,2} < \frac{b\theta_1}{\theta_2}\right) = P\left(F_{2,2} < b\right).$$

Hence, an LR test of size α rejects H_0 when $F < b$ with b being the αth quantile of the F-distribution $F_{2,2}$. ∎

Exercise 47 (#6.98). Let $(X_1, ..., X_n)$ be a random sample from $N(\mu, \sigma^2)$.
(i) Suppose that $\sigma^2 = \gamma\mu^2$ with unknown $\gamma > 0$ and $\mu \in \mathcal{R}$. Find an LR test for testing $H_0 : \gamma = 1$ versus $H_1 : \gamma \neq 1$.
(ii) In the testing problem in (i), find the forms of Wald's test and Rao's score test.
(iii) Repeat (i) and (ii) when $\sigma^2 = \gamma\mu$ with unknown $\gamma > 0$ and $\mu > 0$.
Solution. (i) The likelihood function is

$$\ell(\mu, \gamma) = (\sqrt{2\pi\gamma}|\mu|)^{-n} \exp\left\{-\frac{1}{2\gamma\mu^2}\sum_{i=1}^n (X_i - \mu)^2\right\}.$$

The MLE of (μ, γ) is $(\hat{\mu}, \hat{\gamma}) = (\bar{X}, \hat{\sigma}^2/\bar{X}^2)$, where $\bar{X}$ is the sample mean and $\hat{\sigma}^2 = n^{-1}\sum_{i=1}^n (X_i - \bar{X})^2$. Under H_0, by Exercise 41(viii) in Chapter 4, the MLE of μ is

$$\hat{\mu}_0 = \begin{cases} \mu_+ & \ell(\mu_+, 1) > \ell(\mu_-, 1) \\ \mu_- & \ell(\mu_+, 1) \leq \ell(\mu_-, 1), \end{cases}$$

where

$$\mu_\pm = \frac{-\bar{X} \pm \sqrt{5\bar{X}^2 + 4\hat{\sigma}^2}}{2}.$$

The likelihood ratio is

$$\lambda = \frac{\ell(\hat{\mu}_0, 1)}{\ell(\hat{\mu}, \hat{\gamma})} = \frac{e^{n/2}\hat{\sigma}^n}{|\hat{\mu}_0|^n} \exp\left\{-\frac{n\hat{\sigma}^2 + n(\hat{\mu}_0 - \bar{X})^2}{2\hat{\mu}_0^2}\right\},$$

which is a function of $\bar{X}^2/\hat{\sigma}^2$. Under H_0, the distribution of $\bar{X}^2/\hat{\sigma}^2$ does not depend on any unknown parameter. Hence, an LR test can be constructed with rejection region $\lambda < c$.
(ii) Let

$$s(\mu, \gamma) = \frac{\partial \log \ell(\mu, \gamma)}{\partial(\mu, \gamma)} = \begin{pmatrix} -\frac{n}{\mu} + \frac{n(\bar{X}-\mu)}{\gamma\mu^2} + \frac{\sum_{i=1}^n (X_i-\mu)^2}{\gamma\mu^3} \\ -\frac{n}{2\gamma} + \frac{\sum_{i=1}^n (X_i-\mu)^2}{2\mu^2\gamma^2} \end{pmatrix}.$$

The Fisher information about (μ, γ) is

$$I_n(\mu, \gamma) = E\{s(\mu, \gamma)[s(\mu, \gamma)]^\tau\} = n \begin{pmatrix} \frac{1}{\mu^2\gamma} + \frac{2}{\mu^2} & \frac{1}{\mu\gamma} \\ \frac{1}{\mu\gamma} & \frac{1}{2\gamma^2} \end{pmatrix}.$$

Then, Rao's score test statistic is

$$R_n = [s(\hat{\mu}_0, 1)]^\tau [I_n(\hat{\mu}_0, 1)]^{-1} s(\hat{\mu}_0, 1),$$

where $\hat{\mu}_0$ is given in part (i) of the solution.

Let $R(\mu,\gamma) = \gamma - 1$. Then $\partial R/\partial\mu = 0$ and $\partial R/\partial\gamma = 1$. Wald's test statistic W_n is equal to $[R(\hat\mu,\hat\gamma)]^2$ divided by the last element of the inverse of $I_n(\hat\mu,\hat\gamma)$, where $\hat\mu$ and $\hat\gamma$ are given in part (i) of the solution. Hence,

$$W_n = \frac{n(\hat\gamma - 1)^2}{2\hat\gamma^2 + 4\hat\gamma^3}.$$

(iii) Let $T = n^{-1}\sum_{i=1}^n X_i^2$. The likelihood function is

$$\ell(\mu,\gamma) = (\sqrt{2\pi\gamma\mu})^{-n}\exp\left\{-\frac{nT}{2\gamma\mu} + \frac{n\bar X}{\gamma} - \frac{n\mu}{2\gamma}\right\}.$$

When $\gamma = 1$, it is shown in Exercise 60 of Chapter 4 that the MLE of μ is $\hat\mu_0 = (\sqrt{1+4T} - 1)/2$. For the MLE of (μ,γ), it is equal to $(\hat\mu,\hat\gamma) = (\bar X, \hat\sigma^2/\bar X)$ when $\bar X > 0$. If $\bar X \le 0$, however, the likelihood is unbounded in γ. Hence, the likelihood ratio is

$$\lambda = \begin{cases} e^{n/2}\hat\sigma^n\hat\mu_0)^{-n/2}\exp\left\{-\frac{nT}{2\hat\mu_0} + n\bar X - \frac{n\hat\mu_0}{2}\right\} & \bar X > 0 \\ 0 & \bar X \le 0. \end{cases}$$

To construct an LR test, we note that, under H_0, T is sufficient for μ. Hence, under H_0, we may find a $c(T)$ such that $P(\lambda < c(T)|T) = \alpha$ for every T. The test rejecting H_0 when $\lambda < c(T)$ has size α, since $P(\lambda < c(T)) = \alpha$ under H_0.

Note that

$$s(\mu,\gamma) = \frac{\partial \log\ell(\mu,\gamma)}{\partial(\mu,\gamma)} = \begin{pmatrix} -\frac{n}{2\mu} + \frac{nT}{2\gamma\mu^2} - \frac{n}{2\gamma} \\ -\frac{n}{2\gamma} + \frac{nT}{2\gamma^2\mu} - \frac{n\bar X}{\gamma^2} + \frac{n\mu}{2\gamma^2} \end{pmatrix},$$

$$\frac{\partial^2 \log\ell(\mu,\gamma)}{\partial\mu^2} = \frac{n}{2\mu^2} - \frac{nT}{\gamma\mu^3},$$

$$\frac{\partial^2 \log\ell(\mu,\gamma)}{\partial\mu\partial\gamma} = -\frac{nT}{2\gamma^2\mu^2} + \frac{n}{2\gamma^2},$$

and

$$\frac{\partial^2 \log\ell(\mu,\gamma)}{\partial\gamma^2} = \frac{n}{2\gamma^2} - \frac{nT}{\gamma^3\mu} + \frac{2n\bar X}{\gamma^3} - \frac{n\mu}{\gamma^3}.$$

Hence, the Fisher information about (μ,γ) is

$$I_n(\mu,\gamma) = n\begin{pmatrix} \frac{1}{2\mu^2} + \frac{1}{\gamma\mu} & \frac{1}{2\gamma\mu} \\ \frac{1}{2\gamma\mu} & \frac{1}{2\gamma^2} \end{pmatrix}$$

and Rao's score test statistic is

$$R_n = [s(\hat\mu_0, 1)]^\tau [I_n(\hat\mu_0, 1)]^{-1} s(\hat\mu_0, 1).$$

Similar to that in part (ii), Wald's test statistic W_n is equal to $(\hat{\gamma}-1)^2$ divided by the last element of the inverse of $I_n(\hat{\mu},\hat{\gamma})$, i.e.,

$$W_n = \frac{n(\hat{\gamma}-1)^2}{2\hat{\gamma}^2 + \hat{\gamma}^3/\hat{\mu}}.$$

Note that W_n is not defined when $\bar{X} \leq 0$. But $\lim_n P(\bar{X} \leq 0) = 0$ since $\mu > 0$. ∎

Exercise 48 (#6.100). Suppose that $X = (X_1, ..., X_k)$ has the multinomial distribution with a known size n and an unknown probability vector $(p_1, ..., p_k)$. Consider the problem of testing $H_0 : (p_1, ..., p_k) = (p_{01}, ..., p_{0k})$ versus $H_1 : (p_1, ..., p_k) \neq (p_{01}, ..., p_{0k})$, where $(p_{01}, ..., p_{0k})$ is a known probability vector. Find the forms of Wald's test and Rao's score test.

Solution. The MLE of $\theta = (p_1, ..., p_{k-1})$ is $\hat{\theta} = (X_1/n, ..., X_{k-1}/n)$. The Fisher information about θ is

$$I_n(\theta) = n[D(\theta)]^{-1} + \frac{n}{p_k} J_{k-1} J_{k-1}^\tau,$$

where $D(\theta)$ denotes the $(k-1)\times(k-1)$ diagonal matrix whose $k-1$ diagonal elements are the components of the vector θ and J_{k-1} is the $(k-1)$-vector of 1's. Let $\theta_0 = (p_{01}, ..., p_{0(k-1)})$. Then $H_0 : \theta = \theta_0$ and the Wald's test statistic is

$$\begin{aligned} W_n &= (\hat{\theta}-\theta_0)^\tau I_n(\hat{\theta})(\hat{\theta}-\theta_0) \\ &= n(\hat{\theta}-\theta_0)^\tau [D(\hat{\theta})]^{-1}(\hat{\theta}-\theta_0) + \frac{n^2}{X_k}[J_{k-1}^\tau(\hat{\theta}-\theta_0)]^2 \\ &= \sum_{j=1}^{k} \frac{(X_j - np_{0j})^2}{X_j}, \end{aligned}$$

using the fact that $J_{k-1}^\tau(\hat{\theta}-\theta_0) = p_{0k} - X_k/n$. Let $\ell(\theta)$ be the likelihood function. Then

$$s(\theta) = \frac{\partial \log \ell(\theta)}{\partial \theta} = \left(\frac{X_1}{p_1} - \frac{X_k}{p_k}, ..., \frac{X_{k-1}}{p_{k-1}} - \frac{X_k}{p_k} \right).$$

Note that

$$[I_n(\theta)]^{-1} = n^{-1}D(\theta) - n^{-1}\theta\theta^\tau$$

and

$$\theta^\tau s(\theta) = \sum_{j=1}^{k-1} \left(X_j - \frac{X_k}{p_k} \right) = n - X_k - \frac{X_k}{p_k}(1-p_k) = \frac{np_k - X_k}{p_k}.$$

Also,

$$
\begin{aligned}
[s(\theta)]^\tau D(\theta) s(\theta) &= \sum_{j=1}^{k-1} p_j \left(\frac{X_j}{p_j} - \frac{X_k}{p_k} \right)^2 \\
&= \sum_{j=1}^{k-1} p_j \left(\frac{X_j}{p_j} - n \right)^2 + \sum_{j=1}^{k-1} p_j \left(\frac{X_k}{p_k} - n \right)^2 \\
&\quad + 2 \sum_{j=1}^{k-1} p_j \left(\frac{X_j}{p_j} - n \right) \left(n - \frac{X_k}{p_k} \right) \\
&= \sum_{j=1}^{k-1} \frac{(X_j - np_j)^2}{p_j} + (1 - p_k) \frac{(X_k - np_k)^2}{p_k^2} \\
&\quad + 2[(n - X_k) - n(1 - p_k)] \left(n - \frac{X_k}{p_k} \right) \\
&= \sum_{j=1}^{k-1} \frac{(X_j - np_j)^2}{p_j} + [\theta^\tau s(\theta)]^2 - \frac{(X_k - np_k)^2}{p_k} \\
&\quad + \frac{2(X_k - np_k)^2}{p_k} \\
&= \sum_{j=1}^{k} \frac{(X_j - np_j)^2}{p_j} + [\theta^\tau s(\theta)]^2.
\end{aligned}
$$

Hence, Rao's score test statistic is

$$
\begin{aligned}
R_n &= [s(\theta_0)]^\tau [I_n(\theta_0)]^{-1} s(\theta_0) \\
&= n^{-1} [s(\theta_0)]^\tau D(\theta_0) s(\theta_0) - n^{-1} [\theta_0^\tau s(\theta_0)]^2 \\
&= \sum_{j=1}^{k} \frac{(X_j - np_{0j})^2}{np_{0j}}. \quad ∎
\end{aligned}
$$

Exercise 49 (#6.101). Let A and B be two different events in a probability space related to a random experiment. Suppose that n independent trials of the experiment are carried out and the frequencies of the occurrence of the events are given in the following 2×2 contingency table:

	A	A^c
B	X_{11}	X_{12}
B^c	X_{21}	X_{22}

Consider testing $H_0 : P(A) = P(B)$ versus $H_1 : P(A) \neq P(B)$.
(i) Derive the likelihood ratio λ and the limiting distribution of $-2 \log \lambda$

under H_0.
(ii) Find the forms of Wald's test and Rao's score test.
Solution. Let $p_{ij} = E(X_{ij}/n)$, $i = 1, 2$, $j = 1, 2$, and $\theta = (p_{11}, p_{12}, p_{21})$ be the parameter vector ($p_{22} = 1 - p_{11} - p_{12} - p_{21}$). The likelihood function is proportional to

$$\ell(\theta) = p_{11}^{X_{11}} p_{12}^{X_{12}} p_{21}^{X_{21}} (1 - p_{11} - p_{12} - p_{21})^{n - X_{11} - X_{12} - X_{21}}.$$

Note that H_0 is equivalent to $H_0 : p_{21} = p_{12}$.
(i) The MLE of θ is $\hat{\theta} = n^{-1}(X_{11}, X_{12}, X_{21})$. Under H_0, The MLE of p_{11} is still X_{11}/n, but the MLE of $p_{12} = p_{21}$ is $(X_{12} + X_{21})/(2n)$. Then

$$\lambda = \frac{[(X_{12} + X_{21})/2]^{X_{12}+X_{21}}}{X_{12}^{X_{12}} X_{21}^{X_{21}}}.$$

Note that there are two unknown parameters under H_0. By Theorem 6.5 in Shao (2003), under H_0, $-2 \log \lambda \to_d \chi_1^2$.
(ii) Let $R(\theta) = p_{12} - p_{21}$. Then $C(\theta) = \partial R/\partial\theta = (0, 1, -1)$. The Fisher information matrix about θ is

$$I_n(\theta) = n \begin{pmatrix} p_{11}^{-1} & 0 & 0 \\ 0 & p_{12}^{-1} & 0 \\ 0 & 0 & p_{21}^{-1} \end{pmatrix} + \frac{n}{p_{22}} \begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix}$$

with

$$[I_n(\theta)]^{-1} = \frac{1}{n} \begin{pmatrix} p_{11} & 0 & 0 \\ 0 & p_{12} & 0 \\ 0 & 0 & p_{21} \end{pmatrix} - \frac{1}{n}\theta\theta^\tau.$$

Therefore, Wald's test statistic is

$$\begin{aligned} W_n &= [R(\hat{\theta})]^\tau \{[C(\hat{\theta})]^\tau [I_n(\hat{\theta})]^{-1} C(\hat{\theta})\}^{-1} R(\hat{\theta}) \\ &= \frac{n(X_{12} - X_{21})^2}{n(X_{12} + X_{21}) - (X_{12} - X_{21})^2}. \end{aligned}$$

Note that

$$s(\theta) = \frac{\partial \log \ell(\theta)}{\partial \theta} = \begin{pmatrix} \frac{X_{11}}{p_{11}} - \frac{X_{22}}{1 - p_{11} - p_{12} - p_{21}} \\ \frac{X_{12}}{p_{12}} - \frac{X_{22}}{1 - p_{11} - p_{12} - p_{21}} \\ \frac{X_{21}}{p_{21}} - \frac{X_{22}}{1 - p_{11} - p_{12} - p_{21}} \end{pmatrix}$$

and, hence, $s(\theta)$ evaluated at $\tilde{\theta} = n^{-1}(X_{11}, (X_{12} + X_{21})/2, (X_{12} + X_{21})/2)$ is

$$s(\tilde{\theta}) = \begin{pmatrix} 0 \\ \frac{n(X_{12} - X_{21})}{X_{12} + X_{21}} \\ \frac{n(X_{21} - X_{12})}{X_{12} + X_{21}} \end{pmatrix}.$$

Therefore, Rao's score test statistic is

$$R_n = [s(\tilde{\theta})]^\tau [I_n(\tilde{\theta})]^{-1} s(\tilde{\theta}) = \frac{(X_{12} - X_{21})^2}{X_{12} + X_{21}}. \blacksquare$$

Exercise 50 (#6.102). Consider the $r \times c$ contingency table

	1	2	$\cdots$	c
1	X_{11}	X_{12}	$\cdots$	X_{1c}
2	X_{21}	X_{22}	$\cdots$	X_{2c}
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
r	X_{r1}	X_{r2}	$\cdots$	X_{rc}

with unknown $p_{ij} = E(X_{ij})/n$, where n is a known positive integer.
(i) Let $A_1, ..., A_c$ be disjoint events with $A_1 \cup \cdots \cup A_c = \Omega$ (the sample space of a random experiment), and let $B_1, ..., B_r$ be disjoint events with $B_1 \cup \cdots \cup B_r = \Omega$. Suppose that X_{ij} is the frequency of the occurrence of $A_j \cap B_i$ in n independent trials of the experiment. Derive the χ^2 goodness-of-fit test for testing independence of $\{A_1, ..., A_c\}$ and $\{B_1, ..., B_r\}$, i.e.,

$$H_0 : p_{ij} = p_{i\cdot} p_{\cdot j} \text{ for all } i, j \quad \text{versus} \quad H_1 : p_{ij} \neq p_{i\cdot} p_{\cdot j} \text{ for some } i, j,$$

where $p_{i\cdot} = P(B_i)$ and $p_{\cdot j} = P(A_j)$, $i = 1, ..., r$, $j = 1, ..., c$.
(ii) Let $(X_{1j}, ..., X_{rj})$, $j = 1, ..., c$, be c independent random vectors having the multinomial distributions with sizes n_j and unknown probability vectors $(p_{1j}, ..., p_{rj})$, $j = 1, ..., c$, respectively. Consider the problem of testing whether c multinomial distributions are the same, i.e.,

$$H_0 : p_{ij} = p_{i1} \text{ for all } i, j \quad \text{versus} \quad H_1 : p_{ij} \neq p_{i1} \text{ for some } i, j.$$

Show that the χ^2 goodness-of-fit test is the same as that in (i).
Solution. (i) Using the Lagrange multiplier method, we can obtain the MLE of p_{ij}'s by maximizing

$$\sum_{i=1}^r \sum_{j=1}^c X_{ij} \log p_{ij} - \lambda \left(\sum_{i=1}^r \sum_{j=1}^c p_{ij} - 1 \right),$$

where λ is the Lagrange multiplier. Thus, the MLE of p_{ij} is X_{ij}/n. Under H_0, the MLE's of $p_{i\cdot}$'s and $p_{\cdot j}$'s can be obtained by maximizing

$$\sum_{i=1}^r \sum_{j=1}^c X_{ij} (\log p_{i\cdot} + \log p_{\cdot j}) - \lambda_1 \left(\sum_{i=1}^r p_{i\cdot} - 1 \right) - \lambda_2 \left(\sum_{j=1}^c p_{\cdot j} - 1 \right),$$

where λ_1 and λ_2 are the Lagrange multipliers. Thus, the MLE of $p_{i\cdot}$ is $\bar{X}_{i\cdot} = \sum_{j=1}^{c} X_{ij}/n$ and the MLE of $p_{\cdot j}$ is $\bar{X}_{\cdot j} = \sum_{i=1}^{r} X_{ij}/n$. Hence, the χ^2 statistic is

$$\chi^2 = \sum_{i=1}^{r}\sum_{j=1}^{c} \frac{(X_{ij} - n\bar{X}_{i\cdot}\bar{X}_{\cdot j})^2}{n\bar{X}_{i\cdot}\bar{X}_{\cdot j}}.$$

The number of free parameters is $rc-1$. Under H_0, the number of free parameters is $r-1+c-1 = r+c-2$. The difference of the two is $rc-r-c+1 = (r-1)(c-1)$. By Theorem 6.9 in Shao (2003), under H_0, $\chi^2 \to_d \chi^2_{(r-1)(c-1)}$. Therefore, the χ^2 goodness-of-fit test rejects H_0 when $\chi^2 > \chi^2_{(r-1)(c-1),\alpha}$, where $\chi^2_{(r-1)(c-1),\alpha}$ is the $(1-\alpha)$th quantile of the chi-square distribution $\chi^2_{(r-1)(c-1)}$.
(ii) Since $(X_{1j},...,X_{rj})$ has the multinomial distribution with size n_j and probability vector $(p_{1j},...,p_{rj})$, the MLE of p_{ij} is X_{ij}/n. Let $Y_i = \sum_{j=1}^{c} X_{ij}$. Under H_0, $(Y_1,...,Y_r)$ has the multinomial distribution with size n and probability vector $(p_{11},...,p_{r1})$. Hence, the MLE of p_{i1} under H_0 is $\bar{X}_{i\cdot} = Y_i/n$. Note that $n_j = n\bar{X}_{\cdot j}$, $j = 1,...,c$. Hence, under H_0, the expected (i,j)th frequency estimated by the MLE under H_0 is $n\bar{X}_{i\cdot}\bar{X}_{\cdot j}$. Thus, the χ^2 statistic is the same as that in part (i) of the solution. The number of free parameters in this case is $c(r-1)$. Under H_0, the number of free parameters is $r-1$. The difference of the two is $c(r-1)-(r-1) = (r-1)(c-1)$. Hence, $\chi^2 \to_d \chi^2_{(r-1)(c-1)}$ under H_0 and the χ^2 goodness-of-fit test is the same as that in (i). ∎

Exercise 51 (#6.103). In Exercise 50(i), derive Wald's test and Rao's score test statistics.
Solution. For a set $\{a_{ij}, i = 1,...,r, j = 1,...,c, (i,j) \neq (r,c)\}$ of $rc-1$ numbers, we denote $\text{vec}(a_{ij})$ to be the $(rc-1)$-vector whose components are a_{ij}'s and $D(a_{ij})$ to be the diagonal matrix whose diagonal elements are the components of $\text{vec}(a_{ij})$. Let $\theta = \text{vec}(p_{ij})$, J be the $(rc-1)$-vector of 1's, and $\ell(\theta)$ be the likelihood function. Then,

$$s(\theta) = \frac{\partial \log \ell(\theta)}{\partial \theta} = \text{vec}\left(\frac{X_{ij}}{p_{ij}}\right) - \frac{X_{rc}}{1 - J^\tau\theta} J$$

and

$$\frac{\partial^2 \log \ell(\theta)}{\partial\theta\partial\theta^\tau} = -D\left(\frac{X_{ij}}{p_{ij}^2}\right) - \frac{X_{rc}}{(1-J^\tau\theta)^2} JJ^\tau.$$

Since $E(X_{ij}) = np_{ij}$, the Fisher information about θ is

$$I_n(\theta) = nD(p_{ij}^{-1}) + np_{rc}^{-1}JJ^\tau$$

with

$$[I_n(\theta)]^{-1} = n^{-1}D(p_{ij}) - n^{-1}\theta\theta^\tau.$$

Let $\tilde{p}_{ij}$ be the MLE of p_{ij} under H_0 and $\tilde{\theta} = \text{vec}(\tilde{p}_{ij})$. Then Rao's score test statistic is

$$R_n = n^{-1}[s(\tilde{\theta})]^\tau [D(\tilde{p}_{ij}) - \tilde{\theta}\tilde{\theta}^\tau] s(\tilde{\theta}).$$

Note that

$$\tilde{\theta}^\tau s(\tilde{\theta}) = \sum_{(i,j)\neq(r,c)} \left(X_{ij} - \frac{X_{rc}\tilde{p}_{ij}}{1 - J^\tau\tilde{\theta}} \right) = n - X_{rc} - \frac{X_{rc}J^\tau\tilde{\theta}}{1 - J^\tau\tilde{\theta}} = n - \frac{X_{rc}}{\tilde{p}_{rc}},$$

where $\tilde{p}_{rc} = 1 - J^\tau\tilde{\theta} = \bar{X}_{r\cdot}\bar{X}_{\cdot c}$. Also,

$$\begin{aligned}
[s(\tilde{\theta})]^\tau D(\tilde{p}_{ij}) s(\tilde{\theta}) &= \sum_{(i,j)\neq(r,c)} \tilde{p}_{ij} \left(\frac{X_{ij}}{\tilde{p}_{ij}} - \frac{X_{rc}}{\tilde{p}_{rc}} \right)^2 \\
&= \sum_{(i,j)\neq(r,c)} \tilde{p}_{ij} \left(\frac{X_{ij}}{\tilde{p}_{ij}} - n \right)^2 + (1 - \tilde{p}_{rc}) \left(n - \frac{X_{rc}}{\tilde{p}_{rc}} \right)^2 \\
&\quad + 2 \sum_{(i,j)\neq(r,c)} \tilde{p}_{ij} \left(\frac{X_{ij}}{\tilde{p}_{ij}} - n \right) \left(n - \frac{X_{rc}}{\tilde{p}_{rc}} \right) \\
&= \sum_{(i,j)\neq(r,c)} \tilde{p}_{ij} \left(\frac{X_{ij}}{\tilde{p}_{ij}} - n \right)^2 + (1 + \tilde{p}_{rc}) \left(n - \frac{X_{rc}}{\tilde{p}_{rc}} \right)^2.
\end{aligned}$$

Hence,

$$\begin{aligned}
R_n &= \frac{1}{n} \sum_{(i,j)\neq(r,c)} \tilde{p}_{ij} \left(\frac{X_{ij}}{\tilde{p}_{ij}} - n \right)^2 + \tilde{p}_{rc} \left(n - \frac{X_{rc}}{\tilde{p}_{rc}} \right)^2 \\
&= \frac{1}{n} \sum_{i=1}^r \sum_{j=1}^c \tilde{p}_{ij} \left(\frac{X_{ij}}{\tilde{p}_{ij}} - n \right)^2 \\
&= \sum_{i=1}^r \sum_{j=1}^c \frac{(X_{ij} - n\bar{X}_{i\cdot}\bar{X}_{\cdot j})^2}{n\bar{X}_{i\cdot}\bar{X}_{\cdot j}},
\end{aligned}$$

using the fact that $\tilde{p}_{ij} = \bar{X}_{i\cdot}\bar{X}_{\cdot j}$ (part (i) of the solution to Exercise 50). Hence, R_n is the same as χ^2 in part (i) of the solution to Exercise 50.

Let $\eta(\theta)$ be the $(r-1)(c-1)$-vector obtained by deleting components p_{rj} and p_{ic}, $j = 1, ..., c-1$, $i = 1, ..., r-1$, from the vector θ and let $\zeta(\theta)$ be $\eta(\theta)$ with p_{ij} replaced by $p_{i\cdot}p_{\cdot j}$, $i = 1, ..., r-1$, $j = 1, ..., c-1$. Let $\hat{\theta} = \text{vec}(X_{ij}/n)$, the MLE of θ, $R(\theta) = \eta(\theta) - \zeta(\theta)$, and $C(\theta) = \frac{\partial\eta}{\partial\theta} - \frac{\partial\zeta}{\partial\theta}$. Then, Wald's test statistic is

$$W_n = [R(\hat{\theta})]^\tau \{[C(\hat{\theta})]^\tau [I_n(\hat{\theta})]^{-1} C(\hat{\theta})\}^{-1} R(\hat{\theta}). \quad \blacksquare$$

Exercise 52 (#6.105). Let $(X_1, ..., X_n)$ be a random sample of binary random variables with $\theta = P(X_1 = 1)$.
(i) Let $\Pi(\theta)$ be the beta distribution with parameter (a, b). When Π is used as the prior for θ, find the Bayes factor and the Bayes test for $H_0 : \theta \leq \theta_0$ versus $H_1 : \theta > \theta_0$.
(ii) Let $\pi_0 I_{[\theta_0,\infty)}(\theta) + (1 - \pi_0)\Pi(\theta)$ be the prior cumulative distribution, where Π is the same as that in (i) and $\pi_0 \in (0, 1)$ is a constant. Find the Bayes factor and the Bayes test for $H_0 : \theta = \theta_0$ versus $H_1 : \theta \neq \theta_0$.
Solution. (i) Under prior Π, the posterior of θ is the beta distribution with parameter $(a+T, b+n-T)$, where $T = \sum_{i=1}^n X_i$. Then, the posterior probability of the set $(0, \theta_0]$ (the null hypothesis H_0) is

$$p(T) = \frac{\Gamma(a+b+n)}{\Gamma(a+T)\Gamma(b+n-T)} \int_0^{\theta_0} u^{a+T-1}(1-u)^{b+n-T-1}du.$$

Hence, the Bayes test rejects H_0 if and only if $p(T) < \frac{1}{2}$ and the Bayes factor is

$$\frac{\text{posterior odds ratio}}{\text{prior odds ratio}} = \frac{p(T)[1-\pi(0)]}{[1-p(T)]\pi(0)},$$

where

$$\pi(0) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \int_0^{\theta_0} u^{a-1}(1-u)^{b-1}du$$

is the prior probability of the set $(0, \theta_0]$.
(ii) Let

$$\begin{aligned} m_1(T) &= \int_{\theta \neq \theta_0} \theta^T (1-\theta)^{n-T} d\Pi(\theta) \\ &= \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \int_0^1 \theta^{a+T-1}(1-\theta)^{b+n-T-1}d\theta \\ &= \frac{\Gamma(a+b)\Gamma(a+T)\Gamma(b+n-T)}{\Gamma(a+b+n)\Gamma(a)\Gamma(b)}. \end{aligned}$$

Since the likelihood function is $\ell(\theta) = \binom{n}{T}\theta^T(1-\theta)^{n-T}$, the posterior probability of the set $\{\theta_0\}$ (the null hypothesis H_0) is

$$\begin{aligned} p(T) &= \frac{\int_{\theta=\theta_0} \ell(\theta)d[\pi_0 I_{[\theta_0,\infty)}(\theta) + (1-\pi_0)\Pi(\theta)]}{\int \ell(\theta)d[\pi_0 I_{[\theta_0,\infty)}(\theta) + (1-\pi_0)\Pi(\theta)]} \\ &= \frac{\pi_0\theta_0^T(1-\theta_0)^{n-T}}{\pi_0\theta_0^T(1-\theta_0)^{n-T} + (1-\pi_0)m_1(T)}. \end{aligned}$$

Hence, the Bayes test rejects H_0 if and only if $p(T) < \frac{1}{2}$ and and the Bayes factor is $\frac{p(T)(1-\pi_0)}{[1-p(T)]\pi_0}$. ∎

Exercise 53 (#6.114). Let $(X_1, ..., X_n)$ be a random sample from a continuous distribution F on $\mathcal{R}$, F_n be the empirical distribution, $D_n^+ = \sup_{x\in\mathcal{R}}[F_n(x) - F(x)]$, and $D_n^- = \sup_{x\in\mathcal{R}}[F(x) - F_n(x)]$. Show that $D_n^-(F)$ and $D_n^+(F)$ have the same distribution and, for $t \in (0,1)$,

$$P\big(D_n^+(F) \le t\big) = n! \prod_{i=1}^n \int_{\max\{0,\frac{n-i+1}{n}-t\}}^{u_{n-i+2}} du_1 \cdots du_n.$$

Proof. Let $X_{(i)}$ be the ith order statistic, $i = 1, ..., n$, $X_{(0)} = -\infty$, and $X_{(n+1)} = \infty$. Note that

$$\begin{aligned} D_n^+(F) &= \max_{0\le i\le n} \sup_{X_{(i)}\le x< X_{(i+1)}} \left[\frac{i}{n} - F(x)\right] \\ &= \max_{0\le i\le n} \left[\frac{i}{n} - \inf_{X_{(i)}\le x< X_{(i+1)}} F(x)\right] \\ &= \max_{0\le i\le n} \left[\frac{i}{n} - F(X_{(i)})\right] \\ &= \max_{0\le i\le n} \left[\frac{i}{n} - U_{(i)}\right], \end{aligned}$$

where $U_{(i)} = F(X_{(i)})$ is the ith order statistic of a random sample of size n from the uniform distribution on $(0, 1)$. Similarly,

$$\begin{aligned} D_n^-(F) &= \max_{0\le j\le n} \sup_{X_{(j)}\le x< X_{(j+1)}} \left[F(x) - \frac{j}{n}\right] \\ &= \max_{0\le j\le n} \left[\sup_{X_{(j)}\le x< X_{(j+1)}} F(x) - \frac{j}{n}\right] \\ &= \max_{0\le j\le n} \left[U_{(j+1)} - \frac{j}{n}\right] \\ &= \max_{0\le i\le n} \left[\frac{i}{n} + U_{(n-i+1)} - 1\right]. \end{aligned}$$

Since $(1 - U_{(1)}, ..., 1 - U_{(n)})$ has the same distribution as $(U_{(n)}, ..., U_{(1)})$, we conclude that $D_n^-(F)$ and $D_n^+(F)$ have the same distribution. For $t \in (0, 1)$,

$$\begin{aligned} P(D_n^+(F) \le t) &= P\left(\max_{0\le i\le n} \left[\frac{i}{n} - U_{(i)}\right] \le t\right) \\ &= P\left(U_{(i)} \ge \frac{i}{n} - t, i = 1, ..., n\right) \\ &= n! \prod_{i=1}^n \int_{\max\{0,\frac{n-i+1}{n}-t\}}^{u_{n-i+2}} du_1 \cdots du_n. \quad \blacksquare \end{aligned}$$

Exercise 54 (#6.116). Let $(X_1, ..., X_n)$ be a random sample from a continuous distribution F on $\mathcal{R}$, F_n be the empirical distribution, and $C_n(F) = \int [F_n(x) - F(x)]^2 dF(x)$. Show that the distribution of $C_n(F)$ does not vary with F.

Solution. Note that

$$\begin{aligned} C_n(F) &= \int_0^1 [F_n(F^{-1}(t)) - t]^2 dt \\ &= \int_0^1 \left[\frac{1}{n} \sum_{i=1}^n I_{(-\infty, F^{-1}(t)]}(X_i) - t \right]^2 dt \\ &= \int_0^1 \left[\frac{1}{n} \sum_{i=1}^n I_{(-\infty, t]}(F(X_i)) - t \right]^2 dt, \end{aligned}$$

where the last equality follows from the fact that $X_i \leq F^{-1}(t)$ if and only if $F(X_i) \leq t$. Since $(F(X_1), ..., F(X_n))$ is a random sample from the uniform distribution on $(0,1)$, the distribution of $C_n(F)$ does not depend on F. ∎

Exercise 55 (#6.123). Let $\hat{\theta}_n$ be an estimator of a real-valued parameter θ such that $V_n^{-1/2}(\hat{\theta}_n - \theta) \to_d N(0,1)$ for any θ and let $\hat{V}_n$ be a consistent estimator of V_n. Suppose that $V_n \to 0$.

(i) Show that the test with rejection region $\hat{V}_n^{-1/2}(\hat{\theta}_n - \theta_0) > z_\alpha$ is a consistent asymptotic level α test for testing $H_0 : \theta \leq \theta_0$ versus $H_1 : \theta > \theta_0$, where z_α is the $(1-\alpha)$th quantile of $N(0,1)$.

(ii) Apply the result in (i) to show that the one-sample one-sided t-test for the testing problem in (i) is a consistent asymptotic level α test.

Solution. (i) Under $H_1 : \theta > \theta_0$, $\hat{V}_n^{-1/2}(\hat{\theta}_n - \theta) \to_d N(0,1)$. Therefore, the test with rejection region $\hat{V}_n^{-1/2}(\hat{\theta}_n - \theta_0) > z_\alpha$ is an asymptotic level α test. Also,

$$P\left(\hat{V}_n^{-1/2}(\hat{\theta}_n - \theta_0) > z_\alpha\right) = P\left(\hat{V}_n^{-1/2}(\hat{\theta}_n - \theta) > z_\alpha - \hat{V}_n^{-1/2}(\theta - \theta_0)\right) \to 1$$

as $n \to \infty$, since $\hat{V}_n^{-1/2}(\theta - \theta_0) \to_p \infty$. Hence, the test is consistent.

(ii) Let $(X_1, ..., X_n)$ be a random sample from a population with finite mean θ and variance σ^2. For testing $H_0 : \theta \leq \theta_0$ versus $H_1 : \theta > \theta_0$, the one-sample t-test rejects H_0 if and only if $t(X) = \sqrt{n}(\bar{X} - \theta_0)/S > t_{n-1,\alpha}$, where $\bar{X}$ and S^2 are the sample mean and variance and $t_{n-1,\alpha}$ is the $(1-\alpha)$th quantile of the t-distribution t_{n-1}. By the central limit theorem, $\sqrt{n}(\bar{X} - \theta) \to_d N(0, \sigma^2)$. Hence $V_n = \sigma^2/n \to 0$. By the law of large numbers, $S^2 \to_p \sigma^2$. Hence $\hat{V}_n = S^2/n$ is a consistent estimator of V_n. Note that the t-distribution t_{n-1} converges to $N(0,1)$. Then, by the result in Exercise 28 of Chapter 5, $\lim_n t_{n-1,\alpha} = z_\alpha$. From the result in (i), the one-sample t-test is a consistent asymptotic level α test. ∎

Exercise 56 (#6.124). Let $(X_1, ..., X_n)$ be a random sample from the gamma distribution with shape parameter θ and scale parameter γ, where $\theta > 0$ and $\gamma > 0$ are unknown. Let $T_n = n\sum_{i=1}^n X_i^2/(\sum_{i=1}^n X_i)^2$. Show how to use T_n to obtain an asymptotic level α and consistent test for testing $H_0 : \theta = 1$ versus $H_1 : \theta \neq 1$.

Solution. From the central limit theorem,

$$\sqrt{n}\left[\begin{pmatrix} \frac{1}{n}\sum_{i=1}^n X_i \\ \frac{1}{n}\sum_{i=1}^n X_i^2 \end{pmatrix} - \begin{pmatrix} E(X_1) \\ E(X_1^2) \end{pmatrix}\right] \to_d N_2(0, \Sigma),$$

where

$$\Sigma = \begin{pmatrix} E(X_1^2) - [E(X_1)]^2 & E(X_1^3) - E(X_1)E(X_1^2) \\ E(X_1^3) - E(X_1)E(X_1^2) & E(X_1^4) - [E(X_1^2)]^2 \end{pmatrix}.$$

Since the moment generating function of the gamma distribution is $g(t) = (1-\gamma t)^{-\theta}$,

$$E(X_1) = g'(0) = \theta\gamma,$$

$$E(X_1^2) = g''(0) = \theta(\theta+1)\gamma^2,$$

$$E(X_1^3) = g'''(0) = \theta(\theta+1)(\theta+2)\gamma^3,$$

and

$$E(X_1^4) = g''''(0) = \theta(\theta+1)(\theta+2)(\theta+3)\gamma^4.$$

Hence,

$$\Sigma = \begin{pmatrix} \theta\gamma^2 & 2\theta(\theta+1)\gamma^3 \\ 2\theta(\theta+1)\gamma^3 & 2\theta(\theta+1)(2\theta+3)\gamma^4 \end{pmatrix}.$$

Note that

$$T = \frac{\frac{1}{n}\sum_{i=1}^n X_i^2}{\left(\frac{1}{n}\sum_{i=1}^n X_i\right)^2} = h\left(\frac{1}{n}\sum_{i=1}^n X_i, \frac{1}{n}\sum_{i=1}^n X_i^2\right)$$

with $h(x,y) = y/x^2$,

$$H(x,y) = \begin{pmatrix} \frac{\partial h}{\partial x} \\ \frac{\partial h}{\partial y} \end{pmatrix} = \begin{pmatrix} -\frac{2y}{x^3} \\ \frac{1}{x^2} \end{pmatrix},$$

and

$$H(E(X_1), E(X_1^2)) = H(\theta\gamma, \theta(\theta+1)\gamma^2) = \begin{pmatrix} -\frac{2(\theta+1)}{\theta^2\gamma} \\ \frac{1}{\theta^2\gamma^2} \end{pmatrix}.$$

By the δ-method,

$$\sqrt{n}[T - h(E(X_1), E(X_1^2))] \to_d N(0, \sigma^2),$$

where

$$\begin{aligned}\sigma^2 &= [H(\theta\gamma, \theta(\theta+1)\gamma^2)]^\tau \Sigma H(\theta\gamma, \theta(\theta+1)\gamma^2) \\ &= \frac{4(\theta+1)^2}{\theta^3} + \frac{2(\theta+1)(2\theta+3)}{\theta^3} - \frac{8(\theta+1)^2}{\theta^3} \\ &= \frac{2(\theta+1)}{\theta^3}.\end{aligned}$$

Note that

$$h(E(X_1), E(X_1^2)) = h(\theta\gamma, \theta(\theta+1)\gamma^2) = \frac{\theta(\theta+1)\gamma^2}{\theta^2\gamma^2} = 1 + \frac{1}{\theta}.$$

Combining all the results, we obtain that

$$V_n^{-1/2}\left(T - 1 - \frac{1}{\theta}\right) \to_d N(0, 1)$$

with $V_n = 2(\theta+1)/(\theta^3 n)$. From the asymptotic normality of T,

$$\frac{1}{T-1} \to_p \theta.$$

Hence, a consistent estimator of V_n is

$$\hat{V}_n = \frac{2\left(\frac{1}{T-1} + 1\right)}{\frac{n}{(T-1)^3}} = \frac{2T(T-1)^2}{n}.$$

From Theorem 6.12 in Shao (2003), an asymptotic level α and consistent test for $H_0 : \theta = 1$ versus $H_1 : \theta \neq 1$ rejects H_0 if and only if

$$\hat{V}_n^{-1}(T-2)^2 > \chi^2_{1,\alpha},$$

which is the same as

$$\frac{n(T-2)^2}{2T(T-1)^2} > \chi^2_{1,\alpha},$$

where $\chi^2_{1,\alpha}$ is the $(1-\alpha)$th quantile of the chi-square distribution χ^2_1. ∎

Chapter 7

Confidence Sets

Exercise 1 (#7.4). Let (X_{i1}, X_{i2}), $i = 1, ..., n$, be a random sample from the bivariate normal distribution with unknown $\mu_j = E(X_{1j})$, $\sigma_j^2 = \text{Var}(X_{1j})$, $j = 1, 2$, and $\sigma_{12} = \text{Cov}(X_{11}, X_{12})$. Let $\theta = \mu_2/\mu_1$ be the parameter of interest ($\mu_1 \neq 0$), $Y_i(\theta) = X_{i2} - \theta X_{i1}$, and

$$S^2(\theta) = \frac{1}{n-1}\sum_{i=1}^{n}[Y_i(\theta) - \bar{Y}(\theta)]^2 = S_2^2 - 2\theta S_{12} + \theta^2 S_1^2,$$

where $\bar{Y}(\theta)$ is the average of $Y_i(\theta)$'s and S_i^2 and S_{12} are sample variances and covariance based on X_{ij}'s. Discuss when Fieller's confidence set for θ,

$$C(X) = \{\theta : n[\bar{Y}(\theta)]^2/S^2(\theta) \leq t_{n-1,\alpha/2}^2\},$$

is a finite interval, the complement of a finite interval, or the whole real line.

Solution. Note that $n\bar{Y}^2(\theta)/S^2(\theta) \leq t_{n-1,\alpha/2}^2$ is equivalent to

$$a\theta^2 + b\theta + c \geq 0,$$

where $a = t_{n-1,\alpha/2}^2 S_1^2 - n\bar{X}_1^2$, $b = 2(n\bar{X}_1\bar{X}_2 - t_{n-1,\alpha/2}^2 S_{12})$, and $c = t_{n-1,\alpha/2}^2 S_2^2 - n\bar{X}_2^2$. Then the confidence set $C(X)$ represents the whole real line if $a > 0$ and $b^2 - 4ac < 0$; the complement of a finite interval if $a > 0$ and $b^2 - 4ac \geq 0$; a finite interval if $a < 0$ and $b^2 - 4ac \geq 0$. ∎

Exercise 2 (#7.6). Let $X = (X_1, ..., X_n)$ be a random sample from $N(\theta, \theta)$ with an unknown $\theta > 0$. Find a pivotal quantity and use it to construct a confidence interval for θ.

Solution. Let $\bar{X}$ be the sample mean and $R(X, \theta) = n(\bar{X} - \theta)^2/\theta$. Since $\bar{X}$ is distributed as $N(\theta, \theta/n)$, $R(X, \theta)$ has the chi-square distribution χ_1^2

and, thus, is a pivotal quantity. Let $\chi^2_{1,\alpha}$ be the $(1-\alpha)$th quantile of χ^2_1 and $c = \chi^2_{1,\alpha}/(2n)$. Then

$$C(X) = \left\{\theta : R(X,\theta) \le \chi^2_{1,\alpha}\right\} = \left\{\theta : \theta^2 - 2(\bar{X}+c)\theta + \bar{X}^2 \le 0\right\}$$

is a confidence set for θ with confidence coefficient $1-\alpha$. If $\bar{X} \ge -c/2$, then

$$C(X) = \left[\bar{X} + c - \sqrt{2c\bar{X}+c^2}, \bar{X} + c + \sqrt{2c\bar{X}+c^2}\,\right]$$

is an interval. If $\bar{X} < -c/2$, then $C(X)$ is the empty set. ∎

Exercise 3 (#7.7). Let T be a random variable having cumulative distribution function F. For any $t \in (0,1)$, show that $P(F(T) < t) \le t$ and $P(F(T-) > 1-t) \le t$, where $F(x-)$ denotes the left limit of F at x.
Solution. Let $F^{-1}(t) = \inf\{x : F(x) \ge t\}$. If $F(T) < t$, then, by definition, $T < F^{-1}(t)$. Thus,

$$P(F(T) < t) \le P(T < F^{-1}(t)) = F(F^{-1}(t)-) \le t,$$

since $F(x) < t$ for all $x < F^{-1}(t)$ so that the left limit of F at $F^{-1}(t)$ is no larger than t. Similarly, $F(T-) > 1-t$ implies $T > F^{-1}(1-t)$ and, hence,

$$P(F(T-) > 1-t) \le P(T > F^{-1}(1-t)) = 1 - F(F^{-1}(1-t)) \le t,$$

since $F(F^{-1}(1-t)) \ge 1-t$. ∎

Exercise 4 (#7.9). Let $(X_1,...,X_n)$ be a random sample of random variables having Lebesgue density $\frac{a}{\theta}\left(\frac{x}{\theta}\right)^{a-1} I_{(0,\theta)}(x)$, where $a \ge 1$ is known and $\theta > 0$ is unknown.
(i) Construct a confidence interval for θ with confidence coefficient $1-\alpha$ using the cumulative distribution function of the largest order statistic $X_{(n)}$.
(ii) Show that the confidence interval in (i) can also be obtained using a pivotal quantity.
Solution. (i) The cumulative distribution function of $X_{(n)}$ is

$$F_\theta(t) = \begin{cases} 0 & t \le 0 \\ (t/\theta)^{na} & 0 < t < \theta \\ 1 & t \ge \theta, \end{cases}$$

which is decreasing in θ for any fixed $t \in (0,\theta)$. Also, for $t > 0$, $\lim_{\theta\to 0} F_\theta(t) = 1$ and $\lim_{\theta\to\infty} F_\theta(t) = 0$. By Theorem 7.1 in Shao (2003), a $1-\alpha$ confidence interval for θ has upper limit being the unique solution of $F_\theta(T) = \alpha_1$ and lower limit being the unique solution of $F_\theta(T) = 1-\alpha_2$, where $\alpha_1+\alpha_2 = \alpha$. Consequently, this confidence interval is $[T/(1-\alpha_2)^{(na)^{-1}}, T/\alpha_1^{(na)^{-1}}]$.
(ii) Note that $U(\theta) = \left(X_{(n)}/\theta\right)^{na}$ has the uniform distribution on $(0,1)$

and, hence, is a pivotal quantity. The $1-\alpha$ confidence interval constructed using $U(\theta)$ is the same as that in part (i) of the solution. ∎

Exercise 5 (#7.10). Let $(X_1, ..., X_n)$ be a random sample from the exponential distribution on (a, ∞) with scale parameter 1, where $a \in \mathcal{R}$ is unknown.
(i) Construct a confidence interval for a with confidence coefficient $1-\alpha$ by using the cumulative distribution function of the smallest order statistic $X_{(1)}$.
(ii) Show that the confidence interval in (i) can also be obtained using a pivotal quantity.
Solution. (i) The cumulative distribution function of $X_{(1)}$ is

$$F_a(t) = \begin{cases} 0 & t \leq a \\ 1 - e^{-n(t-a)} & t > a, \end{cases}$$

which is decreasing in a for fixed $t > a$. By Theorem 7.1 in Shao (2003), a $1-\alpha$ confidence interval for a has upper limit being the unique solution of $F_a(T) = \alpha_1$ and lower limit being the unique solution of $F_a(T) = 1-\alpha_2$, where $\alpha_1 + \alpha_2 = \alpha$. Then, $[T + n^{-1}\log(\alpha_2), T + n^{-1}\log(1-\alpha_1)]$ is the resulting confidence interval.
(ii) Note that $W(a) = n(X_{(1)} - a)$ has the exponential distribution on $(0, 1)$ with scale parameter 1 and, hence, it is a pivotal quantity. The $1-\alpha$ confidence interval for a constructed using $W(a)$ is the same as that derived in part (i) of the solution. ∎

Exercise 6 (#7.11). Let X be a single observation from the uniform distribution on $(\theta - \frac{1}{2}, \theta + \frac{1}{2})$, where $\theta \in \mathcal{R}$ is unknown.
(i) Show that $X - \theta$ is a pivotal quantity and that a confidence interval of the form $[X+c, X+d]$ with some constants $-\frac{1}{2} < c < d < \frac{1}{2}$ has confidence coefficient $1-\alpha$ if and only if its length is $1-\alpha$.
(ii) Show that the cumulative distribution function $F_\theta(x)$ of X is nonincreasing in θ for any x and it can be used to construct a confidence interval for θ with confidence coefficient $1-\alpha$.
Solution. (i) The distribution of $\theta - X$ is the uniform distribution on $(-\frac{1}{2}, \frac{1}{2})$. Hence, $\theta - X$ is a pivotal quantity. For $-\frac{1}{2} < c < d < \frac{1}{2}$,

$$P(X + c < \theta < X + d) = P(c < \theta - X < d) = d - c.$$

Hence, $[X + c, X + d]$ is a confidence interval with confidence coefficient $1-\alpha$ if and only if $d - c = 1-\alpha$, i.e., the length of $[X + c, X + d]$ is $1-\alpha$.
(ii) For $x \in \mathcal{R}$,

$$F_\theta(x) = \begin{cases} 1 & \theta \leq x - \frac{1}{2} \\ x + \frac{1}{2} - \theta & x - \frac{1}{2} < \theta < x + \frac{1}{2} \\ 0 & x + \frac{1}{2} \leq \theta, \end{cases}$$

which is nonincreasing in θ. By Theorem 7.1 in Shao (2003), a $1-\alpha$ confidence interval for a has upper limit being the unique solution of $F_\theta(X) = \alpha_1$ and lower limit being the unique solution of $F_\theta(X) = 1 - \alpha_2$, where $\alpha_1 + \alpha_2 = \alpha$. This confidence interval is $[x + \alpha_2 - \frac{1}{2}, X + \frac{1}{2} - \alpha_1]$. ∎

Exercise 7 (#7.12). Let $X_1, ..., X_n$ be a random sample of random variables with Lebesgue density $\theta a^\theta x^{-(\theta+1)} I_{(a,\infty)}(x)$, where $\theta > 0$ and $a > 0$.
(i) When θ is known, derive a confidence interval for a with confidence coefficient $1-\alpha$ by using the cumulative distribution function of the smallest order statistic $X_{(1)}$.
(ii) When both a and θ are unknown and $n \geq 2$, derive a confidence interval for θ with confidence coefficient $1-\alpha$ by using the cumulative distribution function of $T = \prod_{i=1}^n (X_i/X_{(1)})$.
(iii) Show that the confidence intervals in (i) and (ii) can be obtained using pivotal quantities.
(iv) When both a and θ are unknown, construct a confidence set for (a, θ) with confidence coefficient $1-\alpha$ by using a pivotal quantity.
Solution. (i) The cumulative distribution function of $X_{(1)}$ is

$$F_a(x) = \begin{cases} 1 - a^{n\theta}/x^{n\theta} & a < x \\ 0 & a > x, \end{cases}$$

which is nonincreasing in a. By Theorem 7.1 in Shao (2003), a $1-\alpha$ confidence interval for a has upper limit being the unique solution of $F_a(X_{(1)}) = \alpha_1$ and lower limit being the unique solution of $F_a(X_{(1)}) = 1 - \alpha_2$, where $\alpha_1 + \alpha_2 = \alpha$. This leads to the interval $[\alpha_2^{(n\theta)^{-1}} X_{(1)}, (1-\alpha_1)^{(n\theta)^{-1}} X_{(1)}]$.
(ii) Consider $Y_i = \log X_i$. Then $(Y_1, ..., Y_n)$ is a random sample from the exponential distribution on the interval (b, ∞) with scale parameter θ^{-1}, where $b = \log a$. From the result in Exercise 7 of Chapter 2 and the fact that $Y_{(1)} = \log X_{(1)}$ is the smallest order statistic of the Y sample, $2\theta \log T = 2\theta \sum_{i=1}^n (Y_i - Y_{(1)})$ has the chi-square distribution $\chi^2_{2(n-1)}$. Hence, the cumulative distribution function of T is

$$P(T \leq t) = P(2\theta \log T \leq 2\theta \log t) = F_{2(n-1)}(2\theta \log t)$$

for $t > 1$, where $F_{2(n-1)}$ denotes the cumulative distribution function of the chi-square distribution $\chi^2_{2(n-1)}$. From Theorem 7.1 in Shao (2003), a $1-\alpha$ confidence interval for θ has upper limit being the unique solution of $F_{2(n-1)}(2\theta \log T) = 1 - \alpha_2$ and lower limit being the unique solution of $F_{2(n-1)}(2\theta \log T) = \alpha_1$, where $\alpha_1 + \alpha_2 = \alpha$. The resulting interval is $[\chi^2_{2(n-1),1-\alpha_1}/(2\log T), \chi^2_{2(n-1),\alpha_2}/(2\log T)]$, where $\chi^2_{2(n-1),\alpha}$ denotes the $(1-\alpha)$th quantile of $\chi^2_{2(n-1)}$.
(iii) When θ is known, $X_{(1)}/a$ is a pivotal quantity. The confidence interval constructed using $X_{(1)}/a$ is the same as that in part (i) of the solution.

When both a and θ are unknown, it follows from part (ii) of the solution that $2\theta \log T$ is a pivotal quantity. The confidence interval constructed using $2\theta \log T$ is the same as that in part (ii) of the solution.
(iv) From part (i) of the solution,

$$P\left(\frac{X_{(1)}^{\theta}}{a^{\theta}} \leq t\right) = P\left(X_{(1)} \leq a t^{1/\theta}\right) = 1 - \frac{1}{t^n}.$$

Hence, a pivotal quantity is $X_{(1)}^{\theta}/a$. Let $0 < c_1 < c_2$ be constants such that $c_1^{-n} - c_2^{-n} = 1 - \alpha$. Then, a $1 - \alpha$ confidence set for (a, θ) is

$$C(X) = \left\{(a, \theta) : c_1 a \leq X_{(1)}^{\theta} \leq c_2 a\right\}. \blacksquare$$

Exercise 8 (#7.13). Let $X = (X_1, ..., X_n)$ be a random sample from the Weibull distribution with Lebesgue density $\frac{a}{\theta} x^{a-1} e^{-x^a/\theta} I_{(0,\infty)}(x)$, where $a > 0$ and $\theta > 0$ are unknown. Show that $R(X, a, \theta) = \prod_{i=1}^n (X_i^a/\theta)$ is pivotal. Construct a confidence set for (a, θ) with confidence coefficient $1 - \alpha$ by using $R(X, a, \theta)$.
Solution. Let $Y_i = X_i^a/\theta$. Then Y_i has the exponential distribution on $(0, \infty)$ with scale parameter 1. Since $Y_1, ..., Y_n$ are independent, $R(X, a, \theta) = \prod_{i=1}^n (X_i^a/\theta)$ is pivotal. Since the distribution of $R(X, a, \theta)$ is known, we can find positive constants c_1 and c_2 such that $P(c_1 \leq R(X, a, \theta) \leq c_2) = 1 - \alpha$. A confidence set for (a, θ) with confidence coefficient $1 - \alpha$ is

$$C(X) = \left\{(a, \theta) : c_1 \theta^n \leq \prod_{i=1}^n X_i^a \leq c_2 \theta^n\right\}. \blacksquare$$

Exercise 9 (#7.14). Let F and G be two known cumulative distribution functions on $\mathcal{R}$ and X be a single observation from the cumulative distribution function $\theta F(x) + (1 - \theta) G(x)$, where $\theta \in [0, 1]$ is unknown. Construct a level $1 - \alpha$ confidence interval for θ based on the observation X. Find a condition under which the derived confidence interval has confidence coefficient $1 - \alpha$.
Solution. Let $f(x)$ and $g(x)$ be the probability densities of F and G, respectively, with respect to the measure induced by $F + G$. From the solution of Exercise 12 in Chapter 6, the family of densities

$$\{\theta f(x) + (1 - \theta) g(x) : \theta \in [0, 1]\}$$

has monotone likelihood ratio in $Y(X) = f(X)/g(X)$. Let $F_{Y,\theta}$ be the cumulative distribution function for $Y(X)$. By Lemma 6.3 in Shao (2003),

$F_{Y,\theta}(X)$ is nonincreasing in θ for any fixed X. Thus, by Theorem 7.1 in Shao (2003), a level $1-\alpha$ confidence interval for θ is $[\underline{\theta}, \overline{\theta}]$, where

$$\overline{\theta} = \sup\{\theta : F_{Y,\theta}(X) \geq \alpha_1\}, \qquad \underline{\theta} = \inf\{\theta : F_{Y,\theta}(X-) \leq 1-\alpha_2\},$$

and $\alpha_1 + \alpha_2 = \alpha$. A sufficient condition for this interval to have confidence coefficient $1-\alpha$ is that $F_{Y,\theta}(x)$ is continuous in x, which is implied by the condition that F' and G' exist and the set $\{x : F'(x) = cG'(x)\}$ has Lebesgue measure 0 for any fixed c. ∎

Exercise 10 (#7.16). Let $(X_1, ..., X_n)$ be a random sample of binary random variables with $P(X_i = 1) = p$. Using the cumulative distribution function of $T = \sum_{i=1}^n X_i$, show that a level $1-\alpha$ confidence interval for p is

$$\left[\frac{1}{1+\frac{n-T+1}{T}F_{2(n-T+1),2T,\alpha_2}}, \ \frac{\frac{T+1}{n-T}F_{2(T+1),2(n-T),\alpha_1}}{1+\frac{T+1}{n-T}F_{2(T+1),2(n-T),\alpha_1}}\right],$$

where $\alpha_1 + \alpha_2 = \alpha$, $F_{a,b,\alpha}$ is the $(1-\alpha)$th quantile of the F-distribution $F_{a,b}$, and $F_{a,0,\alpha}$ is defined to be ∞.

Solution. Since T has the binomial distribution with size n and probability p and the binomial family has monotone likelihood ratio in T, the cumulative distribution function of T, $F_{T,p}(t)$, is decreasing in p for fixed t. By Theorem 7.1 in Shao (2003), a level $1-\alpha$ confidence interval for p is $[\underline{p}, \overline{p}]$, where $\overline{p}$ is the solution to $F_{T,p}(T) = \alpha_1$ and $\underline{p}$ is the solution to $F_{T,p}(T-) = 1-\alpha_2$. Let Y be a random variable having the beta distribution with parameter $(t, n-t+1)$. Using integral by parts, we obtain that

$$\begin{aligned}
P(Y \leq p) &= \int_0^p \frac{n!}{(t-1)!(n-t)!} y^{t-1}(1-y)^{n-t} dy \\
&= \frac{n!}{t!(n-t)!} p^t (1-p)^{n-t} + \int_0^p \frac{n!}{t!(n-t-1)!} y^t (1-y)^{n-t-1} dy \\
&= \sum_{i=t}^n \frac{n!}{i!(n-i)!} p^i (1-p)^{n-i} \\
&= 1 - F_{T,p}(t-).
\end{aligned}$$

Therefore, $\underline{p}$ is the α_2th quantile of the beta distribution with parameter $(T, n-T+1)$ if $T > 0$ and is equal to 0 if $T = 0$. For $\overline{p}$, it is the solution to $1-\alpha_1 = 1 - F_{T,p}(T+1-)$. Hence, $\overline{p}$ is the $(1-\alpha_1)$th quantile of the beta distribution with parameter $(T+1, n-T)$ if $T < n$ and is equal to 1 if $T = n$. Let $F_{a,b}$ be a random variable having the F-distribution $F_{a,b}$. Then, $\frac{(a/b)F_{a,b}}{1+(a/b)F_{a,b}}$ has the beta distribution with parameter $(a/2, b/2)$. Hence

$$\overline{p} = \frac{\frac{T+1}{n-T}F_{2(T+1),2(n-T),\alpha_1}}{1+\frac{T+1}{n-T}F_{2(T+1),2(n-T),\alpha_1}}$$

when $F_{a,0,\alpha}$ is defined to be ∞. Similarly,

$$\underline{p} = \frac{\frac{T}{n-T+1}F_{2T,2(n-T+1),1-\alpha_2}}{1+\frac{T}{n-T+1}F_{2T,2(n-T+1),1-\alpha_2}}.$$

Note that $F_{a,b}^{-1}$ has the F-distribution $F_{b,a}$. Hence,

$$\underline{p} = \frac{1}{1+\frac{n-T+1}{T}F_{2(n-T+1),2T,\alpha_2}}. \quad \blacksquare$$

Exercise 11 (#7.17). Let X be a sample of size 1 from the negative binomial distribution with a known size r and an unknown probability $p \in (0,1)$. Using the cumulative distribution function of $T = X - r$, show that a level $1-\alpha$ confidence interval for p is

$$\left[\frac{1}{1+\frac{T+1}{r}F_{2(T+1),2r,\alpha_2}}, \frac{\frac{r}{T}F_{2r,2T,\alpha_1}}{1+\frac{r}{T}F_{2r,2T,\alpha_1}}\right],$$

where $\alpha_1 + \alpha_2$ and $F_{a,b,\alpha}$ is the same as that in the previous exercise.
Solution. Since the negative binomial family has monotone likelihood ratio in $-T$, the cumulative distribution function of T, $F_{T,p}(t)$, is increasing in p for fixed t. By Theorem 7.1 in Shao (2003), a level $1-\alpha$ confidence interval for p is $[\underline{p}, \overline{p}]$, where $\underline{p}$ is the solution to $F_{T,p}(T) = \alpha_2$ and $\overline{p}$ is the solution to $F_{T,p}(T-) = 1-\alpha_1$. Let $B_{m,p}$ denote a binomial random variable with size m and probability p and $\beta_{a,b}$ denote a beta random variable with parameter (a,b). Then,

$$F_{T,p}(t) = P(B_{t+r,p} > r-1) = P(B_{t+r,p} \geq r) = P(\beta_{r,t+1} \leq p).$$

Hence, $\underline{p}$ is the α_2th quantile of $\beta_{r,T+1}$. Since $F_{T,p}(T-) = F_{T,p}(T-1)$, $\overline{p}$ is the $(1-\alpha_1)$th quantile of $\beta_{r,T}$ if $T > 0$ and 1 if $T = 0$. Using the same argument as that in the solution of the previous exercise, we conclude that

$$[\underline{p}, \overline{p}] = \left[\frac{1}{1+\frac{T+1}{r}F_{2(T+1),2r,\alpha_2}}, \frac{\frac{r}{T}F_{2r,2T,\alpha_1}}{1+\frac{r}{T}F_{2r,2T,\alpha_1}}\right]. \quad \blacksquare$$

Exercise 12 (#7.18). Let T be a statistic having the noncentral chi-square distribution $\chi_r^2(\theta)$, where the noncentrality parameter $\theta \geq 0$ is unknown and r is a known positive integer. Show that the cumulative distribution function of T, $F_\theta(t)$, is nonincreasing in θ for each fixed $t > 0$ and use this result to construct a confidence interval for θ with confidence coefficient $1-\alpha$.
Solution A. From Exercise 27 of Chapter 1,

$$F_\theta(t) = e^{-\theta/2}\sum_{j=0}^{\infty}\frac{(\theta/2)^j}{j!}G_{2j+r}(t),$$

where $G_{2j+r}(t)$ is the cumulative distribution function of the central chi-square distribution χ^2_{2j+r}, $j = 1, 2, ...$ From the definition of the chi-square distribution χ^2_{2j+r}, it is the distribution of the sum of $2j+r$ independent and identically distributed χ^2_1 random variables. Hence, $G_{2j+r}(t)$ is decreasing in j for any fixed $t > 0$. Let Y be a random variable having the Poisson distribution with mean $\theta/2$ and $h(j) = G_{2j+r}(t)$, $t > 0$, $j = 0, 1, 2, ...$ Then $F_\theta(t) = E_\theta[h(Y)]$, where E_θ is the expectation with respect to the distribution of Y. Since the family of distributions of Y has monotone likelihood ratio in Y and $h(Y)$ is decreasing in Y, by Lemma 6.3 in Shao (2003), $E_\theta[h(Y)]$ is nonincreasing in θ. By Theorem 7.1 in Shao (2003), a $1 - \alpha$ confidence interval for θ is $[\underline{\theta}, \overline{\theta}]$ with

$$\overline{\theta} = \sup\{\theta : F_\theta(T) \geq \alpha_1\} \quad \text{and} \quad \underline{\theta} = \inf\{\theta : F_\theta(T) \leq 1 - \alpha_2\},$$

where $\alpha_1 + \alpha_2 = \alpha$.

Solution B. By definition,

$$F_\theta(t) = P(X + Y \leq t) = \int_0^\infty P(X \leq t - y) f(y) dy,$$

where X has the noncentral chi-square distribution $\chi^2_1(\theta)$, Y has the central chi-square distribution χ^2_{r-1}, $f(y)$ is the Lebesgue density of Y, and X and Y are independent ($Y = 0$ if $r = 1$). From Exercise 9(ii) in Chapter 6, the family of densities of noncentral chi-square distributions $\chi^2_1(\theta)$ has monotone likelihood ratio in X and, hence, $P(X \leq t - y)$ is nonincreasing in θ for any t and y. Hence, $F_\theta(t)$ is is nonincreasing in θ for any $t > 0$. The rest of the solution is the same as that in Solution A. ∎

Exercise 13 (#7.19). Repeat the previous exercise when $\chi^2_r(\theta)$ is replaced by the noncentral F-distribution $F_{r_1,r_2}(\theta)$ with unknown $\theta \geq 0$ and known positive integers r_1 and r_2.

Solution. It suffices to show that the cumulative distribution function of $F_{r_1,r_2}(\theta)$, $F_\theta(t)$, is nonincreasing in θ for any $t > 0$, since the rest of the solution is the same as that in Solution A of the previous exercise. By definition, $F_\theta(t)$ is the cumulative distribution function of $(U_1/r_1)/(U_2/r_2)$, where U_1 has the noncentral chi-square distribution $\chi^2_{r_1}(\theta)$, U_2 has the central chi-square distribution $\chi^2_{r_2}$, and U_1 and U_2 are independent. Let $g(y)$ be the Lebesgue density of $r_1 U_2/r_2$. Then

$$F_\theta(t) = P\big(U_1 \leq t(r_1 U_2/r_2)\big) = \int_0^\infty P(U_1 \leq ty) g(y) dy.$$

From the previous exercise, $P(U_1 \leq ty)$ is nonincreasing in θ for any t and y. Hence, $F_\theta(t)$ is nonincreasing in θ for any t. ∎

Exercise 14 (#7.20). Let X_{ij}, $j = 1, ..., n_i$, $i = 1, ..., m$, be independent random variables having distribution $N(\mu_i, \sigma^2)$, $i = 1, ..., m$. Let

$\bar{\mu} = n^{-1}\sum_{i=1}^m n_i\mu_i$ and $\theta = \sigma^{-2}\sum_{i=1}^m n_i(\mu_i - \bar{\mu})^2$. Construct an upper confidence bound for θ that has confidence coefficient $1-\alpha$ and is a function of $T = (n-m)(m-1)^{-1}\mathrm{SST}/\mathrm{SSR}$, where

$$\mathrm{SSR} = \sum_{i=1}^m\sum_{j=1}^{n_i}(X_{ij} - \bar{X}_{i\cdot})^2, \qquad \mathrm{SST} = \sum_{i=1}^m\sum_{j=1}^{n_i}(X_{ij} - \bar{X})^2,$$

$\bar{X}_{i\cdot}$ is the sample mean based on $X_{i1},...,X_{in_i}$, and $\bar{X}$ is the sample mean based on all X_{ij}'s.

Solution. Note that $\sum_{j=1}^{n_i}(X_{ij} - \bar{X}_{i\cdot})^2$ has the chi-square distribution $\chi^2_{n_i-1}$, $i = 1,...,m$. By the independence of X_{ij}'s, SSR has the chi-square distribution χ^2_{n-m}, where $n = \sum_{i=1}^m n_i$. Let $Y = (\bar{X}_{1\cdot},...,\bar{X}_{m\cdot})$ and A be the $m\times m$ diagonal matrix whose ith diagonal element is $\sqrt{n_i}/\sigma$. Then AY has distribution $N_m(\zeta, I_m)$, where I_m is the identity matrix of order m and $\zeta = (\mu_1\sqrt{n_1}/\sigma,...,\mu_m\sqrt{n_m}/\sigma)$. Note that

$$\mathrm{SSA} = \sum_{i=1}^m n_i(\bar{X}_{i\cdot} - \bar{X})^2 = Y^\tau A(I_m - n^{-1}K_mK_m^\tau)AY,$$

where $K_m = (\sqrt{n_1},...,\sqrt{n_m})$. Since $K_m^\tau K_m = n$, $(I_m - n^{-1}K_mK_m^\tau)^2 = (I_m - n^{-1}K_mK_m^\tau)$ and, by Exercise 22(i) in Chapter 1, SSA has the noncentral chi-square distribution $\chi^2_{m-1}(\delta)$ with

$$\delta = [E(AY)]^\tau(I_m - n^{-1}K_mK_m^\tau)E(AY) = \theta.$$

Also, by Basu's theorem, SSA and SSR are independent. Since SSA = SST − SSR, we conclude that

$$T - \frac{n-m}{m-1} = \frac{\mathrm{SSA}/(m-1)}{\mathrm{SSR}/(n-m)}$$

has the noncentral F-distribution $F_{m-1,n-m}(\theta)$. From the previous exercise, the cumulative distribution function of T, $F_\theta(t)$, is nonincreasing in θ for any t. Hence, by Theorem 7.1 in Shao (2003), an upper confidence bound for θ that has confidence coefficient $1-\alpha$ is

$$\overline{\theta} = \sup\{\theta : F_\theta(T) \geq \alpha\}. \ ∎$$

Exercise 15 (#7.24). Let X_i, $i = 1,2$, be independent random variables distributed as the binomial distributions with sizes n_i and probabilities p_i, $i = 1,2$, respectively, where n_i's are known and p_i's are unknown. Show how to invert the acceptance regions of UMPU tests to obtain a level $1-\alpha$ confidence interval for the odds ratio $\frac{p_2(1-p_1)}{p_1(1-p_2)}$.

Solution. Let $\theta = \frac{p_2(1-p_1)}{p_1(1-p_2)}$. From the solution to Exercise 24 in Chapter 6, a UMPU test for $H_0 : \theta = \theta_0$ versus $H_1 : \theta \neq \theta_0$ has acceptance region

$$A(\theta_0) = \{(Y, U) : c_1(U, \theta_0) \leq Y \leq c_2(U, \theta_0)\},$$

where $Y = X_2$, $U = X_1 + X_2$, and $c_i(U, \theta)$ are some functions. From Exercise 19 in Chapter 6, for each fixed U, $c_i(U, \theta)$ is nondecreasing in θ. Hence, for every θ,

$$\{\theta \in A(\theta)\} = \{\theta : c_1(U, \theta) \leq Y \leq c_2(U, \theta)\} = [c_{2,U}^{-1}(Y), c_{1,U}^{-1}(Y)],$$

where

$$c_{i,U}^{-1}(Y) = \inf\{x : c_i(U, x) \geq Y\}.$$

From Theorem 7.2 in Shao (2003), $[c_{2,U}^{-1}(Y), c_{1,U}^{-1}(Y)]$ is a level $1 - \alpha$ confidence interval for θ. ∎

Exercise 16 (#7.25). Let $X = (X_1, ..., X_n)$ be a random sample from $N(\mu, \sigma^2)$.
(i) Suppose that $\sigma^2 = \gamma\mu^2$ with unknown $\gamma > 0$ and $\mu \in \mathcal{R}$, $\mu \neq 0$. Obtain a confidence set for γ with confidence coefficient $1 - \alpha$ by inverting the acceptance regions of LR tests for $H_0 : \gamma = \gamma_0$ versus $H_1 : \gamma \neq \gamma_0$.
(ii) Repeat (i) when $\sigma^2 = \gamma\mu$ with unknown $\gamma > 0$ and $\mu > 0$.
Solution. (i) The likelihood function is given in part (i) of the solution to Exercise 47 in Chapter 6. The MLE of (μ, γ) is $(\hat{\mu}, \hat{\gamma}) = (\bar{X}, \hat{\sigma}^2/\bar{X}^2)$, where $\bar{X}$ is the sample mean and $\hat{\sigma}^2 = n^{-1}\sum_{i=1}^n (X_i - \bar{X})^2$. When $\gamma = \gamma_0$, using the same argument in the solution of Exercise 41(viii) in Chapter 4, we obtain the MLE of μ as

$$\hat{\mu}(\gamma_0) = \begin{cases} \mu_+(\gamma_0) & \ell(\mu_+(\gamma_0), \gamma_0) > \ell(\mu_-(\gamma_0), \gamma_0) \\ \mu_-(\gamma_0) & \ell(\mu_+(\gamma_0), \gamma_0) \leq \ell(\mu_-(\gamma_0), \gamma_0), \end{cases}$$

where

$$\mu_\pm(\gamma_0) = \frac{-\bar{X} \pm \sqrt{(5\bar{X}^2 + 4\hat{\sigma}^2)/\gamma_0}}{2}.$$

The likelihood ratio is

$$\lambda(\gamma_0) = \frac{e^{n/2}\hat{\sigma}^n}{\gamma_0^{n/2}|\hat{\mu}(\gamma_0)|^n} \exp\left\{-\frac{n\hat{\sigma}^2 + n[\hat{\mu}(\gamma_0) - \bar{X}]^2}{2[\hat{\mu}(\gamma_0)]^2}\right\}.$$

The confidence set obtained by inverting the acceptance regions of LR tests is $\{\gamma : \lambda(\gamma) \geq c(\gamma)\}$, where $c(\gamma)$ satisfies $P(\lambda(\gamma) < c(\gamma)) = \alpha$.
(ii) The likelihood function is given in part (ii) of the solution to Exercise 47 in Chapter 6. The MLE of (μ, γ) is $(\bar{X}, \hat{\sigma}^2/\bar{X})$ when $\bar{X} > 0$. If $\bar{X} \leq 0$, however, the likelihood is unbounded in γ. When $\gamma = \gamma_0$, using the same

argument as that in the solution to Exercise 60 of Chapter 4, we obtain the MLE of μ as $\hat{\mu}(\gamma_0) = (\sqrt{\gamma_0^2 + 4T} - \gamma_0)/2$, where $T = n^{-1}\sum_{i=1}^n X_i^2$. The likelihood ratio is

$$\lambda(\gamma_0) = \begin{cases} \left[\frac{e\hat{\sigma}^2}{\gamma_0\hat{\mu}(\gamma_0)}\right]^{-n/2} \exp\left\{-\frac{nT}{2\gamma_0\hat{\mu}(\gamma_0)} + \frac{n\bar{X}}{\gamma_0} - \frac{n\hat{\mu}(\gamma_0)}{2\gamma_0}\right\} & \bar{X} > 0 \\ 0 & \bar{X} \le 0. \end{cases}$$

The confidence set obtained by inverting the acceptance regions of LR tests is $\{\gamma : \lambda(\gamma) \ge c(T,\gamma)\}$, where $c(T,\gamma)$ satisfies $P(\lambda(\gamma) < c(T,\gamma)|T) = \alpha$. ∎

Exercise 17 (#7.26). Let $X = (X_1, ..., X_n)$ be a random sample from $N(\mu, \sigma^2)$ with unknown μ and σ^2. Discuss how to construct a confidence interval for $\theta = \mu/\sigma$ with confidence coefficient $1-\alpha$ by
(i) inverting the acceptance regions of the tests given in Exercise 22 of Chapter 6;
(ii) applying Theorem 7.1 in Shao (2003).
Solution. (i) From Exercise 22 in Chapter 6, the acceptance region of a test of size α for $H_0 : \theta \le \theta_0$ versus $H_1 : \theta > \theta_0$ is $\{X\colon t(X) \le c_\alpha(\theta_0)\}$, where $t(X) = \sqrt{n}\bar{X}/S$, $\bar{X}$ is the sample mean, S^2 is the sample variance, and $c_\alpha(\theta)$ is the $(1-\alpha)$th quantile of the noncentral t-distribution $t_{n-1}(\sqrt{n}\theta)$. From the solution to Exercise 22 in Chapter 6, the family of densities of $t_{n-1}(\sqrt{n}\theta)$ has monotone likelihood ratio in $t(X)$. By Lemma 6.3 in Shao (2003), $c_\alpha(\theta)$ is increasing in θ and, therefore, $\{\theta : c_\alpha(\theta) \ge t(X)\} = [\underline{\theta}(X), \infty)$ for some $\underline{\theta}(X)$. By Theorem 7.2 in Shao (2003), $[\underline{\theta}(X), \infty)$ is a confidence interval for $\theta = \mu/\sigma$ with confidence coefficient $1-\alpha$. If it is desired to obtain a bounded confidence interval for θ, then we may consider $C(X) = \{\theta : c_{\alpha/2}(\theta) \ge t(X) \ge d_{\alpha/2}(\theta)\}$, where $d_\alpha(\theta)$ is the is the αth quantile of $t_{n-1}(\sqrt{n}\theta)$. By considering the problem of testing $H_0 : \theta \ge \theta_0$ versus $H_1 : \theta < \theta_0$, we conclude that $\{\theta : t(X) \ge d_{\alpha/2}(\theta)\}$ is a confidence interval for θ with confidence coefficient $1-\alpha/2$. Hence, $C(X)$ is a confidence interval for θ with confidence coefficient $1-\alpha$.
(ii) The cumulative distribution function of $t(X)$ is

$$F_\theta(t) = \int_0^\infty \Phi(ty - \theta) f(y) dy,$$

where Φ is the cumulative distribution function of $N(0,1)$, $f(y)$ is the Lebesgue density of $\sqrt{W/(n-1)}$, and W has the chi-square distribution χ^2_{n-1}. Hence, for any fixed t, $F_\theta(t)$ is continuous and decreasing in θ, $\lim_{\theta\to\infty} F_\theta(t) = 0$, and $\lim_{\theta\to-\infty} F_\theta(t) = 1$. By Theorem 7.1 in Shao (2003), $[\underline{\theta}, \overline{\theta}]$ is a confidence interval for θ with confidence coefficient $1-\alpha$, where $\underline{\theta}$ is the unique solution to $F_\theta(t(X)) = 1-\alpha/2$ and $\overline{\theta}$ is the unique solution to $F_\theta(t(X)) = \alpha/2$. ∎

Exercise 18 (#7.27). Let $(X_1, ..., X_n)$ be a random sample from the uniform distribution on $(\theta - \frac{1}{2}, \theta + \frac{1}{2})$, where $\theta \in \mathcal{R}$. Construct a confidence interval for θ with confidence coefficient $1 - \alpha$.

Solution. Note that $X_i + \frac{1}{2} - \theta$ has the uniform distribution on $(0,1)$. Let $X_{(j)}$ be the jth order statistic. Then

$$P\left(X_{(1)} + \frac{1}{2} - \theta \leq c\right) = 1 - (1-c)^n.$$

Hence

$$P\left(X_{(1)} - \frac{1}{2} + \alpha_1^{1/n} \leq \theta\right) = 1 - \alpha_1.$$

Similarly,

$$P\left(X_{(n)} + \frac{1}{2} - \alpha_2^{1/n} \geq \theta\right) = 1 - \alpha_2.$$

Let $\alpha = \alpha_1 + \alpha_2$. A confidence interval for θ with confidence coefficient $1 - \alpha$ is then $[X_{(1)} - \frac{1}{2} + \alpha_1^{1/n}, X_{(n)} + \frac{1}{2} - \alpha_2^{1/n}]$. ∎

Exercise 19 (#7.29). Let $(X_1, ..., X_n)$ be a random sample from $N(\mu, \sigma^2)$ with unknown $\theta = (\mu, \sigma^2)$. Consider the prior Lebesgue density $\pi(\theta) = \pi_1(\mu|\sigma^2)\pi_2(\sigma^2)$, where $\pi_1(\mu|\sigma^2)$ is the density of $N(\mu_0, \sigma_0^2\sigma^2)$,

$$\pi_2(\sigma^2) = \frac{1}{\Gamma(a)b^a}\left(\frac{1}{\sigma^2}\right)^{a+1} e^{-1/(b\sigma^2)} I_{(0,\infty)}(\sigma^2),$$

and μ_0, σ_0^2, a, and b are known.

(i) Find the posterior of μ and construct a level $1-\alpha$ HPD credible set for μ.

(ii) Show that the credible set in (i) converges to the confidence interval $[\bar{X} - t_{n-1,\alpha/2}\frac{S}{\sqrt{n}}, \bar{X} + t_{n-1,\alpha/2}\frac{S}{\sqrt{n}}]$ as σ_0^2, a, and b converge to some limits, where $\bar{X}$ is the sample mean, S^2 is the sample variance, and $t_{n-1,\alpha}$ is the $(1-\alpha)$th quantile of the t-distribution t_{n-1}.

Solution. (i) This is a special case of the problem in Exercise 20(iii) of Chapter 4. Let $\omega = \sigma^{-2}$. Then the posterior density of (μ, ω) is $p(\mu|\omega)p(\omega)$, where $p(\mu|\omega)$ is the density of $N(\mu_*, \omega^{-1}c_*^{-1})$, $\mu_* = (\sigma_0^{-2}\mu_0 + n\bar{X})/(n + \sigma_0^{-2})$, $c_* = n + \sigma_0^{-2}$, and $p(\omega)$ is the density of the gamma distribution with shape parameter $a + n/2$ and scale parameter $\gamma = [b^{-1} + (n-1)S^2/2]^{-1}$. The posterior density for μ is then

$$\begin{aligned}
f(\mu) &= \int_0^\infty p(\mu|\omega)p(\omega)d\omega \\
&= \int_0^\infty \frac{\sqrt{c_*}\gamma^{-(a+n/2)}}{\sqrt{2\pi}\Gamma(a+\frac{n}{2})}\omega^{a+(n-1)/2}e^{-[\gamma^{-1}+c_*(\mu-\mu_*)^2/2]\omega}d\omega \\
&= \frac{\Gamma(a+\frac{n+1}{2})\sqrt{c_*}(2\gamma^{-1})^{a+n/2}}{\sqrt{\pi}\Gamma(a+\frac{n}{2})[2\gamma^{-1}+c_*(\mu-\mu_*)^2]^{a+(n+1)/2}}.
\end{aligned}$$

Since this density is symmetric about μ_*, a level $1-\alpha$ HPD credible set for μ is $[\mu_* - t_*, \mu_* + t_*]$, where $t_* > 0$ satisfies

$$\int_{\mu_*-t_*}^{\mu_*+t_*} f(\mu)d\mu = 1-\alpha.$$

(ii) Let $\sigma_0^2 \to \infty$, $a \to -1/2$, and $b \to \infty$. Then, $\mu_* \to \bar{X}$, $c_* \to n$, $2\gamma^{-1} \to (n-1)S^2$, and

$$f(\mu) \to \frac{\Gamma(\frac{n}{2})\sqrt{n}[(n-1)S^2]^{(n-1)/2}}{\sqrt{\pi}\Gamma(\frac{n-1}{2})[(n-1)S^2 + n(\mu-\bar{X})^2]^{n/2}},$$

which is the density of $(S/\sqrt{n})T$ with T being a random variable having t-distribution t_{n-1}. Hence, $t_* \to t_{n-1,\alpha/2}S/\sqrt{n}$ and the result follows. ∎

Exercise 20 (#7.30). Let $(X_1, ..., X_n)$ be a random sample from a distribution on $\mathcal{R}$ with Lebesgue density $\frac{1}{\sigma}f\left(\frac{x-\mu}{\sigma}\right)$, where f is a known Lebesgue density and $\mu \in \mathcal{R}$ and $\sigma > 0$ are unknown. Let X_0 be a future observation that is independent of X_i's and has the same distribution as X_i. Find a pivotal quantity $R(X, X_0)$ and construct a level $1-\alpha$ prediction set for X_0.
Solution. Let $\bar{X}$ and S^2 be the sample mean and sample variance. Consider $T = (X_0 - \bar{X})/S$. Since

$$T = \frac{X_0 - \bar{X}}{S} = \frac{\frac{X_0-\mu}{\sigma} - \frac{1}{n}\sum_{i=1}^n \frac{X_i-\mu}{\sigma}}{\left[\frac{1}{n-1}\sum_{i=1}^n\left(\frac{X_i-\mu}{\sigma} - \frac{1}{n}\sum_{j=1}^n \frac{X_j-\mu}{\sigma}\right)^2\right]^{1/2}}$$

and the density of $(X_i - \mu)/\sigma$ is f, T is a pivotal quantity. A $1-\alpha$ prediction set for X_0 is $\left\{X_0 : |X_0 - \bar{X}| \le cS\right\}$, where c is chosen such that $P(|T| \le c) = 1-\alpha$. ∎

Exercise 21 (#7.31). Let $(X_1, ..., X_n)$ be a random sample from a continuous cumulative distribution function F on $\mathcal{R}$ and X_0 be a future observation that is independent of X_i's and is distributed as F. Suppose that F is increasing in a neighborhood of $F^{-1}(\alpha/2)$ and a neighborhood of $F^{-1}(1-\alpha/2)$. Let F_n be the empirical distribution. Show that the prediction interval $C(X) = [F_n^{-1}(\alpha/2), F_n^{-1}(1-\alpha/2)]$ for X_0 satisfies $\lim_n P(X_0 \in C(X)) = 1-\alpha$, where P is the joint distribution of $(X_0, X_1, ..., X_n)$.
Solution. Since F is increasing in a neighborhood of $F^{-1}(\alpha/2)$, $F^{-1}(t)$ is continuous at $\alpha/2$. By the result in Exercise 28 of Chapter 5, $\lim_n F_n^{-1}(\alpha/2) = F^{-1}(\alpha/2)$ a.s. Then $X_0 - F_n^{-1}(\alpha/2) \to_d X_0 - F^{-1}(\alpha/2)$ and, thus,

$$\lim_n P\big(F_n^{-1}(\alpha/2) > X_0\big) = P\big(F^{-1}(\alpha/2) > X_0\big) = \alpha/2,$$

since F is continuous. Similarly,

$$\lim_n P\big(X_0 \le F_n^{-1}(1-\alpha/2)\big) = P\big(X_0 \le F^{-1}(1-\alpha/2)\big) = 1-\alpha/2.$$

Hence,

$$\lim_n P\big(X_0 \in C(X)\big) = \lim_n P\big(F_n^{-1}(\alpha) \le X_0 \le F_n^{-1}(1-\alpha/2)\big) = 1-\alpha. \quad \blacksquare$$

Exercise 22 (#7.33). Let $X = (X_1, ..., X_n)$ ($n > 1$) be a random sample from the exponential distribution on the interval (θ, ∞) with scale parameter θ, where $\theta > 0$ is unknown.
(i) Show that both $\bar{X}/\theta$ and $X_{(1)}/\theta$ are pivotal quantities, where $\bar{X}$ is the sample mean and $X_{(1)}$ is the smallest order statistic.
(ii) Obtain confidence intervals (with confidence coefficient $1-\alpha$) for θ based on the two pivotal quantities in (i).
(iii) Discuss which confidence interval in (ii) is better in terms of the length.
Solution. (i) Note that $X_i/\theta - 1$ has the exponential distribution on the interval $(0, \infty)$ with scale parameter 1. Hence, $\bar{X}/\theta - 1$ has the gamma distribution with shape parameter n and scale parameter n^{-1} and $X_{(1)}/\theta - 1$ has the exponential distribution on $(0, \infty)$ with scale parameter n^{-1}. Therefore, both $\bar{X}/\theta$ and $X_{(1)}/\theta$ are pivotal quantities.
(ii) Let $c_{n,\alpha}$ be the αth quantile of the gamma distribution with shape parameter n and scale parameter n^{-1}. Then

$$P\left(c_{n,\alpha/2} \le \frac{\bar{X}}{\theta} - 1 \le c_{n,1-\alpha/2}\right) = 1-\alpha,$$

which leads to the $1-\alpha$ confidence interval

$$C_1(X) = \left[\frac{\bar{X}}{1+c_{n,1-\alpha/2}}, \frac{\bar{X}}{1+c_{n,\alpha/2}}\right].$$

On the other hand,

$$P\left(\frac{1}{n}\log\frac{1}{1-\alpha/2} \le \frac{X_{(1)}}{\theta} - 1 \le \frac{1}{n}\log\frac{1}{\alpha/2}\right) = 1-\alpha,$$

which leads to the $1-\alpha$ confidence interval

$$C_2(X) = \left[\frac{X_{(1)}}{1-n^{-1}\log(1-\alpha/2)}, \frac{X_{(1)}}{1-n^{-1}\log(\alpha/2)}\right].$$

(iii) The length of $C_1(X)$ is

$$\frac{\bar{X}(c_{n,\alpha/2} - c_{n,1-\alpha/2})}{(1+c_{n,\alpha/2})(1+c_{n,1-\alpha/2})}$$

and the length of $C_2(X)$ is

$$\frac{X_{(1)}\log(2/\alpha-1)n^{-1}}{[1-n^{-1}\log(\alpha/2)][1-n^{-1}\log(1-\alpha/2)]}.$$

From the central limit theorem, $\lim_n \sqrt{n}(c_{n,\alpha/2}-c_{n,1-\alpha/2})>0$, assuming $\alpha<\frac{1}{2}$. Hence, for sufficiently large n,

$$\frac{\log(2/\alpha-1)n^{-1}}{[1-n^{-1}\log(\alpha/2)][1-n^{-1}\log(1-\alpha/2)]}<\frac{(c_{n,\alpha/2}-c_{n,1-\alpha/2})}{(1+c_{n,\alpha/2})(1+c_{n,1-\alpha/2})}.$$

Also, $\bar{X}>X_{(1)}$. Hence, for sufficiently large n, the length of $C_2(X)$ is shorter than the length of $C_1(X)$. ∎

Exercise 23 (#7.34). Let $\theta>0$ be an unknown parameter and $T>0$ be a statistic. Suppose that T/θ is a pivotal quantity having Lebesgue density f and that $x^2 f(x)$ is unimodal at x_0 in the sense that $f(x)$ is nondecreasing for $x\le x_0$ and $f(x)$ is nonincreasing for $x\ge x_0$. Consider the following class of confidence intervals for θ:

$$\mathcal{C}=\left\{[b^{-1}T,a^{-1}T]:\ a>0,b>0,\int_a^b f(x)dx=1-\alpha\right\}.$$

Show that if $[b_*^{-1}T,a_*^{-1}T]\in\mathcal{C}$, $a_*^2f(a_*)=b_*^2f(b_*)>0$, and $a_*\le x_0\le b_*$, then the interval $[b_*^{-1}T,a_*^{-1}T]$ has the shortest length within $\mathcal{C}$.

Solution. We need to minimize $\frac{1}{a}-\frac{1}{b}$ under the constraint $\int_a^b f(x)dx=1-\alpha$. Let $t=\frac{1}{x}$, then

$$\int_a^b f(x)dx=\int_{\frac{1}{b}}^{\frac{1}{a}} f\left(\frac{1}{t}\right)\frac{1}{t^2}dt=1-\alpha.$$

Since f is unimodal at x_0, $f(\frac{1}{t})\frac{1}{t^2}$ is unimodal at $t=\frac{1}{x_0}$. The result follows by applying Theorem 7.3(i) in Shao (2003) to the function $f(\frac{1}{t})\frac{1}{t^2}$. ∎

Exercise 24 (#7.35). Let $t_{n-1,\alpha}$ be the $(1-\alpha)$th quantile of the t-distribution t_{n-1} and z_α be the $(1-\alpha)$th quantile of $N(0,1)$, where $0<\alpha<\frac{1}{2}$ and $n=2,3,....$ Show that

$$\frac{\sqrt{2}\Gamma(\frac{n}{2})}{\sqrt{n}\Gamma(\frac{n-1}{2})}t_{n-1,\alpha}\ge z_\alpha,\quad n=2,3,....$$

Solution. Let $X=(X_1,...,X_n)$ be a random sample from $N(\mu,\sigma^2)$. If σ^2 is known, then a $1-2\alpha$ confidence interval obtained by inverting the UMPU tests is $C_1(X)=[\bar{X}-z_\alpha\sigma/\sqrt{n},\bar{X}+z_\alpha\sigma/\sqrt{n}]$, where $\bar{X}$ is the sample mean.

If σ^2 is unknown, then a $1-2\alpha$ confidence interval obtained by inverting the UMPU tests is $C_2(X) = [\bar{X} - t_{n-1,\alpha}S/\sqrt{n}, \bar{X} + t_{n-1,\alpha}S/\sqrt{n}]$, where S^2 is the sample variance. For any fixed σ, both $C_1(X)$ and $C_2(X)$ are unbiased confidence intervals. By Theorem 7.5 in Shao (2003), $C_1(X)$ is the UMAU (uniformly most accurate unbiased) confidence interval. By Pratt's theorem (e.g., Theorem 7.6 in Shao, 2003), the expected length of $C_1(X)$ is no larger than the expected length of $C_2(X)$. The length of $C_2(X)$ is $2t_{n-1,\alpha}S/\sqrt{n}$. Since $(n-1)S^2/\sigma^2$ has the chi-square distribution χ^2_{n-1},

$$E(S) = \frac{\sqrt{2}\Gamma(\frac{n}{2})}{\sqrt{n}\Gamma(\frac{n-1}{2})}\sigma, \quad n = 2, 3, ...,$$

which implies that the expected length of $C_2(X)$ is

$$\frac{2\sigma}{\sqrt{n}} \frac{\sqrt{2}\Gamma(\frac{n}{2})}{\sqrt{n}\Gamma(\frac{n-1}{2})} t_{n-1,\alpha} \geq \frac{2\sigma}{\sqrt{n}} z_\alpha,$$

the length of $C_1(X)$. This proves the result. ∎

Exercise 25 (#7.36(a),(c)). Let $(X_1, ..., X_n)$ be a random sample from $N(\mu, \sigma^2)$, $\mu \in \mathcal{R}$ and $\sigma^2 > 0$.

(i) Suppose that μ is known. Let a_n and b_n be constants satisfying $a_n^2 f_n(a_n) = b_n^2 f_n(b_n) > 0$ and $\int_{a_n}^{b_n} f_n(x)dx = 1-\alpha$, where f_n is the Lebesgue density of the chi-square distribution χ^2_n. Show that the interval $[b_n^{-1}T, a_n^{-1}T]$ has the shortest length within the class of intervals of the form $[b^{-1}T, a^{-1}T]$, $\int_a^b f_n(x)dx = 1-\alpha$, where $T = \sum_{i=1}^n (X_i - \mu)^2$.

(ii) When μ is unknown, show that $[b_{n-1}^{-1}(n-1)S^2, a_{n-1}^{-1}(n-1)S^2]$ has the shortest length within the class of $1-\alpha$ confidence intervals of the form $[b^{-1}(n-1)S^2, a^{-1}(n-1)S^2]$, where S^2 is the sample variance.

(iii) Find the shortest-length interval for σ within the class of confidence intervals of the form $[b^{-1/2}\sqrt{n-1}S, a^{-1/2}\sqrt{n-1}S]$, where $0 < a < b < \infty$, and $\int_a^b f_{n-1}(x)dx = 1-\alpha$.

Solution. (i) Note that T/σ^2 has the chi-square distribution χ^2_n with Lebesgue density f_n and $x^2 f_n(x)$ is unimodal. The result follows from Exercise 23.

(ii) Since $(n-1)S^2$ has the chi-square distribution χ^2_{n-1}, the result follows from part (i) of the solution with n replaced by $n-1$.

(iii) Let $t = 1/\sqrt{x}$. Then

$$\int_a^b f(x)dx = \int_{\frac{1}{\sqrt{b}}}^{\frac{1}{\sqrt{a}}} f_{n-1}\left(\frac{1}{\sqrt{t}}\right)\frac{1}{2t\sqrt{t}}dt.$$

Minimizing $\frac{1}{\sqrt{a}} - \frac{1}{\sqrt{b}}$ under the constraint $\int_a^b f_{n-1}(x)dx = 1-\alpha$ is the same as minimizing $\frac{1}{\sqrt{a}} - \frac{1}{\sqrt{b}}$ under the constraint $\int_{\frac{1}{\sqrt{b}}}^{\frac{1}{\sqrt{a}}} f_{n-1}\left(\frac{1}{\sqrt{t}}\right)\frac{1}{2t\sqrt{t}}dt = 1-\alpha$.

It is easy to check that $f_{n-1}(\frac{1}{\sqrt{t}})\frac{1}{2t\sqrt{t}}$ is unimodal. By Exercise 23, there exist $\frac{1}{\sqrt{b_*}} < \frac{1}{\sqrt{a_*}}$ such that $b_*^3 f_{n-1}(b_*) = a_*^3 f_{n-1}(a_*)$ and the confidence interval $[b_*^{-1/2}\sqrt{n-1}S, a_*^{-1/2}\sqrt{n-1}S]$ has the shortest length within the class of confidence intervals of the form $[b^{-1/2}\sqrt{n-1}S, a^{-1/2}\sqrt{n-1}S]$, where $0 < a < b < \infty$, and $\int_a^b f_{n-1}(x)dx = 1-\alpha$. ∎

Exercise 26 (#7.38). Let f be a Lebesgue density that is nonzero in $[x_-, x_+]$ and is 0 outside $[x_-, x_+]$, $-\infty \le x_- < x_+ \le \infty$.
(i) Suppose that f is decreasing. Show that, among all intervals $[a,b]$ satisfying $\int_a^b f(x)dx = 1-\alpha$, the shortest interval is obtained by choosing $a = x_-$ and b_* so that $\int_{x_-}^{b_*} f(x)dx = 1-\alpha$.
(ii) Obtain a result similar to that in (i) when f is increasing.
(iii) Show that the interval $[X_{(n)}, \alpha^{-1/n}X_{(n)}]$ has the shortest length among all intervals $[b^{-1}X_{(n)}, a^{-1}X_{(n)}]$, where $X_{(n)}$ is the largest order statistic based on a random sample of size n from the uniform distribution on $(0,\theta)$.
Solution. (i) Since f is decreasing, we must have $x_- > -\infty$. Without loss of generality, we consider a and b such that $x_- \le a < b \le x_+$. Assume $a \le b_*$. If $b \le b_*$, then

$$\int_{x_-}^a f(x)dx = \int_{x_-}^{b_*} f(x)dx - \int_a^{b_*} f(x)dx < 1-\alpha - \int_a^b f(x)dx = 0,$$

which is impossible, where the inequality follows from the fact that f is decreasing. If $b > b_*$ but $b - a < b_* - x_-$, then

$$\begin{aligned}
\int_a^b f(x)dx &= \int_a^{b_*} f(x)dx + \int_{b_*}^b f(x)dx \\
&\le \int_a^{b_*} f(x)dx + f(b_*)(b-b_*) \\
&< \int_a^{b_*} f(x)dx + f(a)(b-b_*) \\
&< \int_a^{b_*} f(x)dx + f(a)(a-x_-) \\
&\le \int_a^{b_*} f(x)dx + \int_{x_-}^a f(x)dx \\
&= 1-\alpha,
\end{aligned}$$

which contradicts $\int_a^b f(x)dx = 1-\alpha$. Hence, we must have $b - a \le b_* - x_-$. If $a > b_*$ and $b - a \le b_* - x_-$, then

$$\int_a^b f(x)dx \le f(a)(b-a) < f(b_*)(b_* - x_-) \le \int_{x_-}^{b_*} f(x)dx = 1-\alpha.$$

Hence, we must have $b - a > b_* - x_-$.
(ii) When f is increasing, we must have $x_+ < \infty$. The shortest interval is obtained by choosing $b = x_+$ and a_* so that $\int_{a_*}^{x_+} f(x)dx = 1-\alpha$. The proof is similar to that in part (i) of the solution.
(iii) Let $X_{(n)}$ be the largest order statistic based on a random sample of size n from the uniform distribution on $(0,\theta)$. The Lebesgue density of $X_{(n)}/\theta$ is $nx^{n-1}I_{(0,1)}(x)$. We need to minimize $a^{-1} - b^{-1}$ under the constraint

$$1-\alpha = \int_a^b nx^{n-1}dx = \int_{b^{-1}}^{a^{-1}} \frac{n}{y^{n+1}}dy, \quad 0 \le a < b \le 1.$$

Note that n/y^{n+1} is decreasing in $[1,\infty)$. By (i), the solution is $b^{-1} = 1$ and a^{-1} satisfying $\int_1^{a^{-1}} ny^{-(n+1)}dy = 1-\alpha$, which yields $a = \alpha^{1/n}$. The corresponding confidence interval is $[X_{(n)}, \alpha^{1/n}X_{(n)}]$. ∎

Exercise 27 (#7.39). Let $(X_1,...,X_n)$ be a random sample from the exponential distribution on (a,∞) with scale parameter 1, where $a \in \mathcal{R}$ is unknown. Find a confidence interval for a having the shortest length within the class of confidence intervals $[X_{(1)} + c, X_{(1)} + d]$ with confidence coefficient $1-\alpha$, where $X_{(1)}$ is the smallest order statistic.
Solution. The Lebesgue density of $X_{(1)} - \theta$ is $ne^{-nx}I_{[0,\infty)}(x)$, which is decreasing on $[0,\infty)$. Note that

$$P(X_{(1)} + c \le a \le X_{(1)} + d) = P(-d \le X_{(1)} - a \le -c) = \int_{-d}^{-c} ne^{-nx}dx.$$

Hence, $-c \ge -d \ge 0$. To minimize $d - c$, the length of the $1-\alpha$ confidence interval $[X_{(1)} + c, X_{(1)} + d]$, it follows from Exercise 26(i) that $-d = 0$ and $-c$ satisfies

$$\int_0^{-c} ne^{-nx}dx = 1-\alpha,$$

which yields $c = n^{-1}\log\alpha$. The shortest length confidence interval is then $[X_{(1)} + n^{-1}\log\alpha, X_{(1)}]$. ∎

Exercise 28 (#7.42). Let $(X_1,...,X_n)$ be a random sample from a distribution with Lebesgue density $\theta x^{\theta-1}I_{(0,1)}(x)$, where $\theta > 0$ is unknown.
(i) Construct a confidence interval for θ with confidence coefficient $1-\alpha$, using a sufficient statistic.
(ii) Discuss whether the confidence interval obtained in (i) has the shortest length within a class of confidence intervals.
(iii) Discuss whether the confidence interval obtained in (i) is UMAU.
Solution. (i) The complete and sufficient statistic is $T = -\sum_{i=1}^n \log X_i$. Note that θT has the gamma distribution with shape parameter n and scale

parameter 1. Let $f(x)$ be the Lebesgue density of θT. Then a confidence interval of coefficient $1-\alpha$ can be taken from the following class:

$$\mathcal{C} = \left\{ [(bT)^{-1}, (aT)^{-1}] : \int_a^b \frac{1}{x^2} f\left(\frac{1}{x}\right) dx = 1-\alpha \right\}.$$

(ii) Note that $f(\frac{1}{x})$ is unimodal. By Exercise 23, $[(b_*T)^{-1}, (a_*T)^{-1}]$ with $f(a_*) = f(b_*)$ has the shortest length within $\mathcal{C}$.
(iii) Consider testing hypotheses $H_0 : \theta = \theta_0$ versus $H_1 : \theta \neq \theta_0$. The acceptance region of a UMPU test is $A(\theta_0) = \{X : c_1 \leq \theta_0 T \leq c_2\}$, where c_1 and c_2 are determined by

$$\int_{c_1}^{c_2} f(x)dx = 1-\alpha \quad \text{and} \quad \int_{c_1}^{c_2} x f(x)dx = n(1-\alpha).$$

Thus a UMAU confidence interval is $[c_1/T, c_2/T]$, which is a member of $\mathcal{C}$ but in general different from the one in part (ii). ∎

Exercise 29 (#7.45). Let X be a single observation from $N(\theta-1,1)$ if $\theta<0$, $N(0,1)$ if $\theta=0$, and $N(\theta+1,1)$ if $\theta>0$.
(i) Show that the distribution of X is in a family with monotone likelihood ratio in X.
(ii) Construct a Θ'-UMA (uniformly most accurate) lower confidence bound for θ with confidence coefficient $1-\alpha$, where $\Theta' = (-\infty, \theta)$.
Solution. (i) Let $\mu(\theta)$ be the mean of X. Then

$$\mu(\theta) = \begin{cases} \theta - 1 & \theta < 0 \\ 0 & \theta = 0 \\ \theta + 1 & \theta > 0, \end{cases}$$

which is an increasing function of θ. Let $f_\theta(x)$ be the Lebesgue density of X. For any $\theta_2 > \theta_1$,

$$\frac{f_{\theta_2}(x)}{f_{\theta_1}(x)} = \exp\left\{ [\mu(\theta_2) - \mu(\theta_1)]x - \frac{[\mu(\theta_2)]^2 - [\mu(\theta_1)]^2}{2} \right\}$$

is increasing in x. Therefore, the family $\{f_\theta : \theta \in \mathcal{R}\}$ has monotone likelihood ratio in X.
(ii) Consider testing $H_0 : \theta = \theta_0$ versus $H_1 : \theta > \theta_0$. The UMP test has acceptance region $\{X : X \leq c(\theta_0)\}$, where $P_\theta(X \geq c(\theta)) = \alpha$ and P_θ denotes the distribution of X. Since P_θ is $N(\mu(\theta), 1)$, $c(\theta) = z_\alpha + \mu(\theta)$, where z_α is the $(1-\alpha)$th quantile of $N(0,1)$. Inverting these acceptance regions, we obtain a confidence set

$$C(X) = \{\theta : X < z_\alpha + \mu(\theta)\}.$$

When $X - z_\alpha > 1$, if $\theta \le 0$, then $\mu(\theta) > X - z_\alpha$ cannot occur; if $\theta > 0$, $\mu(\theta) > X - z_\alpha$ if and only if $\theta > X - z_\alpha - 1$; hence, $C(X) = (X - z_\alpha - 1, \infty)$. Similarly, when $X - z_\alpha < -1$, $C(X) = (X - z_\alpha + 1, \infty)$. When $-1 \le X - z_\alpha < 0$, $\mu(\theta) > X - z_\alpha$ if and only if $\theta \ge 0$ and, hence, $C(X) = [0, \infty)$. When $0 \le X - z_\alpha \le 1$, $\mu(\theta) > X - z_\alpha$ if and only if $\theta > 0$ and, hence, $C(X) = (0, \infty)$. Hence, a $(-\infty, \theta)$-UMA confidence lower bound for θ is

$$\underline{\theta} = \begin{cases} X - z_\alpha - 1 & X > z_\alpha + 1 \\ 0 & z_\alpha - 1 \le X \le z_\alpha + 1 \\ X - z_\alpha + 1 & X < z_\alpha - 1. \end{cases} \qquad \blacksquare$$

Exercise 30 (#7.46). Let X be a vector of n observations having distribution $N_n(Z\beta, \sigma^2 I_n)$, where Z is a known $n \times p$ matrix of rank $r \le p < n$, β is an unknown p-vector, and $\sigma^2 > 0$ is unknown. Let $\theta = L\beta$, where L is an $s \times p$ matrix of rank s and all rows of L are in $\mathcal{R}(Z)$,

$$W(X, \theta) = \frac{[\|X - Z\hat{\beta}(\theta)\|^2 - \|X - Z\hat{\beta}\|^2]/s}{\|X - Z\hat{\beta}\|^2/(n - r)},$$

where $\hat{\beta}$ is the LSE of β and, for each fixed θ, $\hat{\beta}(\theta)$ is a solution of

$$\|X - Z\hat{\beta}(\theta)\|^2 = \min_{\beta: L\beta = \theta} \|X - Z\beta\|^2.$$

Show that $C(X) = \{\theta : W(X, \theta) \le c_\alpha\}$ is an unbiased $1 - \alpha$ confidence set for θ, where c_α is the $(1 - \alpha)$th quantile of the F-distribution $F_{s,n-r}$.
Solution. From the discussion in §6.3.2 of Shao (2003), $W(X, \eta)$ has the noncentral F-distribution $F_{s,n-r}(\delta)$, where $\delta = \|\eta - \theta\|^2/\sigma^2$ for any $\eta = L\gamma$, $\gamma \in \mathcal{R}^p$. Hence, when θ is the true parameter value, $W(X, \theta)$ has the central F-distribution $F_{s,n-r}$ and, therefore,

$$P\big(\theta \in C(X)\big) = P\big(W(X, \theta) \le c_\alpha\big) = 1 - \alpha,$$

i.e., $C(X)$ has confidence coefficient $1 - \alpha$. If θ' is not the true parameter value,

$$P\big(\theta' \in C(X)\big) = P\big(W(X, \theta') \le c_\alpha\big) \le P\big(W(X, \theta) \le c_\alpha\big) = 1 - \alpha,$$

where the inequality follows from the fact that the noncentral F-distribution cumulative distribution function is decreasing in its noncentrality parameter (Exercise 13). Hence, $C(X)$ is unbiased. ∎

Exercise 31 (#7.48). Let $X = (X_1, ..., X_n)$ be a random sample from the exponential distribution on (a, ∞) with scale parameter θ, where $a \in \mathcal{R}$ and $\theta > 0$ are unknown. Find a UMA confidence interval for a with confidence coefficient $1 - \alpha$.

Solution. From Exercises 15 and 33 of Chapter 6, for testing $H_0 : a = a_0$ versus $H_1 : a \neq a_0$, a UMP test of size α rejects H_0 when $X_{(1)} \leq a_0$ or $2n(X_{(1)} - a_0)/V > c$, where $X_{(1)}$ is the smallest order statistic, $V = 2\sum_{i=1}^n (X_i - X_{(1)})$, and c satisfies $(n-1)\int_0^c (1+v)^{-n} dv = 1 - \alpha$. The acceptance region of this test is

$$A(a_0) = \left\{ X : 0 \leq \frac{2n(X_{(1)} - a_0)}{V} \leq c \right\}.$$

Then,

$$\begin{aligned} C(X) &= \{a : a \in A(a)\} \\ &= \left\{ a : 0 \leq \frac{2n(X_{(1)} - a)}{V} \leq c \right\} \\ &= \left[X_{(1)} - \frac{cV}{2n}, X_{(1)} \right] \end{aligned}$$

is a UMA confidence interval for a with confidence coefficient $1 - \alpha$. ∎

Exercise 32. Let $X = (X_1, ..., X_n)$ be a random sample from the uniform distribution on $(\theta, \theta + 1)$, where $\theta \in \mathcal{R}$ is unknown. Obtain a UMA lower confidence bound for θ with confidence coefficient $1 - \alpha$.

Solution. When $n \geq 2$, it follows from Exercise 13(i) in Chapter 6 (with the fact that $X_i - \theta_0$ has the uniform distribution on $(\theta - \theta_0, \theta - \theta_0 + 1)$) that a UMP test of size α for testing $H_0 : \theta = \theta_0$ versus $H_1 : \theta \neq \theta_0$ has acceptance region

$$A(\theta_0) = \{X : X_{(1)} - \theta_0 < 1 - \alpha^{1/n} \text{ and } X_{(n)} - \theta_0 < 1\},$$

where $X_{(j)}$ is the jth order statistic. When $n = 1$, by Exercise 8(iv) in Chapter 6, the family of densities of X has monotone likelihood ratio in X and, hence, the UMP test of size α rejects H_0 when $X > c$ and c satisfies $P(X \geq c) = \alpha$ when $\theta = \theta_0$, i.e., $c = 1 - \alpha + \theta_0$. Hence, the acceptance region is $\{X : X < 1 - \alpha + \theta_0\}$, which is still equal to $A(\theta_0)$ since $X_{(1)} = X_{(n)} = X$ when $n = 1$. Therefore, a $(-\infty, \theta)$-UMA confidence set for θ with confidence coefficient $1 - \alpha$ is

$$\begin{aligned} C(X) &= \{\theta : X \in A(\theta)\} \\ &= \{\theta : X_{(1)} - (1 - \alpha^{1/n}) < \theta \text{ and } X_{(n)} - 1 < \theta\} \\ &= \{\theta : \max\{X_{(1)} - (1 - \alpha^{1/n}), X_{(n)} - 1\} < \theta\} \\ &= [\underline{\theta}, \infty], \end{aligned}$$

where

$$\underline{\theta} = \max\{X_{(1)} - (1 - \alpha^{1/n}), X_{(n)} - 1\}$$

is a $(-\infty,\theta)$-UMA lower confidence bound for θ with confidence coefficient $1-\alpha$. ∎

Exercise 33 (#7.51). Let $(X_1,...,X_n)$ be a random sample from the Poisson distribution with an unknown mean $\theta>0$. Find a randomized UMA upper confidence bound for θ with confidence coefficient $1-\alpha$.
Solution. Let $Y=\sum_{i=1}^n X_i$ and $W=Y+U$, where U is a random variable that is independent of Y and has the uniform distribution on $(0,1)$. Note that Y has the Poisson distribution with mean $n\theta$. Then, for $w>0$,

$$
\begin{aligned}
P(W\le w) &= \sum_{j=0}^{\infty} P(W\le w, Y=j) \\
&= \sum_{j=0}^{\infty} P(Y=j)P(U\le w-j) \\
&= \sum_{j=0}^{\infty} \frac{e^{-n\theta}(n\theta)^j}{j!}(w-j)I_{(j,j+1]}(w)
\end{aligned}
$$

and

$$
\begin{aligned}
f_\theta(w) &= \frac{d}{dw}P(W\le w) \\
&= \sum_{j=0}^{\infty} \frac{e^{-n\theta}(n\theta)^j}{j!} I_{(j,j+1]}(w) \\
&= \frac{e^{-n\theta}(n\theta)^{[w]}}{[w]!} I_{(0,\infty)}(w),
\end{aligned}
$$

where $[w]$ is the integer part of w. For $\theta_1<\theta_2$,

$$
\frac{f_{\theta_2}(w)}{f_{\theta_1}(w)} = e^{n(\theta_1-\theta_2)}\left(\frac{\theta_2}{\theta_1}\right)^{[w]}
$$

is increasing in $[w]$ and, hence, increasing in w, i.e., the family $\{f_\theta\colon \theta>0\}$ has monotone likelihood ratio in W. Thus, for testing $H_0:\theta=\theta_0$ versus $H_1:\theta<\theta_0$, the UMP test has acceptance region $\{W: W\ge c(\theta_0)\}$, where $\int_0^{c(\theta_0)} f_{\theta_0}(w)dw=\alpha$. Let $c(\theta)$ be the function defined by $\int_0^{c(\theta)} f_\theta(w)dw=\alpha$. For $\theta_1<\theta_2$, if $c(\theta_1)>c(\theta_2)$, then

$$
\alpha = \int_0^{c(\theta_1)} f_{\theta_1}(w)dw \ge \int_0^{c(\theta_1)} f_{\theta_2}(w)dw > \int_0^{c(\theta_2)} f_{\theta_2}(w)dw = \alpha,
$$

where the first inequality follows from Lemma 6.3 in Shao (2003). Thus, we must have $c(\theta_1)\le c(\theta_2)$, i.e., $c(\theta)$ is nondecreasing in θ. Let $c^{-1}(t)=$

$\inf\{\theta : c(\theta) \geq t\}$. Then $W \geq c(\theta)$ if and only if $c^{-1}(W) \geq \theta$. Hence, a Θ'-UMA upper confidence bound with confidence coefficient $1-\alpha$ is $c^{-1}(W)$, where $\Theta' = (\theta, \infty)$. ∎

Exercise 34 (#7.52). Let X be a nonnegative integer-valued random variable from a population $P \in \mathcal{P}$. Suppose that $\mathcal{P}$ contains discrete probability densities indexed by a real-valued θ and $\mathcal{P}$ has monotone likelihood ratio in X. Let U be a random variable that has the uniform distribution on $(0,1)$ and is independent of X. Show that a UMA lower confidence bound for θ with confidence coefficient $1-\alpha$ is the solution of the equation

$$UF_\theta(X) + (1-U)F_\theta(X-1) = 1-\alpha$$

(assuming that a solution exists), where $F_\theta(x)$ is the cumulative distribution function of X.

Solution. Let $W = X + U$. Using the same argument in the solution of the previous exercise, we conclude that W has Lebesgue density

$$f_\theta(w) = F_\theta([w]) - F_\theta([w]-1),$$

where $[w]$ is the integer part of w. Note that the probability density function of F_θ with respect to the counting measure is $F_\theta(x)-F_\theta(x-1)$, $x = 0, 1, 2, \ldots$. Since $\mathcal{P}$ has monotone likelihood ratio in X,

$$\frac{F_{\theta_2}(x) - F_{\theta_2}(x-1)}{F_{\theta_1}(x) - F_{\theta_1}(x-1)}$$

is nondecreasing in X for any $\theta_1 < \theta_2$ and, hence, the family of densities of W has monotone likelihood ratio in W. For testing $H_0 : \theta = \theta_0$ versus $H_0 : \theta > \theta_0$, a UMP test of size α rejects H_0 when $W > c(\theta_0)$, where $\int_{c(\theta_0)}^\infty f_{\theta_0}(w)dw = \alpha$. Let $c(\theta)$ be the function defined by $\int_{c(\theta)}^\infty f_\theta(w)dw = \alpha$ and

$$\begin{aligned} A(\theta) &= \{W : W \leq c(\theta)\} \\ &= \left\{ W : \int_W^\infty f_\theta(w)dw \geq \int_{c(\theta)}^\infty f_\theta(w)dw = \alpha \right\}. \end{aligned}$$

Since $\int_W^\infty f_\theta(w)dw$ is nondecreasing in θ (Lemma 6.3 in Shao, 2003),

$$\begin{aligned} C(W) &= \{\theta : W \in A(\theta)\} \\ &= \left\{ \theta : \int_W^\infty f_\theta(w)dw \geq \alpha \right\} \\ &= [\underline{\theta}, \infty), \end{aligned}$$

where $\underline{\theta}$ is a solution to $\int_W^\infty f_\theta(w)dw = \alpha$ (assuming that a solution exists), i.e., a solution to $\int_0^{X+U} f_\theta(w)dw = 1-\alpha$. The result follows from

$$\begin{aligned}\int_0^{X+U} f_\theta(w)dw &= \sum_{j=0}^{X-1}\int_j^{j+1} f_\theta(w)dw + \int_X^{X+U} f_\theta(w)dw \\ &= \sum_{j=0}^{X-1}[F_\theta(j) - F_\theta(j-1)] + U[F_\theta(X) - F_\theta(X-1)] \\ &= F_\theta(X-1) + U[F_\theta(X) - F_\theta(X-1)] \\ &= UF_\theta(X) + (1-U)F_\theta(X-1). \quad \blacksquare\end{aligned}$$

Exercise 35 (#7.60(a)). Let X_1 and X_2 be independent random variables from the exponential distributions on $(0,\infty)$ with scale parameters θ_1 and θ_2, respectively. Show that $[\alpha Y/(2-\alpha), (2-\alpha)Y/\alpha]$ is a UMAU confidence interval for θ_2/θ_1 with confidence coefficient $1-\alpha$, where $Y = X_2/X_1$.

Solution. First, we need to find a UMPU test of size α for testing $H_0 : \theta_2 = \lambda\theta_1$ versus $H_1 : \theta_2 \neq \lambda\theta_1$, where $\lambda > 0$ is a known constant. The joint density of X_1 and X_2 is

$$\frac{1}{\theta_1\theta_2}\exp\left\{-\frac{X_1}{\theta_1} - \frac{X_2}{\theta_2}\right\},$$

which can be written as

$$\frac{1}{\theta_1\theta_2}\exp\left\{-X_1\left(\frac{1}{\theta_1} - \frac{\lambda}{\theta_2}\right) - (\lambda X_1 + X_2)\frac{1}{\theta_2}\right\}.$$

Hence, by Theorem 6.4 in Shao (2003), a UMPU test of size α rejects H_0 when $X_1 < c_1(U)$ or $X_1 > c_2(U)$, where $U = \lambda X_1 + X_2$. Note that X_1/X_2 is independent of U under H_0. Hence, by Lemma 6.7 of Shao (2003), the UMPU test is equivalent to the test that rejects H_0 when $X_1/X_2 < d_1$ or $X_1/X_2 > d_2$, which is equivalent to the test that rejects H_0 when $W < b_1$ or $W > b_2$, where $W = \frac{Y/\lambda}{1+Y/\lambda}$ and b_1 and b_2 satisfy $P(b_1 < W < b_2) = 1-\alpha$ (for size α) and $E[WI_{(b_1,b_2)}(W)] = (1-\alpha)E(W)$ (for unbiasedness) under H_0. When $\theta_2 = \lambda\theta_1$, W has the same distribution as $\frac{Z_1/Z_2}{1+Z_1/Z_2}$, where Z_1 and Z_2 are independent and identically distributed random variables having the exponential distribution on $(0,\infty)$ with scale parameter 1. Hence, the distribution of W under H_0 is uniform on $(0,1)$. Then the requirements on b_1 and b_2 become $b_2 - b_1 = 1-\alpha$ and $b_2^2 - b_1^2 = 1-\alpha$, which yield $b_1 = \alpha/2$ and $b_2 = 1-\alpha/2$ (assuming that $0 < \alpha < \frac{1}{2}$). Hence, the acceptance region of the UMPU test is

$$A(\lambda) = \left\{W : \frac{\alpha}{2} \leq W \leq 1 - \frac{\alpha}{2}\right\} = \left\{Y : \frac{\alpha}{2-\alpha} \leq \frac{Y}{\lambda} \leq \frac{2-\alpha}{\alpha}\right\}.$$

Inverting $A(\lambda)$ leads to

$$\{\lambda : \lambda \in A(\lambda)\} = \left[\frac{\alpha Y}{2-\alpha}, \frac{(2-\alpha)Y}{\alpha}\right],$$

which is a UMAU confidence interval for $\lambda = \theta_2/\theta_1$ with confidence coefficient $1-\alpha$. ∎

Exercise 36 (#7.63). Let $(X_{i1},...,X_{in_i})$, $i = 1, 2$, be two independent random samples from $N(\mu_i, \sigma_i^2)$, $i = 1, 2$, respectively, where all parameters are unknown. Let $\theta = \mu_1 - \mu_2$, $\bar{X}_i$ and S_i^2 be the sample mean and sample variance of the ith sample, $i = 1, 2$.
(i) Show that

$$R(X,\theta) = \frac{\bar{X}_1 - \bar{X}_2 - \theta}{\sqrt{n_1^{-1}S_1^2 + n_2^{-1}S_2^2}}$$

is asymptotically pivotal, assuming that $n_1/n_2 \to c \in (0, \infty)$. Construct a $1-\alpha$ asymptotically correct confidence interval for θ using $R(X,\theta)$.
(ii) Show that

$$t(X,\theta) = \frac{(\bar{X}_1 - \bar{X}_2 - \theta)/\sqrt{n_1^{-1} + n_2^{-1}}}{\sqrt{[(n_1-1)S_1^2 + (n_2-1)S_2^2]/(n_1+n_2-2)}}$$

is asymptotically pivotal if either $n_1/n_2 \to 1$ or $\sigma_1 = \sigma_2$ holds.
Solution. (i) Note that

$$R(X,\theta) = \frac{\bar{X}_1 - \bar{X}_2 - \theta}{\sqrt{n_1^{-1}\sigma_1^2 + n_2^{-1}\sigma_2^2}} \frac{\sqrt{\sigma_1^2 + (n_1/n_2)\sigma_2^2}}{\sqrt{S_1^2 + (n_1/n_2)S_2^2}} \to_d N(0,1),$$

because $\bar{X}_1 - \bar{X}_2$ is distributed as $N(\theta, n_1^{-1}\sigma_1^2 + n_2^{-1}\sigma_2^2)$ and

$$\frac{\sqrt{\sigma_1^2 + (n_1/n_2)\sigma_2^2}}{\sqrt{S_1^2 + (n_1/n_2)S_2^2}} \to_p \frac{\sqrt{\sigma_1^2 + c\sigma_2^2}}{\sqrt{\sigma_1^2 + c\sigma_2^2}} = 1$$

by the fact that $S_i^2 \to_p \sigma_i^2$, $i = 1, 2$. Therefore, $R(X,\theta)$ is asymptotically pivotal. A $1-\alpha$ asymptotically correct confidence interval for θ is

$$\left[\bar{X}_1 - \bar{X}_2 - z_{\alpha/2}\sqrt{n_1^{-1}S_1^2 + n_2^{-1}S_2^2}, \bar{X}_1 - \bar{X}_2 + z_{\alpha/2}\sqrt{n_1^{-1}S_1^2 + n_2^{-1}S_2^2}\right],$$

where z_α is the $(1-\alpha)$th quantile of $N(0,1)$.
(ii) If $\sigma_1^2 = \sigma_2^2$, then $t(X,\theta)$ has the t-distribution $t_{n_1+n_2-2}$. Consider now the case where $\sigma_1 \neq \sigma_2$ but $n_1/n_2 \to 1$. Note that

$$t(X,\theta) = \frac{\bar{X}_1 - \bar{X}_2 - \theta}{\sqrt{n_1^{-1}\sigma_1^2 + n_2^{-1}\sigma_2^2}} g(X),$$

where

$$g(X) = \frac{\sqrt{n_1+n_2-2}\sqrt{\sigma_1^2+(n_1/n_2)\sigma_2^2}}{\sqrt{n_1+n_2}\sqrt{[(n_1-1)/n_2]S_1^2+[(n_2-1)/n_1]S_2^2}} \to_p 1$$

when $n_1/n_2 \to 1$. Then, $t(X,\theta) \to_d N(0,1)$ and, therefore, is asymptotically pivotal. ∎

Exercise 37 (#7.64). Let $(X_1, ..., X_n)$ be a random sample of binary random variables with unknown $p = P(X_i = 1)$.
(i) The confidence set for p obtained by inverting acceptance regions of Rao's score tests is

$$C_3(X) = \{p : n(\hat{p}-p)^2 \le p(1-p)\chi^2_{1,\alpha}\}.$$

where $\hat{p} = n^{-1}\sum_{i=1}^n X_i$ and $\chi^2_{1,\alpha}$ is the $(1-\alpha)$th quantile of χ^2_1. Show that $C_3(X) = [p_-, p_+]$ with

$$p_\pm = \frac{2n\hat{p}+\chi^2_{1,\alpha} \pm \sqrt{\chi^2_{1,\alpha}[4n\hat{p}(1-\hat{p})+\chi^2_{1,\alpha}]}}{2(n+\chi^2_{1,\alpha})}.$$

(ii) Compare the length of $C_3(X)$ with

$$C_2(X) = [\,\hat{p} - z_{1-\alpha/2}\sqrt{\hat{p}(1-\hat{p})/n},\ \hat{p} + z_{1-\alpha/2}\sqrt{\hat{p}(1-\hat{p})/n}\,],$$

the confidence set for p obtained by inverting acceptance regions of Wald's tests.
Solution. (i) Let $g(p) = (n+\chi^2_{1,\alpha})p^2 - (2n\hat{p}+\chi^2_{1,\alpha})p + n\hat{p}^2$. Then $C_3(X) = \{p : g(p) \le 0\}$. Since $g(p)$ is a quadratic form of p with $g''(p) > 0$, $C_3(X)$ is an interval whose limits are two real solutions of $g(p)$. The result follows from the fact that $p_\pm$ are the two real solutions to $g(p) = 0$.
(ii) The length of the interval $C_3(X)$ is

$$l_3(X) = \frac{\sqrt{\chi^2_{1,\alpha}[4n\hat{p}(1-\hat{p})+\chi^2_{1,\alpha}]}}{n+\chi^2_{1,\alpha}}$$

and the length of the interval $C_2(X)$ is

$$l_2(X) = \sqrt{\frac{4\chi^2_{1,\alpha}\hat{p}(1-\hat{p})}{n}}.$$

Since

$$[l_2(X)]^2 - [l_3(X)]^2 = \frac{(\chi^2_{1,\alpha})^2[(8n+4\chi^2_{1,\alpha})\hat{p}(1-\hat{p})-n]}{n(n+\chi^2_{1,\alpha})^2},$$

we conclude that $l_2(X) \geq l_3(X)$ if and only if

$$\hat{p}(1-\hat{p}) \geq \frac{n}{4(2n+\chi^2_{1,\alpha})}. \quad \blacksquare$$

Exercise 38 (#7.67). Let $X = (X_1, ..., X_n)$ be a random sample from the Poisson distribution with unknown mean $\theta > 0$ and $\bar{X}$ be the sample mean.
(i) Show that $R(X,\theta) = (\bar{X}-\theta)/\sqrt{\theta/n}$ is asymptotically pivotal. Construct a $1-\alpha$ asymptotically correct confidence interval for θ, using $R(X,\theta)$.
(ii) Show that $R_1(X,\theta) = (\bar{X}-\theta)/\sqrt{\bar{X}/n}$ is asymptotically pivotal. Derive a $1-\alpha$ asymptotically correct confidence interval for θ, using $R_1(X,\theta)$.
(iii) Obtain $1-\alpha$ asymptotically correct confidence intervals for θ by inverting acceptance regions of LR tests, Wald's tests, and Rao's score tests.
Solution. (i) Since $E(X_1) = \text{Var}(X_1) = \theta$, the central limit theorem implies that $\sqrt{n}(\bar{X}-\theta) \to_d N(0,\theta)$. Thus, $R(X,\theta) = \sqrt{n}(\bar{X}-\theta)/\sqrt{\theta} \to_d N(0,1)$ and is asymptotically pivotal. Let z_α be the $(1-\alpha)$th quantile of $N(0,1)$. A $1-\alpha$ asymptotically correct confidence set for θ is

$$C(X) = \{\theta : [R(X,\theta)]^2 \leq z^2_{\alpha/2}\} = \{\theta : n\theta^2 - (2n\bar{X} + z^2_{\alpha/2})\theta + n\bar{X}^2 \leq 0\}.$$

Since the quadratic form $n\theta^2 - (2n\bar{X} + z^2_{\alpha/2})\theta + n\bar{X}^2$ has two real roots

$$\theta_{\pm} = \frac{2n\bar{X} + z^2_{\alpha/2} \pm \sqrt{4n\bar{X}z^2_{\alpha/2} + z^4_{\alpha/2}}}{2n},$$

we conclude that $C(X) = [\theta_-, \theta_+]$ is an interval.
(ii) By the law of large numbers, $\bar{X} \to_p \theta$. By the result in part (i) of the solution and Slutsky's theorem, $R(X,\theta) = \sqrt{n}(\bar{X}-\theta)/\sqrt{\bar{X}} \to_d N(0,1)$ and is asymptotically pivotal. A $1-\alpha$ asymptotically correct confidence set for θ is

$$C_1(X) = \{\theta : [R_1(X,\theta)]^2 \leq z^2_{\alpha/2}\} = \{\theta : n(\bar{X}-\theta)^2 \leq \bar{X}z^2_{\alpha/2}\},$$

which is the interval $[\bar{X} - z_{\alpha/2}\bar{X}/\sqrt{n}, \bar{X} + z_{\alpha/2}\bar{X}/\sqrt{n}]$.
(iii) The likelihood function is

$$\ell(\theta) = e^{-n\theta}\theta^{n\bar{X}} \prod_{i=1}^n \frac{1}{X_i!}.$$

Then

$$\frac{\partial \log \ell(\theta)}{\partial \theta} = -n + \frac{n\bar{X}}{\theta}$$

and

$$\frac{\partial^2 \log \ell(\theta)}{\partial \theta^2} = -\frac{n\bar{X}}{\theta^2}.$$

The MLE of θ is $\bar{X}$. Since $E(\bar{X}) = \theta$, the Fisher information is $I_n(\theta) = n/\theta$. For testing $H_0 : \theta = \theta_0$ versus $H_1 : \theta \neq \theta_0$, Wald's test statistic is

$$(\bar{X} - \theta_0)^2 I_n(\bar{X}) = (\bar{X} - \theta_0)^2 (n/\bar{X}) = [R_1(X, \theta_0)]^2.$$

Hence, the confidence interval for θ obtained by inverting acceptance regions of Wald's tests is $C_1(X)$ given in part (ii) of the solution. Since θ_0 is the MLE of θ under H_0, Rao's score test statistic is

$$\left(-n + \frac{n\bar{X}}{\theta_0}\right)^2 [I_n(\theta_0)]^{-1} = [R(X, \theta_0)]^2.$$

Hence, the confidence interval for θ obtained by inverting acceptance regions of Rao's score tests is $C(X)$ given in part (i) of the solution. The likelihood ratio for testing $H_0 : \theta = \theta_0$ versus $H_1 : \theta \neq \theta_0$ is

$$\lambda = e^{n(\bar{X} - \theta_0)} \left(\frac{\theta_0}{\bar{X}}\right)^{n\bar{X}}.$$

Note that $\lambda \geq c$ for some c is equivalent to $c_1 \leq \bar{X}/\theta_0 \leq c_2$ for some c_1 and c_2. Hence, the confidence interval for θ obtained by inverting acceptance regions of LR tests is $[c_1\bar{X}, c_2\bar{X}]$, where c_1 and c_2 are constants such that $\lim_n P(c_1\bar{X} \leq \theta \leq c_2\bar{X}) = 1 - \alpha$. ∎

Exercise 39 (#7.70). Let $X = (X_1, ..., X_n)$ be a random sample from $N(\mu, \varphi)$ with unknown $\theta = (\mu, \varphi)$. Obtain $1 - \alpha$ asymptotically correct confidence sets for μ by inverting acceptance regions of LR tests, Wald's tests, and Rao's score tests. Are these sets always intervals?
Solution. The log-likelihood function is

$$\log \ell(\theta) = -\frac{1}{2\varphi} \sum_{i=1}^{n} (X_i - \mu)^2 - \frac{n}{2} \log \varphi - \frac{n}{2} \log(2\pi).$$

Note that

$$s_n(\theta) = \frac{\partial \log \ell(\theta)}{\partial \theta} = \left(\frac{n(\bar{X} - \mu)}{\varphi}, \frac{1}{2\varphi^2} \sum_{i=1}^{n} (X_i - \mu)^2 - \frac{n}{2\varphi} \right)$$

and the Fisher information is

$$I_n(\theta) = n \begin{pmatrix} \frac{1}{\varphi} & 0 \\ 0 & \frac{1}{2\varphi^2} \end{pmatrix}.$$

The MLE of θ is $\hat{\theta} = (\bar{X}, \hat{\varphi})$, where $\bar{X}$ is the sample mean and $\hat{\varphi} = n^{-1} \sum_{i=1}^{n} (X_i - \bar{X})^2$.

Consider testing $H_0 : \mu = \mu_0$ versus $H_1 : \mu \neq \mu_0$. For Wald's test, $R(\theta) = \mu - \mu_0$ with $C = \partial R/\partial\theta = (1, 0)$. Hence, Wald's test statistic is

$$[R(\hat{\theta})]^2 \{C^\tau [I_n(\hat{\theta})]^{-1} C\}^{-1} = \frac{n(\bar{X} - \mu_0)^2}{\hat{\varphi}}.$$

Let z_α be the $(1-\alpha)$th quantile of $N(0,1)$. The $1-\alpha$ asymptotically correct confidence set obtained by inverting the acceptance regions of Wald's tests is

$$\left\{ \mu : \frac{n(\bar{X} - \mu)^2}{\hat{\varphi}} \leq z_{\alpha/2}^2 \right\},$$

which is the interval

$$\left[\bar{X} - z_{\alpha/2}\sqrt{\hat{\varphi}/n}, \bar{X} + z_{\alpha/2}\sqrt{\hat{\varphi}/n} \right].$$

Under H_0, the MLE of φ is $n^{-1}\sum_{i=1}^n (X_i - \mu_0)^2 = \hat{\varphi} + (\bar{X} - \mu_0)^2$. Then the likelihood ratio is

$$\lambda = \left(\frac{\hat{\varphi}}{\hat{\varphi} + (\bar{X} - \mu_0)^2} \right)^{n/2}.$$

The asymptotic LR test rejects H_0 when $\lambda < e^{-z_{\alpha/2}^2/2}$, i.e.,

$$(\bar{X} - \mu_0)^2 > (e^{z_{\alpha/2}^2/n} - 1)\hat{\varphi}.$$

Hence, the $1-\alpha$ asymptotically correct confidence set obtained by inverting the acceptance regions of asymptotic LR tests is

$$\left\{ \mu : (\bar{X} - \mu_0)^2 \leq (e^{z_{\alpha/2}^2/n} - 1)\hat{\varphi} \right\},$$

which is the interval

$$\left[\bar{X} - \sqrt{(e^{z_{\alpha/2}^2/n} - 1)\hat{\varphi}}, \bar{X} + \sqrt{(e^{z_{\alpha/2}^2/n} - 1)\hat{\varphi}} \right].$$

Let $\tilde{\theta} = (\mu_0, \hat{\varphi} + (\bar{X} - \mu_0)^2)$ be the MLE of θ under H_0. Then Rao's score test statistic is

$$R_n^2 = [s(\tilde{\theta})]^\tau [I_n(\tilde{\theta})]^{-1} s(\tilde{\theta}).$$

Note that

$$s(\tilde{\theta}) = \left(\frac{n(\bar{X} - \mu_0)}{\hat{\varphi} + (\bar{X} - \mu_0)^2}, \ 0 \right).$$

Hence,

$$R_n^2 = \frac{n(\bar{X} - \mu_0)^2}{\hat{\varphi} + (\bar{X} - \mu_0)^2}$$

and the $1-\alpha$ asymptotically correct confidence set obtained by inverting the acceptance regions of Rao's score tests is

$$\left\{\mu : (n - z_{\alpha/2}^2)(\bar{X} - \mu)^2 \le z_{\alpha/2}^2 \hat{\varphi}\right\},$$

which is the interval

$$\left[\bar{X} - z_{\alpha/2}\sqrt{\hat{\varphi}/(n - z_{\alpha/2}^2)}, \bar{X} + z_{\alpha/2}\sqrt{\hat{\varphi}/(n - z_{\alpha/2}^2)}\right]. \blacksquare$$

Exercise 40. Let $(X_1, ..., X_n)$ be a random sample from a distribution with mean μ, variance σ^2, and finite 4th moment. Derive a $1-\alpha$ asymptotically correct confidence interval for $\theta = \mu/\sigma$.

Solution. Let $\bar{X}$ be the sample mean and $\hat{\sigma}^2 = n^{-1}\sum_{i=1}^n (X_i - \bar{X})^2$. It follows from Example 2.8 in Shao (2003) that

$$\sqrt{n}\left[\begin{pmatrix} \bar{X} \\ \hat{\sigma}^2 \end{pmatrix} - \begin{pmatrix} \mu \\ \sigma^2 \end{pmatrix}\right] \to_d N_2\left(0, \begin{pmatrix} \sigma^2 & \gamma \\ \gamma & \kappa \end{pmatrix}\right),$$

where $\gamma = E(X_1 - \mu)^3$ and $\kappa = E(X_1 - \mu)^4 - \sigma^4$. Let $g(x, y) = x/\sqrt{y}$. Then $\partial g/\partial x = 1/\sqrt{y}$ and $\partial g/\partial y = -x/(2y^{3/2})$. By the δ-method,

$$\sqrt{n}\left(\frac{\bar{X}}{\hat{\sigma}} - \frac{\mu}{\sigma}\right) \to_d N\left(0, 1 + \mu^2\kappa/(4\sigma^6) - \mu\gamma/\sigma^4\right).$$

Let

$$\hat{\gamma} = \frac{1}{n}\sum_{i=1}^n (X_i - \bar{X})^3$$

and

$$\hat{\kappa} = \frac{1}{n}\sum_{i=1}^n (X_i - \bar{X})^4 - \hat{\sigma}^4.$$

By the law of large numbers, $\hat{\gamma} \to_p \gamma$ and $\hat{\kappa} \to_p \kappa$. Let $W = 1 + \bar{X}^2\hat{\kappa}/(4\hat{\sigma}^6) - \bar{X}\hat{\gamma}/\hat{\sigma}^4$. By Slutsky's theorem,

$$\frac{\sqrt{n}}{\sqrt{W}}\left(\frac{\bar{X}}{\hat{\sigma}} - \theta\right) \to_d N(0, 1)$$

and, hence, a $1-\alpha$ asymptotically correct confidence interval for θ is

$$\left[\frac{\bar{X}}{\hat{\sigma}} - z_{\alpha/2}\frac{\sqrt{W}}{\sqrt{n}}, \frac{\bar{X}}{\hat{\sigma}} + z_{\alpha/2}\frac{\sqrt{W}}{\sqrt{n}}\right],$$

where z_α is the $(1-\alpha)$th quantile of $N(0, 1)$. $\blacksquare$

Exercise 41. Consider the linear model $X = Z\beta + \varepsilon$, where Z is an $n \times p$ matrix of full rank, $\beta \in \mathcal{R}^p$, $\varepsilon = (\varepsilon_1, ..., \varepsilon_n)$ with independent and identically distributed ε_i's, $E(\varepsilon_i) = 0$, and $\text{Var}(\varepsilon_i) = \sigma^2$. Let Z_i be the ith row of Z. Assume that $\lim_n \max_{1\leq i\leq n} Z_i^\tau (Z^\tau Z) Z_i = 0$. Find an asymptotically pivotal quantity and construct a $1-\alpha$ asymptotically correct confidence set for β.
Solution. Let $\hat{\beta}$ be the LSE of β. Note that $\text{Var}(\hat{\beta}) = \sigma^2 (Z^\tau Z)^{-1}$. By Theorem 3.12 in Shao (2003),

$$\sigma^{-1}(Z^\tau Z)^{1/2}(\hat{\beta} - \beta) \to_d N_p(0, I_p).$$

Let $\hat{\sigma}^2 = \|X - Z\hat{\beta}\|^2/n$. Since

$$\begin{aligned} X_i - Z_i^\tau \hat{\beta} &= X_i - Z_i^\tau \beta + Z_i^\tau(\beta - \hat{\beta}) \\ &= \varepsilon_i Z_i^\tau (\beta - \hat{\beta}), \end{aligned}$$

we obtain that

$$\begin{aligned} \hat{\sigma}^2 &= \frac{1}{n}\sum_{i=1}^n (X_i - Z_i^\tau \hat{\beta})^2 \\ &= \frac{1}{n}\sum_{i=1}^n \varepsilon_i^2 + \frac{1}{n}\sum_{i=1}^n [Z_i^\tau(\hat{\beta} - \beta)]^2 - \frac{2}{n}\sum_{i=1}^n \varepsilon_i Z_i^\tau (\hat{\beta} - \beta). \end{aligned}$$

By the law of large numbers, $n^{-1}\sum_{i=1}^n \varepsilon_i^2 \to_p \sigma^2$. By the Cauchy-Schwartz inequality,

$$\frac{1}{n}\sum_{i=1}^n [Z_i^\tau(\hat{\beta} - \beta)]^2 \leq \frac{1}{n}\sum_{i=1}^n Z_i^\tau (Z^\tau Z)^{-1} Z_i [(Z^\tau Z)(\hat{\beta} - \beta)]^2,$$

which is bounded by $[(Z^\tau Z)(\hat{\beta} - \beta)]^2 \max_{1\leq i\leq n} Z_i^\tau (Z^\tau Z)^{-1} Z_i \to_p 0$. By the Cauchy-Schwartz inequality again,

$$\left[\frac{1}{n}\sum_{i=1}^n \varepsilon_i Z_i^\tau(\hat{\beta} - \beta)\right]^2 \leq \left(\frac{1}{n}\sum_{i=1}^n \varepsilon_i^2\right)\left\{\frac{1}{n}\sum_{i=1}^n [Z_i^\tau(\hat{\beta} - \beta)]^2\right\} \to_p 0.$$

Hence, $\hat{\sigma}^2 \to_p \sigma^2$ and $\hat{\sigma}^{-1}(Z^\tau Z)^{1/2}(\hat{\beta} - \beta)$ is asymptotically pivotal. A $1-\alpha$ asymptotically correct confidence set for β is then

$$\left\{\beta : (\hat{\beta} - \beta)^\tau (Z^\tau Z)(\hat{\beta} - \beta) \leq \hat{\sigma}^2 \chi^2_{p,\alpha}\right\},$$

where $\chi^2_{p,\alpha}$ is the $(1-\alpha)$th quantile of the chi-square distribution χ^2_p. ∎

Exercise 42 (#7.81). Let $(X_{i1}, ..., X_{in_i})$, $i = 1, 2$, be two independent random samples from $N(\mu_i, \sigma_i^2)$, $i = 1, 2$, respectively, where all parameters

are unknown.
(i) Find $1-\alpha$ asymptotically correct confidence sets for (μ_1, μ_2) by inverting acceptance regions of LR tests, Wald's tests, and Rao's score tests.
(ii) Repeat (i) for the parameter $(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2)$.
(iii) Repeat (i) under the assumption that $\sigma_1^2 = \sigma_2^2 = \sigma^2$.
(iv) Repeat (iii) for the parameter (μ_1, μ_2, σ^2).
Solution. (i) The likelihood function is proportional to

$$\frac{1}{\sigma_1^{n_1}} \frac{1}{\sigma_2^{n_2}} \exp\left\{ -\sum_{i=1}^{2} \frac{1}{2\sigma_i^2} \sum_{j=1}^{n_i} (X_{ij} - \mu_i)^2 \right\}.$$

Hence, the score function is

$$\left(\frac{n_1(\bar{X}_1 - \mu_1)}{\sigma_1^2}, \frac{n_1\sigma_1^2(\mu_1)}{2\sigma_1^4} - \frac{n_1}{2\sigma_1^2}, \frac{n_2(\bar{X}_2 - \mu_2)}{\sigma_2^2}, \frac{n_2\sigma_2^2(\mu_2)}{2\sigma_2^4} - \frac{n_2}{2\sigma_2^2} \right),$$

where $\bar{X}_i$ is the sample mean of the ith sample and

$$\sigma_i^2(t) = \frac{1}{n_i} \sum_{j=1}^{n_i} (X_{ij} - t)^2, \quad i = 1, 2,$$

and the Fisher information is the 4×4 diagonal matrix whose diagonal elements are n_1/σ_1^2, $n_1/(2\sigma_1^4)$, n_2/σ_2^2, and $n_2/(2\sigma_2^4)$. The MLE of $(\mu_1, \sigma_1^2, \mu_2, \sigma_2^2)$ is $(\bar{X}_1, \hat{\sigma}_1^2, \bar{X}_2, \hat{\sigma}_2^2)$, where $\hat{\sigma}_i^2 = \sigma_i^2(\bar{X}_i)$. When μ_i is known, the MLE of σ_i^2 is $\sigma_i^2(\mu_i)$, $i = 1, 2$. Thus, the $1 - \alpha$ asymptotically correct confidence set for (μ_1, μ_2) by inverting acceptance regions of LR tests is

$$\left\{ (\mu_1, \mu_2) : [\sigma_1^2(\mu_1)]^{n_1/2} [\sigma_2^2(\mu_2)]^{n_2/2} \leq \hat{\sigma}_1^{n_1} \hat{\sigma}_2^{n_2} e^{\chi^2_{2,\alpha}/2} \right\},$$

where $\chi^2_{r,\alpha}$ is the $(1-\alpha)$th quantile of the chi-square distribution χ^2_r. The $1-\alpha$ asymptotically correct confidence set for (μ_1, μ_2) by inverting acceptance regions of Wald's tests is

$$\left\{ (\mu_1, \mu_2) : \frac{n_1(\bar{X}_1 - \mu_1)^2}{\hat{\sigma}_1^2} + \frac{n_2(\bar{X}_2 - \mu_2)^2}{\hat{\sigma}_2^2} \leq \chi^2_{2,\alpha} \right\}$$

and the $1-\alpha$ asymptotically correct confidence set for (μ_1, μ_2) by inverting acceptance regions of Rao's score tests is

$$\left\{ (\mu_1, \mu_2) : \frac{n_1(\bar{X}_1 - \mu_1)^2}{\sigma_1^2(\mu_1)} + \frac{n_2(\bar{X}_2 - \mu_2)^2}{\sigma_2^2(\mu_2)} \leq \chi^2_{2,\alpha} \right\}.$$

(ii) The $1 - \alpha$ asymptotically correct confidence set for $\theta = (\mu_1, \mu_2, \sigma_1^2, \sigma_2^2)$ by inverting acceptance regions of LR tests is

$$\left\{ \theta : \sigma_1^{n_1} \sigma_2^{n_2} e^{n_1\sigma_1^2(\mu_1)/(2\sigma_1^2) + n_2\sigma_2^2(\mu_2)/(2\sigma_2^2)} \leq \hat{\sigma}_1^{n_1} \hat{\sigma}_2^{n_2} e^{(n_1+n_2+\chi^2_{4,\alpha})/2} \right\}.$$

The $1-\alpha$ asymptotically correct confidence set for θ by inverting acceptance regions of Wald's tests is

$$\left\{\theta : \sum_{i=1}^{2}\left[\frac{n_i(\bar{X}_i-\mu_i)^2}{\hat{\sigma}_i^2}+\frac{n_i(\hat{\sigma}_i^2-\sigma_i^2)^2}{2\hat{\sigma}_i^4}\right]\leq \chi^2_{4,\alpha}\right\}.$$

The $1-\alpha$ asymptotically correct confidence set for θ by inverting acceptance regions of Rao's score tests is

$$\left\{\theta : \sum_{i=1}^{2}\left[\frac{n_i(\bar{X}_i-\mu_i)^2}{\sigma_i^2}+\frac{2\sigma_i^4}{n_i}\left(\frac{n_i\sigma_i^2(\mu_i)}{2\sigma_i^4}-\frac{n_i}{2\sigma_i^2}\right)^2\right]\leq \chi^2_{4,\alpha}\right\}.$$

(iii) The MLE of (μ_1,μ_2,σ^2) is $(\bar{X}_1,\bar{X}_2,\hat{\sigma}^2)$, where

$$\hat{\sigma}^2=\frac{1}{n_1+n_2}\sum_{i=1}^{2}\sum_{j=1}^{n_i}(X_{ij}-\bar{X}_i)^2.$$

When μ_1 and μ_2 are known, the MLE of σ^2 is

$$\sigma^2(\mu_1,\mu_2)=\frac{1}{n_1+n_2}\sum_{i=1}^{2}\sum_{j=1}^{n_i}(X_{ij}-\mu_i)^2.$$

The $1-\alpha$ asymptotically correct confidence set for (μ_1,μ_2) by inverting acceptance regions of LR tests is

$$\left\{(\mu_1,\mu_2) : \sigma^2(\mu_1,\mu_2)\leq e^{-\chi^2_{2,\alpha}/[2(n_1+n_2)]}\hat{\sigma}^2\right\}.$$

The score function in this case is

$$\left(\frac{n_1(\bar{X}_1-\mu_1)}{\sigma^2},\frac{n_2(\bar{X}_2-\mu_2)}{\sigma^2},\frac{(n_1+n_2)\sigma^2(\mu_1,\mu_2)}{2\sigma^4}-\frac{n_1+n_2}{2\sigma^2}\right).$$

The Fisher information is the 3×3 diagonal matrix whose diagonal elements are n_1/σ^2, n_2/σ^2 and $(n_1+n_2)/(2\sigma^4)$. Hence, the $1-\alpha$ asymptotically correct confidence set for (μ_1,μ_2) by inverting acceptance regions of Wald's tests is

$$\left\{(\mu_1,\mu_2) : n_1(\bar{X}_1-\mu_1)^2+n_2(\bar{X}_2-\mu_2)^2\leq \hat{\sigma}^2\chi^2_{2,\alpha}\right\}.$$

The $1-\alpha$ asymptotically correct confidence set for (μ_1,μ_2) by inverting acceptance regions of Rao's score tests is

$$\left\{(\mu_1,\mu_2) : n_1(\bar{X}_1-\mu_1)^2+n_2(\bar{X}_2-\mu_2)^2\leq \sigma^2(\mu_1,\mu_2)\chi^2_{2,\alpha}\right\}.$$

(iv) The $1-\alpha$ asymptotically correct confidence set for $\theta = (\mu_1, \mu_2, \sigma^2)$ by inverting acceptance regions of LR tests is

$$\left\{\theta : \sigma^2 e^{\sigma^2(\mu_1,\mu_2)/\sigma_2^2} \leq \hat{\sigma}^2 e^{1+\chi^2_{3,\alpha}/(n_1+n_2)}\right\}.$$

The $1-\alpha$ asymptotically correct confidence set for θ by inverting acceptance regions of Wald's tests is

$$\left\{\theta : \sum_{i=1}^{2} \frac{n_i(\bar{X}_i - \mu_i)^2}{\hat{\sigma}^2} + \frac{(n_1+n_2)(\hat{\sigma}^2 - \sigma^2)^2}{2\hat{\sigma}^4} \leq \chi^2_{3,\alpha}\right\}.$$

The $1-\alpha$ asymptotically correct confidence set for θ by inverting acceptance regions of Rao's score tests is

$$\left\{\theta : \sum_{i=1}^{2} \frac{n_i(\bar{X}_i - \mu_i)^2}{\sigma^2} + \frac{n_1+n_2}{2}\left[\frac{\sigma^2(\mu_1,\mu_2)}{\sigma^2} - 1\right]^2 \leq \chi^2_{3,\alpha}\right\}. \blacksquare$$

Exercise 43 (#7.83). Let X be a vector of n observations having distribution $N_n(Z\beta, \sigma^2 I_n)$, where Z is a known $n \times p$ matrix of rank $r \leq p < n$, β is an unknown p-vector, and $\sigma^2 > 0$ is unknown. Find $1-\alpha$ asymptotically correct confidence sets for $\theta = L\beta$ by inverting acceptance regions of LR tests, Wald's tests, and Rao's score tests, where L is an $s \times p$ matrix of rank s and all rows of L are in $\mathcal{R}(Z)$.

Solution. Since the rows of L are in $\mathcal{R}(Z)$, $L = AZ$ for an $s \times n$ matrix A with rank s. Since Z is of rank r, there is an $n \times r$ matrix Z_* of rank r such that $Z\beta = Z_* Q\beta$, where Q is $r \times p$. Then, $L\beta = AZ\beta = AZ_*\beta_*$ with $\beta_* = Q\beta \in \mathcal{R}^r$. Hence, without loss of generality, in the following we assume that $r = p$. The likelihood function is

$$\ell(\beta, \sigma^2) = \left(\frac{1}{2\pi\sigma^2}\right)^{n/2} \exp\left\{-\frac{1}{2\sigma^2}\|X - Z\beta\|^2\right\}.$$

Then,

$$\frac{\partial \log \ell(\beta, \sigma^2)}{\partial \beta} = \frac{Z^\tau(X - Z\beta)}{\sigma^2},$$

$$\frac{\partial \log \ell(\beta, \sigma^2)}{\partial \sigma^2} = \frac{\|X - Z\beta\|^2}{2\sigma^4} - \frac{n}{2\sigma^2},$$

and the Fisher information matrix is

$$I_n(\beta, \sigma^2) = \begin{pmatrix} \frac{Z^\tau Z}{\sigma^2} & 0 \\ 0 & \frac{n}{2\sigma^4} \end{pmatrix}.$$

The MLE of β is the LSE $\hat{\beta} = (Z^\tau Z)^{-1} Z^\tau X$ and the MLE of σ^2 is $\hat{\sigma}^2 = \|X - Z\hat{\beta}\|^2/n$. Hence, the $1-\alpha$ asymptotic correct confidence set obtained by inverting acceptance regions of Wald's tests is

$$\left\{\theta : (L\hat{\beta} - \theta)^\tau [L(Z^\tau Z)^{-1} L^\tau]^{-1} (L\hat{\beta} - \theta) \leq \hat{\sigma}^2 \chi^2_{s,\alpha}\right\}$$

and the $1-\alpha$ asymptotic correct confidence set obtained by inverting acceptance regions of Rao's score tests is

$$\left\{\theta : n[X - Z\hat{\beta}(\theta)]^\tau Z(Z^\tau Z)^{-1} Z^\tau [X - Z\hat{\beta}(\theta)] \leq \|X - Z\hat{\beta}(\theta)\|^2 \chi^2_{s,\alpha}\right\},$$

where $\hat{\beta}(\theta)$ is defined as

$$\|X - Z\hat{\beta}(\theta)\|^2 = \min_{\beta : L\beta = \theta} \|X - Z\beta\|^2.$$

Following the discussion in Example 6.20 of Shao (2003), the likelihood ratio for testing $H_0 : L\beta = \theta$ versus $H_1 : L\beta \neq \theta$ is

$$\lambda(\theta) = \left[\frac{sW(X,\theta)}{n-r} + 1\right]^{-n/2},$$

where $W(X,\theta)$ is given in Exercise 30. Hence, the $1-\alpha$ asymptotic correct confidence set obtained by inverting acceptance regions of LR tests is

$$\left\{\theta : sW(X,\theta) \leq (n-r)(e^{\chi^2_{s,\alpha}/n} - 1)\right\}. \blacksquare$$

Exercise 44 (#7.85, #7.86). Let $(X_1, ..., X_n)$ be a random sample from a continuous cumulative distribution function F on $\mathcal{R}$ that is twice differentiable at $\theta = F^{-1}(p)$, $0 < p < 1$, with $F'(\theta) > 0$.
(i) Let $\{k_n\}$ be a sequence of integers satisfying $k_n/n = p + cn^{-1/2} + o(n^{-1/2})$ with a constant c. Show that

$$\sqrt{n}(X_{(k_n)} - \hat{\theta}) = c/F'(\theta) + o(1) \text{ a.s.},$$

where $X_{(j)}$ is the jth order statistic and $\hat{\theta}$ is the sample pth quantile.
(ii) Show that $\sqrt{n}(X_{(k_n)} - \theta)F'(\theta) \to_d N(c, p(1-p))$.
(iii) Let $\{k_{1n}\}$ and $\{k_{2n}\}$ be two sequences of integers satisfying $1 \leq k_{1n} < k_{2n} \leq n$,

$$k_{1n}/n = p - z_{\alpha/2}\sqrt{p(1-p)/n} + o(n^{-1/2}),$$

and

$$k_{2n}/n = p + z_{\alpha/2}\sqrt{p(1-p)/n} + o(n^{-1/2}),$$

where z_α is the $(1-\alpha)$th quantile of $N(0,1)$. Let $C(X) = [X_{(k_{1n})}, X_{(k_{2n})}]$. Show that $\lim_n P(\theta \in C(X)) = 1 - \alpha$, using the result in part (ii).

(iv) Construct a consistent estimator of the asymptotic variance of the sample median, using the interval $C(X)$.

Solution. (i) By the Bahadur representation (e.g., Theorem 7.8 in Shao, 2003),

$$X_{(k_n)} = \theta + \frac{(k_n/n) - F_n(\theta)}{F'(\theta)} + o\left(\frac{1}{\sqrt{n}}\right) \quad \text{a.s.},$$

where F_n is the empirical distribution. For the pth sample quantile, the Bahadur representation is

$$\hat{\theta} = \theta + \frac{p - F_n(\theta)}{F'(\theta)} + o\left(\frac{1}{\sqrt{n}}\right) \quad \text{a.s.}$$

The result follows by taking the difference of the two previous equations.

(ii) Note that

$$\sqrt{n}(X_{(k_n)} - \theta)F'(\theta) = \sqrt{n}(X_{(k_n)} - \hat{\theta})F'(\theta) + \sqrt{n}(\hat{\theta} - \theta)F'(\theta).$$

By (i),

$$\lim_n \sqrt{n}(X_{(k_n)} - \hat{\theta}_p)F'(\theta) = c \quad \text{a.s.}$$

By Theorem 5.10 in Shao (2003),

$$\sqrt{n}(\hat{\theta} - \theta)F'(\theta) \to_d N(0, p(1-p)).$$

Then, by Slusky's Theorem,

$$\sqrt{n}(X_{(k_n)} - \theta)F'(\theta) \to_d N(c, p(1-p)).$$

(iii) By (ii),

$$\sqrt{n}(X_{(k_{1n})} - \theta)F'(\theta) \to_d N\left(-z_{\alpha/2}\sqrt{p(1-p)}, p(1-p)\right)$$

and

$$\sqrt{n}(X_{(k_{2n})} - \theta)F'(\theta) \to_d N\left(z_{\alpha/2}\sqrt{p(1-p)}, p(1-p)\right).$$

Let Φ be the cumulative distribution function of $N(0,1)$. Then

$$P(X_{(k_{1n})} > \theta) = P(\sqrt{n}(X_{(k_{1n})} - \theta)F'(\theta) > 0) \to 1 - \Phi(z_{\alpha/2}) = \alpha/2$$

and

$$P(X_{(k_{2n})} < \theta) = P(\sqrt{n}(X_{(k_{2n})} - \theta)F'(\theta) < 0) \to \Phi(-z_{\alpha/2}) = \alpha/2.$$

The result follows from

$$P(\theta \in C(X)) = 1 - P(X_{(k_{1n})} > \theta) - P(X_{(k_{2n})} < \theta).$$

(iv) Let $p = \frac{1}{2}$. If F' exists and is positive at θ, the median of F, then the asymptotic variance of the sample median is $\{4n[F'(\theta)]^2\}^{-1}$. The length of the interval $C(X)$ is $X_{(k_{2n})} - X_{(k_{1n})}$. By the result in (i), this length is equal to

$$\frac{z_{\alpha/2}}{\sqrt{n}F'(\theta)} + o\left(\frac{1}{\sqrt{n}}\right) \quad \text{a.s.,}$$

i.e.,

$$\frac{[X_{(k_{2n})} - X_{(k_{1n})}]^2}{4z_{\alpha/2}^2} = \frac{1}{4n[F'(\theta)]^2} + o\left(\frac{1}{n}\right) \quad \text{a.s.}$$

Therefore, $[X_{(k_{2n})} - X_{(k_{1n})}]^2/(4z_{\alpha/2}^2)$ is a consistent estimator of the asymptotic variance of the sample median. ∎

Exercise 45 (#7.102). Let $C_{t,\alpha}(X)$ be a confidence interval for θ_t with confidence coefficient $1-\alpha$, $t = 1, ..., k$. Suppose that $C_{1,\alpha}(X), ..., C_{k,\alpha}(X)$ are independent for any α. Show how to construct simultaneous confidence intervals for θ_t, $t = 1, ..., k$, with confidence coefficient $1-\alpha$.
Solution. Let $a_k = 1-(1-\alpha)^{1/k}$. By the independence of $C_{t,a_k}(X)$, $t = 1, ..., k$,

$$\begin{aligned} P\left(\theta_t \in C_{t,a_k}(X), \quad t = 1, ..., k\right) &= \prod_{t=1}^{k} P\left(\theta_t \in C_{t,a_k}(X)\right) \\ &= \prod_{t=1}^{k} (1-\alpha)^{1/k} \\ &= 1-\alpha. \end{aligned}$$

Hence, $C_{t,a_k}(X)$, $t = 1, ..., k$, are simultaneous confidence intervals for θ_t, $t = 1, ..., k$, with confidence coefficient $1-\alpha$. ∎

Exercise 46 (#7.105). Let $x \in \mathcal{R}^k$ and A be a $k \times k$ positive definite matrix. Show that

$$x^\tau A^{-1} x = \max_{y \in \mathcal{R}^k, y \neq 0} \frac{(y^\tau x)^2}{y^\tau A y}.$$

Solution. If $x = 0$, then the equality holds. Assume that $x \neq 0$. By the Cauchy-Schwarz inequality,

$$(y^\tau x)^2 = (y^\tau A^{1/2} A^{-1/2} x)^2 \leq (y^\tau A y)(x^\tau A^{-1} x).$$

Hence,

$$x^\tau A^{-1} x \geq \max_{y \in \mathcal{R}^k, y \neq 0} \frac{(y^\tau x)^2}{y^\tau A y}.$$

Let $y_* = A^{-1}x$. Then

$$\frac{(y_*^\tau x)^2}{y_*^\tau A y_*} = \frac{(x^\tau A^{-1} x)^2}{x^\tau A^{-1} A A^{-1} x} = x^\tau A^{-1} x.$$

Hence,

$$x^\tau A^{-1}x \le \max_{y\in\mathcal{R}^k, y\neq 0} \frac{(y^\tau x)^2}{y^\tau A y}$$

and, thus, the equality holds. ∎

Exercise 47 (#7.106). Let $x \in \mathcal{R}^k$ and A be a $k\times k$ positive definite matrix.
(i) Suppose that $y^\tau A^{-1}x = 0$, where $y \in \mathcal{R}^k$. Show that

$$x^\tau A^{-1}x = \max_{c\in\mathcal{R}^k, c\neq 0, c^\tau y=0} \frac{(c^\tau x)^2}{c^\tau A c}.$$

(ii) Let X be a vector of n observations having distribution $N_n(Z\beta, \sigma^2 I_n)$, where Z is a known $n\times p$ matrix of rank $p < n$, β is an unknown p-vector, and $\sigma^2 > 0$ is unknown. Using the result in (i), construct simultaneous confidence intervals (with confidence coefficient $1-\alpha$) for $c^\tau\beta$, $c \in \mathcal{R}^p$, $c \neq 0$, $c^\tau y = 0$, where $y \in \mathcal{R}^p$ satisfies $Z^\tau Z y = 0$.
Solution. (i) If $x = 0$, then the equality holds. Assume that $x \neq 0$. Let D be the $k \times (k-1)$ matrix which spans the linear subspace

$$\mathcal{Y} = \{c : c \in \mathcal{R}^k, c^\tau y = 0\}.$$

For any $c \in \mathcal{Y}$ ($c \neq 0$), $c = Dt$ for some $t \in \mathcal{R}^{k-1}$, $t \neq 0$. Since $y^\tau A^{-1}x = 0$, $A^{-1}x \in \mathcal{Y}$ and, hence, $A^{-1}x = Dl$ or $x = ADl$ for some $l \in \mathcal{R}^{k-1}$, $l \neq 0$. Then

$$\begin{aligned}
\max_{c\in\mathcal{R}^k, c\neq 0, c^\tau y=0} \frac{(c^\tau x)^2}{c^\tau A c} &= \max_{c\in\mathcal{Y}, c\neq 0} \frac{(c^\tau x)^2}{c^\tau A c} \\
&= \max_{t\in\mathcal{R}^{k-1}, t\neq 0} \frac{(t^\tau D^\tau A D l)^2}{t^\tau D^\tau A D t} \\
&= (D^\tau A D l)^\tau (D^\tau A D)^{-1} (D^\tau A D l) \\
&= l^\tau (D^\tau A D) l \\
&= x^\tau A^{-1} x,
\end{aligned}$$

where the third equality follows from the previous exercise.
(ii) Let $\hat\beta$ be the LSE of β and $\hat\sigma^2 = \|X - Z\hat\beta\|^2/(n-p)$. Note that $(\hat\beta-\beta)^\tau(Z^\tau Z)(\hat\beta-\beta)/(p\hat\sigma^2)$ has the F-distribution $F_{p,n-p}$. Since $Z^\tau Z y = 0$, by the result in (i),

$$\max_{c\in\mathcal{R}^p, c\neq 0, c^\tau y=0} \frac{[c^\tau(\hat\beta-\beta)]^2}{\hat\sigma^2 c^\tau (Z^\tau Z)^{-1} c} = \frac{(\hat\beta-\beta)^\tau(Z^\tau Z)(\hat\beta-\beta)}{\hat\sigma^2}.$$

Thus, the $1-\alpha$ simultaneous confidence intervals for $c^\tau\beta$, $c \in \mathcal{R}^p$, $c \neq 0$, $c^\tau y = 0$, are

$$\mathcal{I}_c = \left[c^\tau\hat\beta - \hat\sigma\sqrt{pF_{p,n-p,\alpha} c^\tau (Z^\tau Z)^{-1} c},\ c^\tau\hat\beta + \hat\sigma\sqrt{pF_{p,n-p,\alpha} c^\tau (Z^\tau Z)^{-1} c} \right],$$

for $c \in \mathcal{R}^p$, $c \neq 0$, $c^\tau y = 0$, where $F_{p,n-p,\alpha}$ is the $(1-\alpha)$th quantile of the F-distribution $F_{p,n-p}$. This is because

$$\begin{aligned} & P\Big(c^\tau \beta \in \mathcal{I}_c, \quad c \in \mathcal{R}^p, c \neq 0, c^\tau y = 0\Big) \\ &= P\left(\max_{c \in \mathcal{R}^p, c\neq 0, c^\tau y = 0} \frac{[c^\tau(\hat{\beta} - \beta)]^2}{p\hat{\sigma}^2 c^\tau (Z^\tau Z)^{-1} c} \leq F_{p,n-p,\alpha} \right) \\ &= P\left(\frac{(\hat{\beta}-\beta)^\tau (Z^\tau Z)(\hat{\beta}-\beta)}{p\hat{\sigma}^2} \leq F_{p,n-p,\alpha} \right) \\ &= 1 - \alpha. \quad \blacksquare \end{aligned}$$

Exercise 48 (#7.111). Let $(X_1, ..., X_n)$ be independently distributed as $N(\beta_0 + \beta_1 z_i, \sigma^2)$, $i = 1, ..., n$, where β_0, β_1, and σ^2 are unknown and z_i's are known constants satisfying $S_z = \sum_{i=1}^n (z_i - \bar{z})^2 > 0$, $\bar{z} = n^{-1} \sum_{i=1}^n z_i$. Show that

$$\mathcal{I}_z = \left[\hat{\beta}_0 + \hat{\beta}_1 z - \hat{\sigma}\sqrt{2F_{2,n-2,\alpha} D(z)},\ \hat{\beta}_0 + \hat{\beta}_1 z + \hat{\sigma}\sqrt{2F_{2,n-2,\alpha} D(z)} \right], \quad z \in \mathcal{R},$$

are simultaneous confidence intervals for $\beta_0 + \beta_1 z$, $z \in \mathcal{R}$, with confidence coefficient $1-\alpha$, where $(\hat{\beta}_0, \hat{\beta}_1)$ is the LSE of (β_0, β_1), $D(z) = (z-\bar{z})^2/S_z + n^{-1}$, and $\hat{\sigma}^2 = (n-2)^{-1} \sum_{i=1}^n (X_i - \hat{\beta}_0 - \hat{\beta}_1 z_i)^2$.

Solution. Let $\beta = (\beta_0, \beta_1)$ and $\hat{\beta} = (\hat{\beta}_0, \hat{\beta}_1)$. Scheffé's $1-\alpha$ simultaneous confidence intervals for $t^\tau \beta$, $t \in \mathcal{R}^2$ are

$$\left[t^\tau \hat{\beta} - \hat{\sigma}\sqrt{2F_{2,n-2,\alpha} t^\tau A t},\ t^\tau \hat{\beta} + \hat{\sigma}\sqrt{2F_{2,n-2,\alpha} t^\tau A t} \right], \quad t \in \mathcal{R}^2$$

(e.g., Theorem 7.10 in Shao, 2003), where

$$A = \begin{pmatrix} n & n\bar{z} \\ n\bar{z} & \sum_{i=1}^n z_i^2 \end{pmatrix}^{-1} = \frac{1}{nS_z} \begin{pmatrix} \sum_{i=1}^n z_i^2 & -n\bar{z} \\ -n\bar{z} & n \end{pmatrix}.$$

From the solution to Exercise 46, $(t^\tau \hat{\beta} - t^\tau \beta)^2 / t^\tau A t$ is maximized at $t_* = A^{-1}(\hat{\beta} - \beta)$. Note that t_*/c still maximizes $(t^\tau \hat{\beta} - t^\tau \beta)^2 / t^\tau A t$ as long as $c \neq 0$, where c is the first component of t_*. Since the first component of $A^{-1}(\hat{\beta} - \beta)$ is $n(\hat{\beta}_0 - \beta_0) + n(\hat{\beta}_1 - \beta_1)\bar{z}$ and

$$P\Big(n(\hat{\beta}_0 - \beta_0) + n(\hat{\beta}_1 - \beta_1)\bar{z} \neq 0\Big) = 1,$$

the quantity $(t^\tau \hat{\beta} - t^\tau \beta)^2 / t^\tau A t$ is maximized at

$$\frac{A^{-1}}{n(\hat{\beta}_0 - \beta_0) + n(\hat{\beta}_1 - \beta_1)\bar{z}} \begin{pmatrix} \hat{\beta}_0 - \beta_0 \\ \hat{\beta}_1 - \beta_1 \end{pmatrix}$$

whose first component is 1. Note that $t^\tau At = D(z)$ when $t = (1, z)$. Therefore,

$$\max_{z\in\mathcal{R}} \frac{(\hat\beta_0 + \hat\beta_1 z - \beta_0 - \beta_1 z)^2}{D(z)} = \max_{t\in\mathcal{R}^2, t\neq 0} \frac{(t^\tau\hat\beta - t^\tau\beta)^2}{t^\tau At}.$$

Consequently,

$$\begin{aligned} P(\beta_0+\beta_1 z \in \mathcal{I}_z,\ z\in\mathcal{R}) &= P\left(\max_{z\in\mathcal{R}} \frac{(\hat\beta_0 + \hat\beta_1 z - \beta_0 - \beta_1 z)^2}{s\hat\sigma^2 D(z)} \leq F_{2,n-2,\alpha}\right) \\ &= P\left(\max_{t\in\mathcal{R}, t\neq 0} \frac{(t^\tau\hat\beta - t^\tau\beta)^2}{2\hat\sigma^2 t^\tau At} \leq F_{2,n-2,\alpha}\right) \\ &= 1-\alpha, \end{aligned}$$

where the last equality follows from

$$\max_{t\in\mathcal{R}, t\neq 0} \frac{(t^\tau\hat\beta - t^\tau\beta)^2}{2\hat\sigma^2 t^\tau At} = \frac{(\hat\beta-\beta)^\tau A^{-1}(\hat\beta-\beta)}{2\hat\sigma^2}$$

by Exercise 46 and the fact that the right hand side of the previous equation has the F-distribution $F_{2,n-2}$. ∎

Exercise 49 (#7.117). Let X_{0j} ($j = 1, ..., n_0$) and X_{ij} ($i = 1, ..., m$, $j = 1, ..., n_0$) represent independent measurements on a standard and m competing new treatments. Suppose that X_{ij} is distributed as $N(\mu_i, \sigma^2)$ with unknown μ_i and $\sigma^2 > 0$, $j = 1, ..., n_0$, $i = 0, 1, ..., m$. For $i = 0, 1, ..., m$, let $\bar X_{i\cdot}$ be the sample mean based on X_{ij}, $j = 1, ..., n_0$. Define $\hat\sigma^2 = [(m+1)(n_0-1)]^{-1}\sum_{i=0}^m\sum_{j=1}^{n_0}(X_{ij} - \bar X_{i\cdot})^2$.
(i) Show that the distribution of

$$R_{st} = \max_{i=1,...,m} |(\bar X_{i\cdot} - \mu_i) - (\bar X_{0\cdot} - \mu_0)|/\hat\sigma$$

does not depend on any unknown parameter.
(ii) Show that Dunnett's intervals

$$\left[\sum_{i=0}^m c_i\bar X_{i\cdot} - q_\alpha\hat\sigma\sum_{i=1}^m |c_i|, \sum_{i=0}^m c_i\bar X_{i\cdot} + q_\alpha\hat\sigma\sum_{i=1}^m |c_i|\right]$$

for all $c_0, c_1, ..., c_m$ satisfying $\sum_{i=0}^m c_i = 0$ are simultaneous confidence intervals for $\sum_{i=0}^m c_i\mu_i$ with confidence coefficient $1-\alpha$, where q_α is the $(1-\alpha)$th quantile of R_{st}.
Solution. (i) The distributions of $\hat\sigma/\sigma$ and $(\bar X_{i\cdot} - \mu_i)/\sigma$, $i = 0, 1, ..., m$, do not depend on any unknown parameter. The result follows from the fact that these random variables are independent so that their joint distribution does not depend on any unknown parameter and R_{st} is a function of these

random variables.
(ii) Let $Y_i = (\bar{X}_{i\cdot} - \mu_i)/\hat{\sigma}$. Then $R_{st} = \max_{i=1,...,m} |Y_i - Y_0|$. Note that

$$
\begin{aligned}
&P\left(\sum_{i=0}^{m} c_i\mu_i \text{ are in Dunnett's intervals for all } c_i\text{'s with } \sum_{i=0}^{m} c_i = 0\right) \\
&\quad = P\left(\left|\sum_{i=0}^{m} c_i(\bar{X}_{i\cdot} - \mu_i)\right| \leq q_\alpha\hat{\sigma}\sum_{i=1}^{m}|c_i|, \text{ all } c_i\text{'s with } \sum_{i=0}^{m} c_i = 0\right) \\
&\quad = P\left(\left|\sum_{i=0}^{m} c_iY_i\right| \leq q_\alpha\sum_{i=1}^{m}|c_i|, \text{ all } c_i\text{'s with } \sum_{i=0}^{m} c_i = 0\right).
\end{aligned}
$$

Hence, the result follows if we can show that

$$\max_{i=1,...,m} |Y_i - Y_0| \leq q_\alpha$$

is equivalent to

$$\left|\sum_{i=0}^{m} c_iY_i\right| \leq q_\alpha\sum_{i=1}^{m}|c_i| \quad \text{for all } c_0, c_1, ..., c_m \text{ satisfying } \sum_{i=0}^{m} c_i = 0.$$

Suppose that $\left|\sum_{i=0}^{m} c_iY_i\right| \leq q_\alpha\sum_{i=1}^{m}|c_i|$ for all $c_0, c_1, ..., c_m$ satisfying $\sum_{i=0}^{m} c_i = 0$. For any fixed i, let $c_0 = 1$, $c_i = -1$, and $c_j = 0$, $j \neq i$. Then these c_i's satisfy

$$\sum_{i=0}^{m} c_i = 0, \quad \sum_{i=1}^{m}|c_i| = 1, \quad \text{and} \quad \left|\sum_{i=0}^{n} c_iY_i\right| = |Y_i - Y_0|.$$

Hence, $|Y_i - Y_0| \leq q_\alpha$ for $i = 1, ..., m$. Thus, $\max_{i=1,...,m} |Y_i - Y_0| \leq q_\alpha$.

Assume now that $\max_{i=1,...,m} |Y_i - Y_0| \leq q_\alpha$. For all $c_0, c_1, ..., c_m$ satisfying $\sum_{i=0}^{m} c_i = 0$,

$$
\begin{aligned}
\left|\sum_{i=0}^{m} c_iY_i\right| &= \left|\sum_{i=1}^{m} c_iY_i + c_0Y_0\right| \\
&= \left|\sum_{i=1}^{m} c_iY_i - \sum_{i=1}^{m} c_iY_0\right| \\
&= \left|\sum_{i=1}^{m} c_i(Y_i - Y_0)\right| \\
&\leq \sum_{i=1}^{m} |c_i||Y_i - Y_0| \\
&\leq q_\alpha\sum_{i=1}^{m}|c_i|. \quad \blacksquare
\end{aligned}
$$

Exercise 50 (#7.118). Let $(X_1, ..., X_n)$ be a random sample from the uniform distribution on $(0, \theta)$, where $\theta > 0$ is unknown. Construct simultaneous confidence intervals for $F_\theta(t)$, $t > 0$, with confidence coefficient $1-\alpha$, where $F_\theta(t)$ is the cumulative distribution function of X_1.

Solution. The cumulative distribution function of X_1 is

$$F_\theta(t) = \begin{cases} 0 & t \leq 0 \\ \frac{t}{\theta} & 0 < t < \theta \\ 1 & t \geq \theta, \end{cases}$$

which is nonincreasing in θ for any fixed t. Note that $F_{\theta_1}(t) \geq F_{\theta_2}(t)$ for all $t > 0$ if and only if $\theta_1 \leq \theta_2$. From Exercise 26(iii), a $1-\alpha$ confidence interval for θ is $[X_{(n)}, c_n X_{(n)}]$, where $X_{(n)}$ is the largest order statistic and $c_n = \alpha^{-1/n}$. Hence,

$$\begin{aligned} P\left(F_{c_n X_{(n)}}(t) \leq F_\theta(t) \leq F_{X_{(n)}}(t),\ t > 0\right) &= P\left(X_{(n)} \leq \theta \leq c_n X_{(n)}\right) \\ &= 1 - \alpha, \end{aligned}$$

i.e.,

$$\left[F_{c_n X_{(n)}}(t),\ F_{X_{(n)}}(t)\right], \quad t > 0,$$

are simultaneous confidence intervals for $F_\theta(t)$, $t > 0$, with confidence coefficient $1 - \alpha$. ∎

References

In addition to the main text, *Mathematical Statistic* (Shao, 2003), some other textbooks are listed here for further readings.

Barndorff-Nielsen, O. E. (1978). *Information and Exponential Families in Statistical Theory.* Wiley, New York.

Barndorff-Nielsen, O. E. and Cox, D. R. (1994). *Inference and Asymptotics.* Chapman & Hall, London.

Berger, J. O. (1985). *Statistical Decision Theory and Bayesian Analysis*, second edition. Springer-Verlag, New York.

Bickel, P. J. and Doksum, K. A. (2002). *Mathematical Statistics*, second edition. Holden Day, San Francisco.

Casella, G. and Berger, R. L. (1990). *Statistical Inference.* Wadsworth, Belmont, CA.

Chung, K. L. (1974). *A Course in Probability Theory*, second edition. Academic Press, New York.

Ferguson, T. S. (1967). *Mathematical Statistics.* Academic Press, New York.

Lehmann, E. L. (1986). *Testing Statistical Hypotheses*, second edition. Springer-Verlag, New York.

Lehmann, E. L. and Casella, G. (1998). *Theory of Point Estimation*, second edition. Springer-Verlag, New York.

Rao, C. R. (1973). *Linear Statistical Inference and Its Applications*, second edition. Wiley, New York.

Rohatgi, V. K. (1976). *An Introduction to Probability Theory and Mathematical Statistics.* Wiley, New York.

Royden, H. L. (1968). *Real Analysis*, second edition. Macmillan, New York.

Searle, S. R. (1971). *Linear Models*. Wiley, New York.

Serfling, R. J. (1980). *Approximation Theorems of Mathematical Statistics*. Wiley, New York.

Shao, J. (2003). *Mathematical Statistics*, second edition. Springer-Verlag, New York.

Index