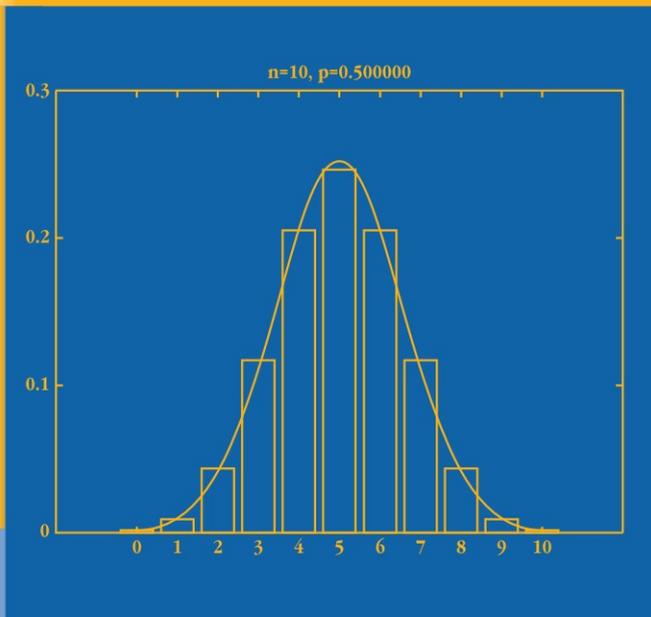


F. Biagini
M. Campanino

ELEMENTI DI PROBABILITÀ E STATISTICA



 Springer

*A Thilo, F.B.
A mio fratello Vittorio, M.C.*

*Nous ne possédons une ligne, un surface,
un volume, que si notre amour l'occupe.*
M. Proust

F. Biagini, M. Campanino

Elementi di Probabilità e Statistica

FRANCESCA BIAGINI
Dipartimento di Matematica
Università di Bologna, Bologna

MASSIMO CAMPANINO
Dipartimento di Matematica
Università di Bologna, Bologna

Springer-Verlag fa parte di Springer Science+Business Media

springer.it

© Springer-Verlag Italia, Milano 2006

ISBN 10 88-470-0330-X

ISBN 13 978-88-470-0330-9

Quest'opera è protetta dalla legge sul diritto d'autore. Tutti i diritti, in particolare quelli relativi alla traduzione, alla ristampa, all'uso di figure e tabelle, alla citazione orale, alla trasmissione radiofonica o televisiva, alla riproduzione su microfilm o in database, alla diversa riproduzione in qualsiasi altra forma (stampa o elettronica) rimangono riservati anche nel caso di utilizzo parziale. Una riproduzione di quest'opera, oppure di parte di questa, è anche nel caso specifico solo ammessa nei limiti stabiliti dalla legge sul diritto d'autore, ed è soggetta all'autorizzazione dell'Editore. La violazione delle norme comporta sanzioni previste dalla legge.

L'utilizzo di denominazioni generiche, nomi commerciali, marchi registrati, ecc., in quest'opera, anche in assenza di particolare indicazione, non consente di considerare tali denominazioni o marchi liberamente utilizzabili da chiunque ai sensi della legge sul marchio.

Riprodotta da copia camera-ready fornita dagli Autori
Progetto grafico della copertina: Simona Colombo, Milano
Stampato in Italia: Signum, Bollate (Mi)

Fondamenti del Calcolo delle Probabilità

I numeri aleatori

1.1 Introduzione

Il Calcolo delle Probabilità si occupa di quantificare il nostro grado di incertezza. Il suo oggetto fondamentale di studio sono gli enti aleatori e, in particolare, i numeri aleatori. Che cosa si intende per *numero aleatorio*? Si tratta di un numero ben definito, ma non necessariamente conosciuto. Ad esempio il risultato di un determinato esperimento, una quotazione azionaria ad un istante prefissato, il valore di una grandezza meteorologica ad un istante fissato. Tutte queste quantità hanno un valore ben definito, ma possono non essere conosciute o perché si riferiscono al futuro e non si hanno i mezzi per poterle prevedere con certezza o anche se si riferiscono al passato non fanno parte delle informazioni conosciute. Indicheremo i numeri aleatori con le lettere maiuscole. Anche se il valore di un numero aleatorio numero aleatorio non è in generale conosciuto, si potrà sapere con certezza *l'insieme dei suoi valori possibili*, che sarà denotato con $I(X)$. I numeri certi si possono considerare come casi particolari di numeri aleatori il cui insieme dei valori possibili contiene un solo elemento.

Esempio 1.1.1. Siano X, Y due numeri aleatori rappresentanti i risultati del lancio di una moneta e di un dado. Indicando croce con 0 e testa con 1, si ottiene

$$\begin{aligned}I(X) &= \{0, 1\}, \\I(Y) &= \{1, 2, 3, 4, 5, 6\}.\end{aligned}$$

Un numero aleatorio X si dice:

- *superiormente limitato* se l'insieme dei valori possibili $I(X)$ è superiormente limitato ($\sup I(X) < +\infty$);
- *inferiormente limitato* se l'insieme dei valori possibili $I(X)$ è inferiormente limitato ($\inf I(X) > -\infty$);

- *limitato* se l'insieme dei valori possibili $I(X)$ è sia superiormente che inferiormente limitato ($\sup I(X) < +\infty$, $\inf I(X) > -\infty$).

Dati X e Y numeri aleatori, si definisce $I(X, Y)$ l'insieme delle coppie possibili. In generale, si indica con $I(X_1, \dots, X_n)$ l'insieme delle n -uple possibili. X e Y si dicono *logicamente indipendenti* se

$$I(X, Y) = I(X) \times I(Y).$$

dove $I(X) \times I(Y)$ indica il prodotto cartesiano fra l'insieme dei valori possibili $I(X)$ e l'insieme dei valori possibili $I(Y)$.

Esempio 1.1.2 (di non indipendenza logica). Si considerino due estrazioni senza reimpulso di numeri da 1 a 90 (gioco del lotto), siano X_1 ed X_2 due numeri aleatori che rappresentano rispettivamente il risultato della prima e della seconda estrazione. L'insieme dei valori possibili è allora

$$I(X, Y) = \{(i, j) | 1 \leq i \leq 90, i \neq j\}.$$

Chiaramente, $I(X, Y) \neq I(X) \times I(Y)$ perché $I(X, Y)$ non contiene le coppie del tipo (i, i) , $i \in \{1, \dots, 90\}$. I due numeri aleatori X e Y non sono quindi logicamente indipendenti.

I numeri aleatori X_1, \dots, X_n si dicono *logicamente indipendenti* se

$$I(X_1, \dots, X_n) = I(X_1) \times \dots \times I(X_n).$$

Con i numeri aleatori si possono effettuare le usuali operazioni aritmetiche. Inoltre, definiamo le seguenti operazioni che saranno frequentemente utilizzate:

1. $X \vee Y := \max(X, Y)$;
2. $X \wedge Y := \min(X, Y)$;
3. $\bar{X} := 1 - X$.

Tali operazioni hanno le seguenti proprietà:

1. Proprietà distributiva

$$X \vee (Y \wedge Z) = (X \vee Y) \wedge (X \vee Z) \tag{1.1}$$

$$X \wedge (Y \vee Z) = (X \wedge Y) \vee (X \wedge Z) \tag{1.2}$$

2. Proprietà associativa

$$X \vee (Y \vee Z) = (X \vee Y) \vee Z \tag{1.3}$$

$$X \wedge (Y \wedge Z) = (X \wedge Y) \wedge Z \tag{1.4}$$

3. Proprietà commutativa

$$X \vee Y = Y \vee X \quad (1.5)$$

$$X \wedge Y = Y \wedge X \quad (1.6)$$

4. Proprietà connesse alla \sim

$$\tilde{\tilde{X}} = X \quad (1.7)$$

$$(X \vee Y)^\sim = \tilde{X} \wedge \tilde{Y} \quad (1.8)$$

$$(X \wedge Y)^\sim = \tilde{X} \vee \tilde{Y} \quad (1.9)$$

1.2 Eventi

Un caso particolare di numero aleatorio è dato dagli eventi. Un evento E è un numero aleatorio tale che $I(E) \subset \{0, 1\}$. Nel caso di eventi, dati due eventi E e F , $E \vee F$ si dice *somma logica* e $E \wedge F$ *prodotto logico*. Si verifica facilmente che:

1. $E \vee F = E + F - EF$;
2. $E \wedge F = EF$.

Dato un evento E , si definisce *complementare* di E l'evento

$$\tilde{E} = 1 - E.$$

Si ha che $\tilde{\tilde{E}} = E$.

Dalle proprietà dell'operazione di completamentazione abbiamo

$$(E \vee F)^\sim = \tilde{E} \wedge \tilde{F} = (1 - E)(1 - F) = 1 - E - F + EF,$$

da cui segue

$$E \vee F = E + F - EF.$$

Analogamente

$$\begin{aligned} (E \vee F \vee G)^\sim &= \tilde{E} \wedge \tilde{F} \wedge \tilde{G} = (1 - E)(1 - F)(1 - G) \\ &= 1 - E - F - G + EF + EG + FG - EFG, \end{aligned}$$

da cui segue

$$E \vee F \vee G = E + F + G - EF - EG - FG + EFG.$$

Altre due operazioni fra eventi sono:

differenza: $E \setminus F = E - EF$;

differenza simmetrica: $E \triangle F = (E \setminus F) \vee (F \setminus E) = E + F \pmod{2}$.

D'ora in avanti useremo il simbolo \vdash per dire che la proposizione che segue è sicuramente vera. Per esempio, $\vdash X \leq Y$ se $I(X, Y) \subset \{(x, y) \mid x \leq y\}$.

Usiamo la notazione

$$E \subset F \text{ se } \vdash E \leq F,$$

Inoltre, $\vdash E = F$ si scrive in modo equivalente come $E \equiv F$ se $E \subset F$ e $F \subset E$.

Definizione 1.2.1. *Si definiscono le seguenti proprietà:*

1. **Incompatibilità:** E, F si dicono incompatibili se $\vdash EF = 0$;
2. **Esaustività:** E_1, \dots, E_n si dicono esaustivi se $\vdash E_1 + \dots + E_n \geq 1$;
3. **Partizione:** E_1, \dots, E_n si dicono una partizione se $\vdash E_1 + \dots + E_n = 1$ (esaustivi e incompatibili).

Esempio 1.2.2. Un evento E ed il suo complementare \tilde{E} sono una partizione.

Dati E_1, \dots, E_n eventi, per costruire una partizione a partire da E_1, \dots, E_n si usa il metodo dei costituenti. Si definisce *costituente* di E_1, \dots, E_n l'evento

$$Q = E_1^* \cdots E_n^*,$$

dove E_i^* può essere uguale ad E_i o al suo complementare \tilde{E}_i .

In generale, non tutti i costituenti sono possibili. Sono possibili tutti i costituenti solo quando gli E_i sono logicamente indipendenti. I costituenti possibili costituiscono una partizione. Infatti

$$1 = (E_1 + \tilde{E}_1) \cdots (E_n + \tilde{E}_n) = \sum_{Q \text{ costituente}} Q$$

Dalla somma si possono escludere tutti i costituenti impossibili.

Se E_1, \dots, E_n sono una partizione, allora i costituenti possibili sono:

$$\begin{aligned} & E_1 \tilde{E}_2 \cdots \tilde{E}_n, \\ & \tilde{E}_1 E_2 \tilde{E}_3 \cdots \tilde{E}_n, \\ & \quad \cdots, \\ & \tilde{E}_1 \cdots \tilde{E}_{n-1} E_n, \end{aligned}$$

che si possono identificare con gli eventi stessi.

Definiamo ora quando un evento è logicamente indipendente da altri eventi. I costituenti sono classificabili nel seguente modo rispetto ad un dato evento E :

- I tipo** $Q \subset E$;
- II tipo** $Q \subset \tilde{E}$;
- III tipo** altrimenti.

L'evento E è *logicamente dipendente* da E_1, \dots, E_n se tutti i costituenti di E_1, \dots, E_n sono del primo o del secondo tipo.

E è *logicamente indipendente* da E_1, \dots, E_n se tutti i costituenti di E_1, \dots, E_n sono del terzo tipo.

Altrimenti E si dice *logicamente semidipendente*.

Se E è logicamente dipendente da E_1, \dots, E_n , si può scrivere

$$E = \sum_{\substack{Q \text{ di I tipo} \\ Q \subset E}} Q.$$

Esempio 1.2.3. Consideriamo due eventi E_1, E_2 . L'evento somma logica ($E_1 \vee E_2$) si può scrivere come

$$E_1 \vee E_2 = E_1 E_2 + \tilde{E}_1 E_2 + E_1 \tilde{E}_2.$$

In generale se un evento E è logicamente dipendente da E_1, \dots, E_n se e solo se E si può scrivere come $E = \Phi(E_1, \dots, E_n)$ per qualche funzione Φ .

Esempio 1.2.4. Supponiamo di effettuare cinque lanci di una moneta. Sia E_i l'evento che corrisponde all'esito "testa" all' i -esimo lancio. Posto $Y = E_1 + E_2 + E_3 + E_4 + E_5$, considero l'evento

$$E = (Y \geq 3).$$

E è logicamente semidipendente dai primi tre eventi. Infatti

I tipo: $E_1 E_2 E_3 \subset E$;

II tipo: $\tilde{E}_1 \tilde{E}_2 \tilde{E}_3 \subset \tilde{E}$;

III tipo: $\tilde{E}_1 E_2 \tilde{E}_3$.

1.3 La previsione

Dato un numero aleatorio X , cerchiamo un valore certo che esprima la nostra valutazione su X . In termini economici se pensiamo a X come a un guadagno aleatorio, vogliamo scegliere un guadagno certo che riteniamo equivalente a X .

Seguendo l'impostazione di de Finetti in [dF] definiamo in modo operativo la previsione $\mathbf{P}(X)$ ¹ che un individuo assegna ad un numero aleatorio X .

Esistono due modi operativi equivalenti per definire la previsione:

1. **Metodo della scommessa:** si pensa X come il guadagno (o la perdita, se negativo) derivante da una scommessa. La previsione $\mathbf{P}(X)$ è allora il guadagno certo che si giudica equivalente alla quantità aleatoria X .

Posto $\mathbf{P}(X) = \bar{x}$, si accetta una scommessa pari a

¹ Prende anche il nome di *media, attesa o speranza*, e si indica anche con $\mathbb{E}[X]$.

$$\lambda(X - \bar{x}),$$

dove $\lambda \in \mathbb{R}$ è un coefficiente di proporzionalità che può essere scelto da chi ci propone la scommessa. Il corrispondente criterio di coerenza è che non si possa scegliere \bar{x} in modo che ci sia una perdita certa. Nella finanza matematica questo prende il nome di *Principio di Non Arbitraggio*.

2. **Metodo della penalità:** si suppone di dover pagare una penalità pari a

$$-\lambda(X - \bar{x})^2,$$

dove $\lambda \in \mathbb{R}^+$ è un coefficiente di proporzionalità. Anche qui vi è un criterio di coerenza: non deve esistere un valore \bar{x}' tale che la corrispondente penalità sia sicuramente minore. Tale \bar{x} si dice previsione $\mathbf{P}(X)$ del numero aleatorio X .

Proposizione 1.3.1 (Proprietà della previsione). *Dal principio di coerenza segue che la previsione ha le seguenti proprietà:*

1. *Monotonia:* $\inf I(X) \leq \mathbf{P}(X) \leq \sup I(X)$;
2. *Linearità:* se $X = \alpha_1 X_1 + \dots + \alpha_n X_n$, allora $\mathbf{P}(X) = \alpha_1 \mathbf{P}(X_1) + \dots + \alpha_n \mathbf{P}(X_n)$.

Dimostrazione. 1. *Monotonia:* La previsione \bar{x} deve essere tale che non si possa scegliere λ in modo tale che si abbia un guadagno certo od una perdita certa. Se fosse $\bar{x} < \inf I(X)$, allora per $\lambda < 0$

$$\vdash \lambda(X - \bar{x}) < 0.$$

Se invece fosse $\bar{x} > \sup I(X)$, per $\lambda > 0$ si avrebbe

$$\vdash \lambda(X - \bar{x}) < 0.$$

Ne segue che

$$\inf I(X) \leq \bar{x} \leq \sup I(X).$$

Tale proprietà si dimostra analogamente in base al secondo criterio.

2. *Linearità:* Per la dimostrazione, si procede utilizzando il principio di Non Arbitraggio. Consideriamo il numero aleatorio $Z = X + Y$. Posto $\bar{z} = \mathbf{P}(Z)$, $\bar{x} = \mathbf{P}(X)$, $\bar{y} = \mathbf{P}(Y)$, sia G il guadagno

$$\begin{aligned} G &= c_1(X - \bar{x}) + c_2(Y - \bar{y}) + c_3(Z - \bar{z}) = \\ &= (c_1 + c_3)X + (c_2 + c_3)Y - c_1\bar{x} - c_2\bar{y} - c_3\bar{z}. \end{aligned}$$

Scegliendo c_1, c_2, c_3 in modo tale da annullare la parte aleatoria

$$c_1 = c_2 = -c_3,$$

si ottiene il guadagno complessivo: $G = c_3(\bar{x} + \bar{y} - \bar{z})$. Per evitare che si possa scegliere c_3 in modo che $\vdash G < 0$, dovrà essere $\bar{x} + \bar{y} - \bar{z} = 0$, ovvero

$\bar{z} = \bar{x} + \bar{y}$. Se si procede invece con il secondo criterio, si è sottoposti ad una penalità (guadagno negativo)

$$-[(X - \bar{x})^2 + (Y - \bar{y})^2 + (Z - \bar{z})^2] = -[(X - \bar{x})^2 + (Y - \bar{y})^2 + (X + Y - \bar{z})^2].$$

Per ogni punto P , la proiezione ortogonale P' di P sul piano $z = x + y$ ha distanza minore di P da ogni punto (X, Y, Z) possibile (che si deve trovare sul piano). In base al principio di coerenza dovrà essere $P = P'$, ovvero che P deve appartenere al piano $z = x + y$. Ne segue che $\bar{z} = \bar{x} + \bar{y}$. Analogamente, per $Z = \alpha X$, $\alpha \in \mathbb{R}$, si ottiene $\bar{z} = \alpha \bar{x}$. In generale, se $X = \alpha_1 X_1 + \dots + \alpha_n X_n$, allora

$$\mathbf{P}(X) = \alpha_1 \mathbf{P}(X_1) + \dots + \alpha_n \mathbf{P}(X_n).$$

La proprietà di monotonia si può descrivere anche nel seguente modo:

$$\begin{aligned} \vdash X \geq c &\implies \mathbf{P}(X) \geq c; \\ \text{Se } c_1 \leq c_2, \vdash c_1 \leq X \leq c_2 &\implies c_1 \leq \mathbf{P}(X) \leq c_2; \\ \vdash X = c &\implies \mathbf{P}(X) = c. \end{aligned}$$

Osservazione 1.3.2. Per i numeri aleatori illimitati per cui $\inf I(X)$, $\sup I(X)$ o entrambi non esistono finiti può non esistere nessun valore finito corrispondente alla nostra valutazione di $\mathbf{P}(X)$. Rimandiamo a [dF] per una discussione di questo argomento e di questo approccio alla definizione della previsione e della probabilità.

1.4 Probabilità di eventi

Nel caso di un evento, la previsione $\mathbf{P}(E)$ si chiama anche *probabilità* di E . Dalla proprietà di monotonia, segue che:

1. la probabilità di un evento è un numero compreso fra 0 ed 1, ovvero $0 \leq \mathbf{P}(E) \leq 1$.
2. $E \equiv 0 \implies \mathbf{P}(E) = 0$.
3. $E \equiv 1 \implies \mathbf{P}(E) = 1$.

Quando $E \equiv 1$, E si dice *evento certo*. Se $E \equiv 0$, E è un *evento impossibile*. Si ha inoltre che:

$$\begin{aligned} \text{somma logica: } \mathbf{P}(E_1 \vee E_2) &= \mathbf{P}(E_1 + E_2 - E_1 E_2) \leq \mathbf{P}(E_1 + E_2); \\ \text{somma: } \mathbf{P}(E_1 + E_2) &= \mathbf{P}(E_1) + \mathbf{P}(E_2). \end{aligned}$$

Le due previsioni coincidono *se e solo se* E_1 e E_2 sono incompatibili. Da $\vdash E_1 + E_2 \leq E_1 \vee E_2$ per la monotonia della previsione si ha che

$$\mathbf{P}(E_1 \vee E_2) \leq \mathbf{P}(E_1 + E_2).$$

Per una partizione

$$\vdash E_1 + \dots + E_n = 1 \implies \sum \mathbf{P}(E_i) = 1$$

La funzione che assegna agli eventi di una partizione le loro probabilità si dice *distribuzione di probabilità* della partizione. Se E dipende logicamente da una partizione di eventi $\{E_1, \dots, E_n\}$ possiamo trovare la probabilità di E a partire da quella degli E_i .

$$\mathbf{P}(E) = \sum_{E_i \subset E} \mathbf{P}(E_i).$$

Vediamo ora una formula per calcolare la previsione. Sia X un numero aleatorio con $I(X) = \{x_1, \dots, x_n\}$ e sia $E_i := (X = x_i)$. Vale che:

$$\mathbf{P}(X) = \sum_{i=1}^n x_i \mathbf{P}(X = x_i). \quad (1.10)$$

Infatti

$$\mathbf{P}(X) = \mathbf{P}(X(E_1 + \dots + E_n)) =$$

$$\mathbf{P}(X E_1) + \dots + \mathbf{P}(X E_n) =$$

$$\sum_{i=1}^n \mathbf{P}(X E_i) = \sum_{i=1}^n \mathbf{P}(x_i E_i) =$$

$$\sum_{i=1}^n x_i \mathbf{P}(E_i) = \sum_{i=1}^n x_i \mathbf{P}(X = x_i).$$

Basta infatti notare che $X E_i$ è un numero aleatorio che assume il valore x_i oppure 0. L'uguaglianza $\mathbf{P}(x_i E_i) = x_i \mathbf{P}(E_i)$ è una conseguenza della proprietà di linearità della previsione.

In generale, se $I(X)$ è finito e $\phi: \mathbb{R} \rightarrow \mathbb{R}$ vale che

$$\mathbf{P}(\phi(X)) = \sum_{i=1}^n \phi(x_i) \mathbf{P}(X = x_i). \quad (1.11)$$

La dimostrazione è analoga a quella per la formula (1.10). Si noti inoltre che (1.10) è un caso particolare di (1.11) quando $\phi(x) = x$.

Esempio 1.4.1. Sia X il numero rappresentante il risultato del lancio di un dado. Se ogni faccia ha la stessa probabilità di uscire, la previsione di X è data da:

$$\mathbf{P}(X) = \frac{1}{6} \sum_{i=1}^6 i = \frac{6 \cdot 7}{6 \cdot 2} = \frac{7}{2}.$$

Esempio 1.4.1. Sia X il numero rappresentante il risultato del lancio di un dado. Se ogni faccia ha la stessa probabilità di uscire, la previsione di X è data da:

$$\mathbf{P}(X) = \frac{1}{6} \sum_{i=1}^6 i = \frac{6 \cdot 7}{6 \cdot 2} = \frac{7}{2}.$$

Esempio 1.4.2. Sia X il numero aleatorio che rappresenta il risultato del lancio di una moneta simmetrica. Indicando con $I(X) = \{0, 1\}$, si ottiene che la previsione di X è data da:

$$\mathbf{P}(X) = \frac{1}{2}.$$

1.5 Partizioni in eventi equiprobabili

In alcune situazioni, per ragioni di simmetria, è naturale attribuire la stessa probabilità a tutti gli eventi di una partizione, come nel caso dei giochi d'azzardo. Se E_1, \dots, E_n sono gli eventi di una partizione con *distribuzione uniforme*, vale che

$$\mathbf{P}(E_i) = \frac{1}{n}.$$

Sia E un evento che dipende logicamente dalla partizione E_1, \dots, E_n . La previsione di E è data da:

$$\mathbf{P}(E) = \mathbf{P} \left(\sum_{E_i \subset E} E_i \right) = \frac{\#\{i | E_i \subset E\}}{n}.$$

Si ottiene dunque la nota formula

$$\mathbf{P}(E) = \frac{\#\text{ casi favorevoli}}{\#\text{ casi possibili}}. \quad (1.12)$$

Tale identità è valida unicamente nel caso in cui gli eventi della partizione sono valutati equiprobabili.

Esempio 1.5.1. Si effettuano n lanci di una moneta equilibrata. Sia X il numero aleatorio che rappresenta il numero di teste che si ottengono considerando n lanci. Sia E_i l'evento corrispondente all'uscita di una *testa* all' i -esimo lancio. Considero l'evento

$$E := (X = k) = \sum_{Q \subset E} Q,$$

dove $Q = E_1^* \dots E_n^*$ sono i costituenti degli eventi E_1, \dots, E_n ; tali costituenti determinano una partizione e sono tutti possibili in quanto gli E_i sono tutti logicamente indipendenti. La simmetria della moneta porta ad attribuire la stessa probabilità ad ogni costituente. La probabilità di E si ottiene usando

n pari: Il massimo valore di $\mathbf{P}(X = k)$ si ha per $k = \frac{n}{2}$;

n dispari: Il massimo valore di $\mathbf{P}(X = k)$ si ha per $k = \frac{n-1}{2}$ e per $k = \frac{n+1}{2}$.

Esempio 1.5.2. Si fanno n estrazioni con reimbussolamento da un'urna con H palline bianche e $N - H$ palline nere. Sia X il numero aleatorio che conta il numero di palline bianche estratte. Si calcola

$$\mathbf{P}(X = k) = \frac{\# \text{ casi favorevoli}}{\# \text{ casi possibili}},$$

dove il numero di *casi possibili* è pari a N^n ed il numero di *casi favorevoli* è pari a

$$\binom{n}{k} H^k (N - H)^{n-k}.$$

Si può pensare un costituente come una sequenza di palline bianche e nere. I casi favorevoli sono quelli in cui tale sequenza presenta una pallina bianca in k posizioni; per ciascuna di queste posizioni si può scegliere tra H palline bianche, essendo le estrazioni con reimbussolamento.

Considerando invece delle estrazioni senza reimbussolamento, possiamo porre come numero dei casi possibili

$$\binom{N}{n}.$$

Possiamo infatti non tener conto dell'ordine in quanto l'evento considerato non dipende dall'ordine di estrazione delle n palline. Il numero di casi favorevoli è dato da

$$\binom{H}{k} \binom{N - H}{n - k}.$$

Si devono infatti scegliere k palline fra le H bianche senza tener conto dell'ordine ed $n - k$ palline fra le $N - H$ nere senza tener conto dell'ordine. Alternativamente avremmo potuto tener conto dell'ordine (ovviamente sia nel conteggio dei casi favorevoli sia in quello dei casi possibili), ottenendo lo stesso risultato.

1.6 Probabilità e previsione subordinata

Si tratta della probabilità (e della previsione) subordinata (o condizionata) al verificarsi di un dato evento. Sia X un numero aleatorio ed H un evento. Per definire la previsione subordinata, si utilizzano due metodi operativi.

1. Metodo della scommessa:

La scommessa vale quando H si verifica, altrimenti è annullata e quindi il guadagno è uguale a 0. Si sceglie \bar{x} sapendo che si può essere sottoposti ad una scommessa con un guadagno:

$$G = \lambda H(X - \bar{x}),$$

dove $\lambda \in \mathbb{R}$ rappresenta un coefficiente di proporzionalità che può essere scelto da chi ci propone la scommessa. Il numero reale \bar{x} si dice *previsione subordinata* di X rispetto ad H e si denota con $\mathbf{P}(X|H)$.

2. *Metodo della penalità:*

Anche qui la penalità viene inflitta se H si verifica. Si sceglie \bar{x} sapendo di dover pagare una penalità

$$p = -H(X - \bar{x})^2.$$

Il numero reale \bar{x} è la *previsione subordinata* di X rispetto ad H e si denota con $\mathbf{P}(X|H)$.

Nel caso in cui si considera la previsione subordinata di un evento E rispetto ad H , si parla di probabilità subordinata di E dato H . Considero l'insieme dei valori possibili $I(X|H)$ di X dato H . Si ha che $I(X|H) \subset I(X)$. Si ha che la previsione subordinata ha le stesse proprietà della previsione, ovvero:

- $\inf I(X|H) \leq \mathbf{P}(X|H) \leq \sup I(X|H)$,
- $\mathbf{P}(X + Y|H) = \mathbf{P}(X|H) + \mathbf{P}(Y|H)$,
- $\mathbf{P}(\lambda X|H) = \lambda \mathbf{P}(X|H)$,

che seguono nello stesso modo dal principio di coerenza.

1.7 Formula delle probabilità composte

Vale la *formula delle probabilità composte*

$$\mathbf{P}(XH) = \mathbf{P}(H)\mathbf{P}(X|H).$$

Per dimostrarla, si pongano $z = P(XH)$, $x = P(H)$ e $y = P(X|H)$. Utilizzando il metodo della scommessa, si ottiene:

$$\begin{aligned} G &= c_1(H - x) + c_2H(X - y) + c_3(XH - z) \\ &= H(c_1 + (c_2 + c_3)X - c_2y) - c_1x - c_3z. \end{aligned}$$

Ponendo $c_2 = -c_3$ e $c_1 = c_2y$ si ottiene

$$G = -c_1x - c_3z = c_2(z - xy).$$

Avendo annullato la parte aleatoria, per non poter esser sottoposti a perdite certe, si dovrà avere:

$$z = xy$$

In modo analogo, è possibile dimostrare la formula con il metodo delle penalità. Se $\mathbf{P}(H) > 0$, vale

$$\mathbf{P}(X|H) = \frac{\mathbf{P}(XH)}{\mathbf{P}(H)}.$$

Se X è un evento, $X = E$, allora

$$\mathbf{P}(E|H) = \frac{\mathbf{P}(EH)}{\mathbf{P}(H)}.$$

Come casi particolari si ha:

1. $E \subset H \Rightarrow \mathbf{P}(E|H) = \frac{\mathbf{P}(E)}{\mathbf{P}(H)}$;
2. $H \subset E$, ovvero $I(E|H) = \{1\} \Rightarrow \mathbf{P}(E|H) = 1$;
3. $H \subset \tilde{E}$, ovvero $I(E|H) = \{0\} \Rightarrow \mathbf{P}(E|H) = 0$.

1.8 Formula delle probabilità totali

Sia H_1, \dots, H_n una partizione e X un numero aleatorio. Vale che:

$$\mathbf{P}(X) = \sum_{i=1}^n \mathbf{P}(X|H_i)\mathbf{P}(H_i)$$

Infatti,

$$\begin{aligned} \mathbf{P}(X) &= \mathbf{P}(X \cdot 1) = \mathbf{P}(X(H_1 + \dots + H_n)) = \\ \mathbf{P}(XH_1 + XH_2 + \dots + XH_n) &= \sum_{i=1}^n \mathbf{P}(XH_i) = \sum_{i=1}^n \mathbf{P}(X|H_i)\mathbf{P}(H_i). \end{aligned}$$

1.9 Formula di Bayes

Siano E, H due eventi tali che $\mathbf{P}(H) > 0$. Vale la *Formula di Bayes*

$$\mathbf{P}(E|H) = \frac{\mathbf{P}(H|E)\mathbf{P}(E)}{\mathbf{P}(H)}.$$

Dalla formula della probabilità condizionata si ha che $\mathbf{P}(EH) = \mathbf{P}(H|E)\mathbf{P}(E)$. Quindi:

$$\mathbf{P}(E|H) = \frac{\mathbf{P}(EH)}{\mathbf{P}(H)} = \frac{\mathbf{P}(H|E)\mathbf{P}(E)}{\mathbf{P}(H)}.$$

Esempio 1.9.1. Consideriamo un'urna di composizione ignota contenente N palline bianche e nere. Sia Y il numero aleatorio di palline bianche nell'urna. Gli eventi $H_i = (Y = i)$ determinano una partizione. Sia E l'evento corrispondente all'estrazione di una pallina bianca. Si vuole calcolare la probabilità di

E e la probabilità che nell'urna vi siano i palline bianche se si è estratta una pallina bianca, ovvero se si è verificato l'evento E . Si usa la formula delle probabilità totali per calcolare la probabilità di E nel modo seguente:

$$\mathbf{P}(E) = \sum_{i=0}^N \mathbf{P}(E|H_i)\mathbf{P}(H_i) = \sum_{i=0}^N \frac{i}{N}\mathbf{P}(H_i).$$

Se è nota la composizione dell'urna, la probabilità di E è data dal numero di palline bianche, cioè i casi favorevoli, diviso il numero totale delle palline, i casi possibili.

Supponiamo ora che non si conosca la composizione dell'urna. Se si assume che gli H_i siano equiprobabili, cioè che sia $\mathbf{P}(H_i) = \frac{1}{N+1}$, si ottiene:

$$\mathbf{P}(E) = \sum_{i=0}^N \frac{i}{N(N+1)} = \frac{1}{2}.$$

Dalla formula di Bayes segue che

$$\mathbf{P}(H_i|E) = \frac{\mathbf{P}(E|H_i)\mathbf{P}(H_i)}{\mathbf{P}(E)} = \frac{\frac{i}{N}\frac{1}{N+1}}{\frac{1}{2}} = \frac{2i}{N(N+1)}.$$

1.10 Correlazione tra eventi

Un evento E si dice *correlato positivamente* con H se

$$\mathbf{P}(E|H) > \mathbf{P}(E).$$

Analogamente, un evento E si dice *correlato negativamente* con H se

$$\mathbf{P}(E|H) < \mathbf{P}(E).$$

Se $\mathbf{P}(E|H) = \mathbf{P}(E)$, si dice che E non è correlato con H ; in tal caso si dice anche che E ed H sono *stocasticamente indipendenti*.

In questo caso, l'informazione che H si è verificato non cambia la valutazione delle probabilità di E e viceversa. Se invece E è correlato positivamente con H , l'informazione che H si è verificato aumenta la valutazione della probabilità di E .

Se $\mathbf{P}(H) > 0$ e $\mathbf{P}(E) > 0$, si può dare una definizione simmetrica della correlazione. E e H si dicono

- *correlati positivamente* se $\mathbf{P}(EH) > \mathbf{P}(E)\mathbf{P}(H)$,
- *correlati negativamente* se $\mathbf{P}(EH) < \mathbf{P}(E)\mathbf{P}(H)$,
- *non correlati* se $\mathbf{P}(EH) = \mathbf{P}(E)\mathbf{P}(H)$.

Se E è correlata positivamente con H , si ha che \tilde{E} è correlato negativamente con H

$$\mathbf{P}(\tilde{E}|H) = 1 - \mathbf{P}(E|H) < 1 - \mathbf{P}(E) = \mathbf{P}(\tilde{E}).$$

Se E non è correlato con H , nemmeno \tilde{E} lo è.

Esempio 1.10.1. Consideriamo un'urna con H palline bianche e $N - H$ palline nere; si effettuano due estrazioni. Si denotano con E_1, E_2 gli eventi che venga estratta una pallina bianca rispettivamente alla prima ed alla seconda estrazione. Nel caso di estrazioni con reimbussolamento, si ottiene

$$\mathbf{P}(E_1) = \mathbf{P}(E_2) = \frac{H}{N}.$$

Infatti la composizione dell'urna è la stessa sia alla prima che alla seconda estrazione. Si verifica subito che le due estrazioni sono indipendenti (come ci si aspettava!) in quanto

$$\mathbf{P}(E_1 E_2) = \frac{H^2}{N^2} = \mathbf{P}(E_1)\mathbf{P}(E_2).$$

Se invece si effettuano le estrazioni senza reimbussolamento si ottiene $\mathbf{P}(E_1) = \frac{H}{N}$ e

$$\begin{aligned} \mathbf{P}(E_2) &= \mathbf{P}(E_2|E_1)\mathbf{P}(E_1) + \mathbf{P}(E_2|\tilde{E}_1)\mathbf{P}(\tilde{E}_1) \\ &= \frac{H-1}{N-1} \frac{H}{N} + \frac{H}{N-1} \left(1 - \frac{H}{N}\right) = \frac{H}{N}. \end{aligned}$$

Le probabilità delle due estrazioni sono dunque le stesse, ma i due eventi risultano correlati negativamente in quanto

$$\mathbf{P}(E_2|E_1) = \frac{H-1}{N-1} < \frac{H}{N} = \mathbf{P}(E_2)$$

per ogni $H < N$.

È possibile estendere la definizione di indipendenza anche al caso di un numero n , generico, di eventi.

Definizione 1.10.2. E_1, \dots, E_n si dicono stocasticamente indipendenti se per ogni scelta finita di indici $\{i_1, \dots, i_k\}$ in $\{1, \dots, n\}$ si ha che

$$\mathbf{P}(E_{i_1} \cdots E_{i_k}) = \mathbf{P}(E_{i_1}) \cdots \mathbf{P}(E_{i_k}). \quad (1.13)$$

Non basta verificare la (1.13) solamente per le coppie!

Vedremo che se E_1, \dots, E_n sono stocasticamente indipendenti, anche gli eventi E_1^*, \dots, E_n^* sono stocasticamente indipendenti per ogni scelta possibile di E_i^* fra E_i ed \tilde{E}_i .

Definizione 1.10.3. Sia $\mathcal{H} = \{H_1, \dots, H_n\}$ una partizione; due eventi E_1, E_2 si dicono stocasticamente indipendenti subordinatamente alla partizione \mathcal{H} se

$$\forall i = 1, \dots, n \quad \mathbf{P}(E_1 E_2 | H_i) = \mathbf{P}(E_1 | H_i) \mathbf{P}(E_2 | H_i).$$

Esempio 1.10.4. Consideriamo un'urna con composizione incognita contenente N palline bianche e nere. Sia Y il numero aleatorio che rappresenta il numero di palline nell'urna. Si effettuano due estrazioni con reimbussolamento. Sia E_1 l'evento "esce una pallina bianca alla prima estrazione" e sia E_2 l'evento "esce una pallina bianca alla seconda estrazione".

Consideriamo la partizione \mathcal{H} determinata dagli eventi

$$H_i = (Y = i) \quad i = 0, \dots, N.$$

Si assume che $\mathbf{P}(H_i) = \frac{1}{N+1}$. Gli eventi E_1 ed E_2 sono stocasticamente indipendenti data \mathcal{H} , ovvero

$$\mathbf{P}(E_1 E_2 | H_i) = \mathbf{P}(E_1 | H_i) \mathbf{P}(E_2 | H_i).$$

per ogni $i = 0, \dots, N$. Ci si chiede se essi siano anche stocasticamente indipendenti. Si calcola

1. La probabilità della prima estrazione

$$\begin{aligned} \mathbf{P}(E_1) &= \sum_{i=0}^N \mathbf{P}(E_1 | H_i) \mathbf{P}(H_i) \\ &= \frac{1}{N+1} \sum_{i=0}^N \frac{i}{N} \\ &= \frac{1}{N+1} \frac{N(N+1)}{2} \\ &= \frac{1}{2}. \end{aligned}$$

2. La probabilità della seconda estrazione è pari a quella della prima come si è già notato. Quindi $\mathbf{P}(E_2) = \mathbf{P}(E_1) = \frac{1}{2}$.

3. Per la probabilità di estrarre due palline bianche si ha

$$\begin{aligned} \mathbf{P}(E_1 E_2) &= \sum_{i=0}^N \mathbf{P}(E_1 E_2 | H_i) \mathbf{P}(H_i) \\ &= \frac{1}{N+1} \sum_{i=0}^N \mathbf{P}(E_1 | H_i) \mathbf{P}(E_2 | H_i) \\ &= \frac{1}{N+1} \sum_{i=0}^N \frac{i^2}{N^2}. \end{aligned}$$

Per calcolare $\sum_{i=0}^N k^2$ si utilizza il fatto che

$$(n+1)^3 - n^3 = 3n^2 + 3n + 1.$$

Ne segue che $\sum_{i=0}^N i^2 = \frac{(N+1)^3}{3} - \frac{N(N+1)}{2} - \frac{(N+1)}{3}$, da cui

$$\mathbf{P}(E_1 E_2) = \frac{2N+1}{6N}.$$

Tale probabilità tende ad $\frac{1}{3}$ per N che tende all'infinito. Quindi $\mathbf{P}(E_1 E_2) \neq \mathbf{P}(E_1)\mathbf{P}(E_2)$ ovvero l'indipendenza stocastica rispetto ad una partizione non implica l'indipendenza stocastica.

1.11 L'indipendenza stocastica attraverso i costituenti

Proposizione 1.11.1. E_1, \dots, E_n sono stocasticamente indipendenti se e solo se per ogni costituente $Q = E_1^* \cdots E_n^*$ di E_1, \dots, E_n , dove

$$E_i^* = \begin{cases} E_i \\ \tilde{E}_i \end{cases}$$

vale che

$$\mathbf{P}(Q) = \mathbf{P}(E_1^*) \cdots \mathbf{P}(E_n^*). \quad (1.14)$$

Dimostrazione. \Rightarrow) Sia $Q = E_1^* \cdots E_n^*$ un costituente di E_1, \dots, E_n . Se si sviluppano i prodotti fra gli eventi, si ottiene che Q è dato da un polinomio ϕ in n variabili di grado 1 in ogni variabile calcolato in E_1, \dots, E_n , ovvero

$$E_1^* \cdots E_n^* = \phi(E_1, \dots, E_n).$$

Per esempio, se si considerano tre eventi E_1, E_2, E_3 , il costituente

$$Q = \tilde{E}_1 E_2 E_3 = (1 - E_1) E_2 E_3 = E_2 E_3 - E_1 E_2 E_3$$

è dato dal polinomio in tre variabili $\phi(x_1, x_2, x_3) = x_2 x_3 - x_1 x_2 x_3$ calcolato in E_1, E_2, E_3 .

Se gli E_i sono stocasticamente indipendenti, le probabilità dei prodotti si fattorizzano e si ottiene

$$\begin{aligned} \mathbf{P}(Q) &= \mathbf{P}(\phi(E_1, \dots, E_n)) \\ &= \phi(\mathbf{P}(E_1), \dots, \mathbf{P}(E_n)) \\ &= \mathbf{P}(E_1^*) \cdots \mathbf{P}(E_n^*). \end{aligned}$$

Ritornando all'esempio, la probabilità di $Q = \tilde{E}_1 E_2 E_3$ è quindi

$$\begin{aligned}
\mathbf{P}(Q) &= \mathbf{P}(\tilde{E}_1 E_2 E_3) \\
&= \mathbf{P}(E_2 E_3 - E_1 E_2 E_3) \\
&= \mathbf{P}(E_2)\mathbf{P}(E_3) - \mathbf{P}(E_1)\mathbf{P}(E_2)\mathbf{P}(E_3) = \phi(\mathbf{P}(E_1), \mathbf{P}(E_2), \mathbf{P}(E_3)).
\end{aligned}$$

⇐ Viceversa, si supponga che valga (1.14). Dati E_1, \dots, E_n essi sono stocasticamente indipendenti se e solo se per ogni scelta di indici i_1, \dots, i_k in $\{1, \dots, n\}$ vale che

$$\mathbf{P}(E_{i_1} \cdots E_{i_k}) = \mathbf{P}(E_{i_1}) \cdots \mathbf{P}(E_{i_k})$$

Usando i costituenti, si ha che:

$$\begin{aligned}
\mathbf{P}(E_{i_1} \cdots E_{i_k}) &= \mathbf{P}\left(\sum_{Q \subset E_{i_1} \cdots E_{i_k}} Q\right) \\
&= \mathbf{P}(E_{i_1}) \cdots \mathbf{P}(E_{i_k}) \cdot \sum \left(\mathbf{P}(E_{i_{k+1}}^*) \cdots \mathbf{P}(E_{i_N}^*)\right).
\end{aligned}$$

La sommatoria nell'ultimo termine deve essere considerata su tutte le possibilità in cui si possono presentare gli altri $(N-k)$ eventi. Tale sommatoria vale quindi 1, da cui la tesi.

1.12 Covarianza e varianza

Dati due numeri aleatori X e Y , si definisce *covarianza* di X e Y

$$\mathbf{cov}(X, Y) = \mathbf{P}((X - \mathbf{P}(X))(Y - \mathbf{P}(Y))).$$

X e Y si dicono

- *correlati positivamente* se $\mathbf{cov}(X, Y) > 0$,
- *correlati negativamente* se $\mathbf{cov}(X, Y) < 0$,
- *non correlati* se $\mathbf{cov}(X, Y) = 0$.

Sviluppando la formula precedente, si ottiene

$$\mathbf{cov}(X, Y) = \mathbf{P}(XY - \mathbf{P}(X)Y - X\mathbf{P}(Y) + \mathbf{P}(X)\mathbf{P}(Y)) = \mathbf{P}(XY) - \mathbf{P}(X)\mathbf{P}(Y).$$

La *varianza* è definita come

$$\sigma^2(X) = \mathbf{cov}(X, X).$$

Si ottiene che $\sigma^2(X) = \mathbf{P}(X^2) - \mathbf{P}(X)^2$ ovvero $\sigma^2(X) = \mathbf{P}((X - \mathbf{P}(X))^2)$. Se $\sigma^2(X)$ è 0, allora tutta la probabilità è concentrata nella previsione $\mathbf{P}(X)$. In senso probabilistico si può dire che X è equivalente alla costante $\mathbf{P}(X)$. Data la varianza, si introducono inoltre

- *Scarto quadratico medio:*

$$\sigma(X) = \sqrt{\sigma^2(X)}.$$

- *Previsione quadratica:*

$$P_Q(X) = \sqrt{\mathbf{P}(X^2)}.$$

Proposizione 1.12.1 (Proprietà della covarianza e varianza). *La covarianza e la varianza rispettano le seguenti proprietà:*

1. *Bilinearità:*

$$\mathbf{cov}(X + Y, Z) = \mathbf{cov}(X, Z) + \mathbf{cov}(Y, Z). \quad (1.15)$$

2. *Proprietà rispetto ad una trasformazione lineare:*

$$\mathbf{cov}(aX + b, cY + d) = ac \mathbf{cov}(X, Y), \quad (1.16)$$

$$\sigma^2(aX + b) = a^2 \sigma^2(X). \quad (1.17)$$

Dimostrazione. 1. Basta utilizzare il fatto che $\mathbf{cov}(X, Z) = \mathbf{P}(XZ) - \mathbf{P}(X)\mathbf{P}(Z)$ e la linearità della previsione. Si ottiene

$$\begin{aligned} \mathbf{cov}(X + Y, Z) &= \mathbf{P}[(X + Y)Z] - \mathbf{P}(X + Y)\mathbf{P}(Z) \\ &= \mathbf{P}(XZ + YZ) - [\mathbf{P}(X) + \mathbf{P}(Y)]\mathbf{P}(Z) \\ &= \mathbf{P}(XZ) - \mathbf{P}(X)\mathbf{P}(Z) + \mathbf{P}(YZ) - \mathbf{P}(Y)\mathbf{P}(Z) \\ &= \mathbf{cov}(X, Z) + \mathbf{cov}(Y, Z). \end{aligned}$$

2. Basta utilizzare la definizione di covarianza

$$\begin{aligned} \mathbf{cov}(aX + b, cY + d) &= \mathbf{P}((aX + b - \mathbf{P}(aX + b))(cY + d - \mathbf{P}(cY + d))) \\ &= \mathbf{P}((aX + b - a\mathbf{P}(X) - b)(cY + d - c\mathbf{P}(Y) - d)) \\ &= \mathbf{P}(a(X - \mathbf{P}(X))c(Y - \mathbf{P}(Y))) \\ &= ac \mathbf{cov}(X, Y). \end{aligned}$$

La proprietà (1.17) segue immediatamente dalla (1.16) sostituendo $(cY + d)$ con $(aX + b)$.

Proposizione 1.12.2 (La varianza nella somma di numeri aleatori).

Siano X_1, \dots, X_n n numeri aleatori. Si ha che

$$\begin{aligned} \sigma^2(X_1 + \dots + X_n) &= \sum_{i=1}^n \sigma^2(X_i) + \sum_{\substack{i,j \\ i \neq j}} \mathbf{cov}(X_i, X_j) \\ &= \sum_{i=1}^n \sigma^2(X_i) + 2 \sum_{i < j} \mathbf{cov}(X_i, X_j). \end{aligned}$$

Dimostrazione. Basta usare la definizione di varianza:

$$\begin{aligned}\sigma^2(X_1 + \dots + X_n) &= \mathbf{P} \left(\left(\sum_{i=1}^n X_i - \sum_{i=1}^n \mathbf{P}(X_i) \right)^2 \right) = \\ &= \mathbf{P} \left[((X_1 - \mathbf{P}(X_1)) + \dots + (X_n - \mathbf{P}(X_n)))^2 \right] = \\ &= \sum_{i=1}^n \underbrace{\mathbf{P} \left((X_i - \mathbf{P}(X_i))^2 \right)}_{\sigma^2(X_i)} + \sum_{\substack{i,j \\ i \neq j}} \underbrace{\mathbf{P} \left((X_i - \mathbf{P}(X_i))(X_j - \mathbf{P}(X_j)) \right)}_{\text{cov}(X_i, X_j)} = \\ &= \sum_{i=1}^n \sigma^2(X_i) + \sum_{\substack{i,j \\ i \neq j}} \text{cov}(X_i, X_j).\end{aligned}$$

1.13 Il coefficiente di correlazione

Dati due numeri aleatori X, Y , si definisce *coefficiente di correlazione* di X, Y

$$\rho(X, Y) = \frac{\text{cov}(X, Y)}{\sigma(X) \sigma(Y)}.$$

Le proprietà del coefficiente di correlazione sono:

1. $\rho(aX + b, cY + d) = \text{segno}(ac) \rho(X, Y)$.

Utilizzando le proprietà della covarianza si ha che:

$$\begin{aligned}\rho(aX + b, cY + d) &= \frac{\text{cov}(aX + b, cY + d)}{\sqrt{\sigma^2(aX + b) \sigma^2(cY + d)}} \\ &= \frac{ac \text{cov}(X, Y)}{|ac| \sqrt{\sigma^2(X) \sigma^2(Y)}} \\ &= \text{segno}(ac) \rho(X, Y).\end{aligned}$$

2. $-1 \leq \rho(X, Y) \leq 1$.

Si considerano i numeri aleatori standardizzati

$$X^* = \frac{X - \mathbf{P}(X)}{\sigma(X)}, \quad Y^* = \frac{Y - \mathbf{P}(Y)}{\sigma(Y)},$$

ovvero tali che $\mathbf{P}(X^*) = \mathbf{P}(Y^*) = 0$ e $\sigma^2(X^*) = \sigma^2(Y^*) = 1$. Dalla Proposizione 1.12.1, si ha che

$$\text{cov}(X^*, Y^*) = \mathbf{P}(X^* Y^*) = \frac{\mathbf{P}((X - \mathbf{P}(X))(Y - \mathbf{P}(Y)))}{\sigma(X) \sigma(Y)} = \rho(X, Y).$$

Calcolando la varianza di $X^* + Y^*$ si ottiene

$$\begin{aligned}\sigma^2(X^* + Y^*) &= \sigma^2(X^*) + \sigma^2(Y^*) + 2 \operatorname{cov}(X^*, Y^*) = \\ &= 2 + 2 \operatorname{cov}(X^*, Y^*) = 2 + 2\rho(X, Y) \geq 0,\end{aligned}$$

in quanto la varianza di un numero aleatorio è una quantità sempre positiva. Analogamente dalla varianza di $X^* - Y^*$ segue che

$$\sigma^2(X^* - Y^*) = 2 - 2\rho(X, Y) \geq 0.$$

Quindi vale

$$-1 \leq \rho(X, Y) \leq 1.$$

1.14 La disuguaglianza di Chebychev

Valgono le seguenti disuguaglianze, dette di Chebychev:

1. Sia X numero aleatorio tale che $P_Q(X) > 0$. Per ogni $t > 0$ vale che

$$\mathbf{P}(|X| \geq tP_Q(X)) \leq \frac{1}{t^2}.$$

2. Sia X numero aleatorio con $\sigma^2(X) > 0$. Posto $m = \mathbf{P}(X)$, per $\forall t > 0$ si ha che:

$$\mathbf{P}(|X - m| \geq \sigma(X)t) \leq \frac{1}{t^2}.$$

Dimostrazione. 1. Sia E l'evento $E = (|X| \geq tP_Q(X))$. Calcoliamo $\mathbf{P}(X^2)$ con la formula delle probabilità totali:

$$\mathbf{P}(X^2) = \mathbf{P}(X^2|E) \mathbf{P}(E) + \mathbf{P}(X^2|\tilde{E}) \mathbf{P}(\tilde{E}).$$

Per la proprietà di monotonia della previsione, si ha che $\mathbf{P}(X^2|\tilde{E}) \geq 0$ in quanto X^2 è un numero aleatorio sempre positivo. Ne segue che

$$\mathbf{P}(X^2) \geq \mathbf{P}(X^2|E) \mathbf{P}(E) \geq t^2 P_Q(X)^2 \mathbf{P}(E).$$

Poiché $P_Q(X)^2 = \mathbf{P}(X^2)$ si ottiene

$$\mathbf{P}(E) \leq \frac{1}{t^2}$$

ovvero

$$\mathbf{P}(|X| \geq tP_Q(X)) \leq \frac{1}{t^2}.$$

2. La seconda disuguaglianza segue dalla prima sostituendo ad X il numero aleatorio $Y = X - m$.

1.15 La legge debole dei grandi numeri

Theorem 1.15.1 (Legge dei grandi numeri). Sia $(X_n)_{n \in \mathbb{N}}$ una successione di numeri aleatori a due a due non correlati con stessa previsione $\mathbf{P}(X_i) = m$ e varianza $\sigma^2(X_i) = \sigma^2$.

Posto $S_n = X_1 + \dots + X_n$, si ha che

$$\forall \lambda > 0 \quad \lim_{n \rightarrow +\infty} \mathbf{P} \left(\left| \frac{S_n}{n} - m \right| \geq \lambda \right) = 0.$$

Il numero aleatorio $\frac{S_n}{n}$ si dice *media campionaria*.

Dimostrazione. Si dimostra il teorema utilizzando la seconda disuguaglianza di Chebychev. Si calcola la previsione di $\frac{S_n}{n}$

$$\mathbf{P} \left(\frac{S_n}{n} \right) = \frac{1}{n} (\mathbf{P}(X_1) + \dots + \mathbf{P}(X_n)) = m$$

e la varianza di $\frac{S_n}{n}$

$$\sigma^2 \left(\frac{S_n}{n} \right) = \frac{1}{n^2} \sigma^2(S_n) = \frac{1}{n^2} \left(\sum_{i=1}^n \sigma^2(X_i) + \sum_{\substack{i,j=1 \\ i \neq j}}^n \text{cov}(X_i, X_j) \right) = \frac{\sigma^2}{n}.$$

Dalla seconda disuguaglianza di Chebychev

$$\mathbf{P} \left(\left| \frac{S_n}{n} - m \right| \geq \frac{\sigma}{\sqrt{n}} t \right) \leq \frac{1}{t^2}.$$

Posto $\lambda = \frac{\sigma}{\sqrt{n}} t$, si ricava $\frac{1}{t^2} = \frac{\sigma^2}{n\lambda^2}$. Ne segue che

$$\mathbf{P} \left(\left| \frac{S_n}{n} - m \right| \geq \lambda \right) \leq \frac{\sigma^2}{n\lambda^2}$$

tende a 0 se $n \rightarrow +\infty$.

Esempio 1.15.2. $X_i = E_i$ eventi non correlati con $\mathbf{P}(E_i) = p$. Dalla legge dei grandi numeri segue che

$$\mathbf{P} \left(\left| \frac{E_1 + \dots + E_n}{n} - p \right| \geq \lambda \right) \xrightarrow{n \rightarrow +\infty} 0$$

In questo caso $\frac{S_n}{n} = \frac{E_1 + \dots + E_n}{n}$ prende il nome di *frequenza*. Per un numero grande di prove, la frequenza approssima la probabilità di un evento.

Distribuzioni discrete

2.1 Numeri aleatori con distribuzione discreta

Un numero aleatorio X si dice con *distribuzione discreta* se la cardinalità dell'insieme dei valori possibili $I(X)$ è finita o numerabile; la *distribuzione di probabilità* di X è data da

$$\mathbf{P}(X = x_i) = p(x_i) \quad x_i \in I(X),$$

dove $\sum_{x_i \in I(X)} \mathbf{P}(X = x_i) = 1$. Vediamo ora alcune delle distribuzioni discrete più importanti.

2.2 Schema di Bernoulli

Sia $(E_i)_{i \in \mathbb{N}}$ una successione di eventi stocasticamente indipendenti ed equiprobabili, ovvero tali che $\mathbf{P}(E_i) = p \forall i \in \mathbb{N}$, con $0 < p < 1$. Indipendenti vuol dire che, per ogni n , E_1, \dots, E_n sono stocasticamente indipendenti. Tale successione prende il nome di *schema di Bernoulli*.

Esempio 2.2.1. Un esempio di schema di Bernoulli è dato dalla successione di numeri aleatori che rappresentano il risultato dei lanci ripetuti di una moneta simmetrica.

2.3 Distribuzione binomiale

Dato $(E_i)_{i \in \mathbb{N}}$ uno schema di Bernoulli con $\mathbf{P}(E_i) = p$, sia S_n il numero aleatorio che conta i successi ottenuti su n prove. S_n si può scrivere come

$$S_n = E_1 + \dots + E_n.$$

L'insieme dei valori possibili per S_n è quindi $I(S_n) = \{0, \dots, n\}$.

Calcoliamo, attraverso i costituenti, la distribuzione di probabilità di S_n .

$$\mathbf{P}(S_n = k) = \sum_{Q \subset (S_n = k)} \mathbf{P}(Q).$$

Bisogna dunque calcolare la probabilità di un costituente del primo tipo dell'evento $(S_n = k)$. Un esempio ne è

$$Q = E_1 \cdots E_k \tilde{E}_{k+1} \cdots \tilde{E}_n,$$

che rappresenta l'evento in cui i k successi si sono ottenuti con le prime k prove, mentre le restanti corrispondono ad insuccessi.

Analogamente, ogni altro costituente di $(S_n = k)$ conterrà k eventi che si sono verificati ed $(n - k)$ che non si sono verificati. Poichè gli E_i sono *iid*¹, si ottiene che ogni costituente Q ha la stessa probabilità, pari a

$$\mathbf{P}(Q) = \underbrace{p \cdots p}_{k \text{ volte}} \underbrace{(1 - p) \cdots (1 - p)}_{(n - k) \text{ volte}} = p^k (1 - p)^{n - k}.$$

Basta quindi contare quanti sono tali costituenti: essi sono $\binom{n}{k}$, pari al numero di modi di scegliere i k posti degli eventi che si verificano nella sequenza degli n eventi che compongono il costituente stesso. Si ottiene quindi

$$\mathbf{P}(S_n = k) = \binom{n}{k} p^k (1 - p)^{n - k}.$$

Si dice che S_n ha *distribuzione binomiale* $\mathcal{B}(n, p)$ di parametri n, p .

Si verifica che $\sum_{k=0}^n \mathbf{P}(S_n = k) = 1$. Infatti, utilizzando le proprietà del binomio di Newton, si ottiene:

$$\sum_{k=0}^n \binom{n}{k} p^k (1 - p)^{n - k} = (p + 1 - p)^n = 1.$$

Calcoliamo infine la previsione di X sapendo che $X = E_1 + \cdots + E_n$:

$$\mathbf{P}(X) = \mathbf{P}(E_1 + \cdots + E_n) = \sum_{i=1}^n \mathbf{P}(E_i) = np.$$

Esempio 2.3.1. Consideriamo un'urna contenente N palline, di cui H bianche ed $N - H$ nere. Si fanno delle estrazioni con reimpulso. La successione $(E_i)_{i \in \mathbb{N}}$ di eventi $E_i =$ (si ottiene una pallina bianca all' i -esima estrazione) è uno schema di Bernoulli, mentre il numero aleatorio che conta il numero di palline bianche ottenute nelle prime n estrazioni ha distribuzione binomiale di parametri $n, \frac{H}{N}$. Si veda l'esempio 1.5.2.

¹ Si indica con *iid* la proprietà di essere indipendenti e identicamente distribuiti.

2.4 Distribuzione geometrica

Sia E_i uno schema di Bernoulli; sia T il numero aleatorio che rappresenta istante del primo successo in una serie di prove, ovvero $T = \inf\{n \mid E_n = 1\}$. L'insieme dei valori possibili per il numero aleatorio T è dato da:

$$I(T) = \mathbb{N} \setminus \{0\}.$$

L'evento $(T = i)$ si può scrivere in termini degli E_i come

$$(T = i) = \tilde{E}_1 \cdots \tilde{E}_{i-1} E_i.$$

Calcoliamo la distribuzione di probabilità:

$$\mathbf{P}(T = i) = \mathbf{P}(\tilde{E}_1 \cdots \tilde{E}_{i-1} E_i) = \mathbf{P}(\tilde{E}_1) \cdots \mathbf{P}(\tilde{E}_{i-1}) \mathbf{P}(E_i) = (1-p)^{i-1} p.$$

Si dice che T ha *distribuzione geometrica* di parametro p . Utilizzando la somma della serie geometrica, si verifica che:

$$\sum_{i=1}^{+\infty} \mathbf{P}(T = i) = \sum_{i=1}^{+\infty} (1-p)^{i-1} p = p \sum_{k=0}^{+\infty} (1-p)^k = p \cdot \frac{1}{1-(1-p)} = 1.$$

Da questo segue inoltre facilmente che $\mathbf{P}(T = \infty) = 0$. Si calcola la previsione di T utilizzando la formula che estende al caso con un'infinità numerabile di valori la formula che abbiamo ottenuto per un numero finito di valori possibili. Questa estensione può essere giustificata come conseguenza di un'ipotesi di regolarità.

$$\mathbf{P}(T) = \sum_{i=1}^{+\infty} i \mathbf{P}(T = i) = \sum_{i=1}^{+\infty} i (1-p)^{i-1} p = p \sum_{i=1}^{+\infty} i (1-p)^{i-1} = p \frac{1}{p^2} = \frac{1}{p},$$

dove si è utilizzando il fatto che

$$\sum_{i=1}^{+\infty} i x^{i-1} = \sum_{i=1}^{+\infty} \frac{d}{dx} [x^i] = \frac{d}{dx} \left(\sum_{i=0}^{+\infty} x^i \right) = \frac{d}{dx} \left(\frac{1}{1-x} \right) = \frac{1}{(1-x)^2}.$$

La distribuzione geometrica gode della proprietà di “*assenza di memoria*”. Vale infatti che

$$\mathbf{P}(T > m+n \mid T > n) = \mathbf{P}(T > m)$$

per ogni $m, n \in \mathbb{N}$. La proprietà di *assenza di memoria* ci dice che la probabilità di non avere un successo fino all'istante $m+n$ se non si era ancora ottenuto un successo fino all'istante n , è pari alla probabilità di non avere un successo fino all'istante m . Per dimostrare tale proprietà, basta osservare che

$$\mathbf{P}(T > m+n \mid T > n) = \frac{\mathbf{P}(T > m+n, T > n)}{\mathbf{P}(T > n)} = \frac{\mathbf{P}(T > m+n)}{\mathbf{P}(T > n)}$$

e che $\mathbf{P}(T > n) = (1-p)^n$ in quanto l'evento $(T > n)$ si verifica se e solo se i primi n eventi non si verificano. Ne segue allora che

$$\mathbf{P}(T > m+n \mid T > n) = \frac{\mathbf{P}(T > m+n)}{\mathbf{P}(T > n)} = \frac{(1-p)^{m+n}}{(1-p)^n} = (1-p)^m = \mathbf{P}(T > m).$$

2.5 Distribuzione di Poisson

Un numero aleatorio X si dice avere *distribuzione di Poisson* di parametro λ , $\lambda \in \mathbb{R}_+$, se $I(X) = \mathbb{N}$ e vale che

$$\mathbf{P}(X = i) = \frac{\lambda^i}{i!} e^{-\lambda}.$$

Si verifica che si tratta di una distribuzione di probabilità propria, ovvero che $\sum_{i=0}^{+\infty} \mathbf{P}(X = i) = 1$. Si ha che

$$\sum_{i=0}^{+\infty} \mathbf{P}(X = i) = \sum_{i=0}^{+\infty} \frac{\lambda^i}{i!} e^{-\lambda} = e^{-\lambda} \sum_{i=0}^{+\infty} \frac{\lambda^i}{i!} = e^{-\lambda} e^{\lambda} = 1.$$

Calcoliamo la previsione di X sotto l'ipotesi di regolarità di cui abbiamo parlato e che d'ora in poi supporremo sempre verificata.

$$\begin{aligned} \mathbf{P}(X) &= \sum_{i=0}^{+\infty} i \mathbf{P}(X = i) = \sum_{i=0}^{+\infty} i \frac{\lambda^i}{i!} e^{-\lambda} = \lambda e^{-\lambda} \sum_{i=1}^{+\infty} \frac{\lambda^{i-1}}{(i-1)!} \\ &= \lambda e^{-\lambda} \sum_{k=0}^{+\infty} \frac{\lambda^k}{k!} = \lambda e^{-\lambda} e^{\lambda} = \lambda. \end{aligned}$$

2.6 La distribuzione ipergeometrica

Si consideri un'urna contenente N palline di cui H bianche ed $N - H$ nere. Si fanno n estrazioni senza reimbussolamento. Sia X il numero aleatorio che conta il numero di palline bianche nel campione.

Il minimo numero di palline bianche fra le n estratte sarà pari a 0 se il numero delle palline nere nell'urna $N - H$ è maggiore o uguale ad n , ed $n - (N - H)$ altrimenti. Viceversa, il numero massimo di palline bianche nel campione è dato dal minimo fra n ed il numero delle palline bianche nell'urna, ovvero H . Si ottiene che

$$I(X) = \{0 \vee n - (N - H), \dots, n \wedge H\}.$$

Sia $i \in I(X)$. Si vuole calcolare la distribuzione di probabilità di X utilizzando la formula

$$\mathbf{P}(X = i) = \frac{\# \text{ casi favorevoli}}{\# \text{ casi possibili}}.$$

Nella definizione dei casi possibili si può non tenere conto dell'ordine, dato che l'evento considerato non dipende dall'ordine. Il numero di casi possibili coincide con il numero di modi di scegliere n palline fra le N presenti nell'urna senza ripetizione e senza tener conto dell'ordine, ovvero

$$\# \text{ casi possibili} = \binom{N}{n}.$$

Per avere i palline bianche nel campione, bisogna prendere i palline bianche fra le H contenute nell'urna e scegliere le restanti $(n - i)$ fra le $(N - H)$ nere. Ne segue che

$$\# \text{ casi favorevoli} = \binom{H}{i} \binom{N - H}{n - i}.$$

Si dice che X possiede *distribuzione ipergeometrica* e si ha quindi

$$\mathbf{P}(X = i) = \frac{\binom{H}{i} \binom{N - H}{n - i}}{\binom{N}{n}}.$$

In questo caso nella definizione dei casi possibili bisogna tener conto dell'ordine. Consideriamo l'evento $E_i =$ (esce una pallina bianca alla i -esima estrazione). La probabilità di ottenere una pallina bianca alla i -esima estrazione è data da

$$\mathbf{P}(E_i) = \frac{\# \text{ casi favorevoli}}{\# \text{ casi possibili}} = \frac{HD_{n-1}^{N-1}}{D_n^N} = \frac{H}{N}.$$

Infatti, se si considerano le n palline estratte come ordinate in una n -upla, il numero di casi favorevoli è dato dalle n -uple ordinate che hanno una pallina bianca all' i -esimo posto, mentre il numero dei casi possibili sono tutte le n -uple ordinate di n elementi scelti su N .

Poiché $X = E_1 + \dots + E_n$, usando la linearità della previsione si ottiene

$$\mathbf{P}(X) = \sum_{i=1}^n \mathbf{P}(E_i) = n \frac{H}{N}.$$

2.7 Indipendenza di partizioni

Si considerino due partizioni:

$$\mathcal{H} = (H_1, \dots, H_m), \quad \mathcal{L} = (L_1, \dots, L_n).$$

\mathcal{H} e \mathcal{L} si dicono *stocasticamente indipendenti* se per ogni i, j tali che $1 \leq i \leq m$, $1 \leq j \leq n$ vale

$$\mathbf{P}(H_i L_j) = \mathbf{P}(H_i) \mathbf{P}(L_j).$$

Date r partizioni $\mathcal{H}_1, \dots, \mathcal{H}_r$, ciascuna formata da n_i ($i = 1, \dots, r$) eventi, esse si dicono *stocasticamente indipendenti* se per ogni scelta di indici i_1, \dots, i_r tali che $1 \leq i_1 \leq n_1, \dots, 1 \leq i_r \leq n_r$ vale

$$\mathbf{P}\left(H_{i_1}^{(1)} \dots H_{i_r}^{(r)}\right) = \mathbf{P}\left(H_{i_1}^{(1)}\right) \dots \mathbf{P}\left(H_{i_r}^{(r)}\right),$$

dove $H_{i_k}^{(k)} \in \mathcal{H}_k$, $k = 1, \dots, r$. Si può pensare ad una partizione come ad un *plurievento*, corrispondente ad un esperimento che può avere un certo numero di risultati.

2.8 Schema di Bernoulli generalizzato

Siano $\mathcal{H}_1, \dots, \mathcal{H}_n$ partizioni contenenti lo stesso numero di eventi, $\mathcal{H}_i = \{E_1^{(i)}, \dots, E_r^{(i)}\}$ ($i = 1, \dots, n$), tali che per ogni i valgano

1. $\forall j = 1, \dots, r \quad \mathbf{P}(E_j^{(i)}) = p_j$,
2. $p_1 + \dots + p_r = 1$.

Si suppone che $\mathcal{H}_1, \dots, \mathcal{H}_n$ siano stocasticamente indipendenti. Si parla in questo caso di *schema di Bernoulli generalizzato*. Si può rappresentare lo schema di Bernoulli generalizzato aiutandoci in modo grafico:

$$\begin{array}{c} E_1^{(1)}, \dots, E_r^{(1)} \\ E_1^{(2)}, \dots, E_r^{(2)} \\ \vdots, \dots, \vdots \\ \vdots, \dots, \vdots \\ E_1^{(n)}, \dots, E_r^{(n)}. \end{array}$$

Gli eventi appartenenti alla stessa colonna hanno la stessa probabilità e quelli sulla stessa riga appartengono ad una stessa partizione, quindi le loro probabilità sommano ad uno.

La definizione si estende ad una successione infinita di partizioni $(\mathcal{H}_i)_{i \in \mathbb{N}}$ richiedendo che $\mathcal{H}_1, \dots, \mathcal{H}_m$ soddisfino le condizioni predette per ogni m .

2.9 La distribuzione multinomiale

Dato uno schema di Bernoulli generalizzato introdotto nella sezione precedente si può definire la *distribuzione multinomiale*. Consideriamo Y_1, \dots, Y_r numeri aleatori definiti come

$$Y_i = \sum_{k=1}^n E_i^{(k)}.$$

Nella rappresentazione grafica precedente, si vede che gli Y_i si ottengono sommando gli eventi sulle colonne. Si ha

$$\sum_{i=1}^r Y_i = \sum_{i=1}^r \sum_{k=1}^n E_i^{(k)} = \sum_{k=1}^n \underbrace{\sum_{i=1}^r E_i^{(k)}}_1 = n.$$

Il concetto di costituente si può estendere in modo naturale dal caso di eventi al caso di partizioni. Date n partizioni $\mathcal{H}_1, \dots, \mathcal{H}_n$ si dirà un loro costituente un evento

$$Q = \prod_{j=1}^n H_j^*,$$

dove H_j^* è un evento della partizione \mathcal{H}_j . Se le partizioni sono stocasticamente indipendenti (come nel caso dello schema di Bernoulli generalizzato) si ha che

$$\mathbf{P}(Q) = \mathbf{P}(H_1^*) \cdots \mathbf{P}(H_n^*).$$

Vogliamo calcolare

$$\mathbf{P}(Y_1 = k_1, \dots, Y_r = k_r) = \sum_{Q \text{ I tipo}} \mathbf{P}(Q),$$

dove Q varia tra i costituenti di $\mathcal{H}_1, \dots, \mathcal{H}_n$. In un costituente di I tipo nel prodotto dovranno apparire k_i volte eventi di tipo i con $1 \leq i \leq r$. Quindi dato che le partizioni sono stocasticamente indipendenti si ha che per ogni costituente Q di I tipo

$$\mathbf{P}(Q) = p_1^{k_1} \cdots p_r^{k_r}.$$

Il numero dei costituenti di I tipo è pari al numero di modi di suddividere n elementi in r sottogruppi di k_1, \dots, k_r elementi ciascuno, ovvero $\frac{n!}{k_1! \cdots k_r!}$. Si ha dunque

$$\mathbf{P}(Y_1 = k_1, \dots, Y_r = k_r) = \sum_{Q \text{ I tipo}} \mathbf{P}(Q) = \underbrace{\frac{n!}{k_1! \cdots k_r!}}_{\text{numero di costituenti}} \underbrace{p_1^{k_1} \cdots p_r^{k_r}}_{\mathbf{P}(Q)}.$$

La distribuzione multinomiale dipende quindi dal parametro r e dalle probabilità p_1, \dots, p_{r-1} (p_r è determinabile conoscendo le altre $(r-1)$ probabilità). Per $r = 2$ è equivalente alla distribuzione binomiale.

2.10 Indipendenza stocastica per numeri aleatori con distribuzione discreta

Siano X e Y due numeri aleatori con $I(X) = \{x_1, \dots, x_m\}$ e $I(Y) = \{y_1, \dots, y_n\}$. Si considerino le partizioni \mathcal{H} generata dagli eventi $H_i = (X = x_i)$, $x_i \in I(X)$, e \mathcal{L} generata dagli eventi $L_j = (Y = y_j)$, $y_j \in I(Y)$.

I numeri aleatori X e Y si dicono *stocasticamente indipendenti* se lo sono le partizioni \mathcal{H} e \mathcal{L} .

2.11 Distribuzione congiunta

Consideriamo il *vettore aleatorio* (X, Y) con insieme dei valori possibili $I(X, Y)$. Si definisce *distribuzione congiunta* di (X, Y) la funzione $p(x_i, y_j)$

$$p(x_i, y_j) = \mathbf{P}(X = x_i, Y = y_j)$$

dove $(x_i, y_j) \in I(X, Y)$. Si può associare alla distribuzione congiunta la matrice

$$\begin{pmatrix} p(x_1, y_1) & \cdots & p(x_1, y_n) \\ \vdots & \ddots & \vdots \\ p(x_m, y_1) & \cdots & p(x_m, y_n) \end{pmatrix}.$$

Si definisce *distribuzione marginale* di X

$$p_1(x_i) = \mathbf{P}(X = x_i).$$

Tale distribuzione marginale si ottiene dalla congiunta nel modo seguente:

$$p_1(x_i) = \mathbf{P}(X = x_i) = \sum_{j=1}^n \mathbf{P}(X = x_i, Y = y_j) = \sum_{j=1}^n p(x_i, y_j).$$

Analogamente, si definisce la *distribuzione marginale* di Y

$$p_2(y_j) = \mathbf{P}(Y = y_j) = \sum_{i=1}^m p(x_i, y_j).$$

Ne segue che due numeri aleatori sono stocasticamente indipendenti se e solo se

$$\forall(i, j) \quad p(x_i, y_j) = p_1(x_i)p_2(y_j). \quad (2.1)$$

Data $\psi : \mathbb{R}^2 \longrightarrow \mathbb{R}$, la previsione del numero aleatorio $Z = \psi(X, Y)$ si ottiene utilizzando la distribuzione congiunta di (X, Y) nel modo seguente:

$$\mathbf{P}(Z) = \mathbf{P}(\psi(X, Y)) = \sum_{(x_i, y_j) \in I(X, Y)} \psi(x_i, y_j) \mathbf{P}(X = x_i, Y = y_j), \quad (2.2)$$

se la sommatoria a destra esiste. La dimostrazione è analoga al caso di un solo numero aleatorio. Per esempio, si può calcolare $\mathbf{P}(XY)$ usando (2.2). Si ottiene

$$\mathbf{P}(XY) = \sum_{(x_i, y_j) \in I(X, Y)} x_i y_j \mathbf{P}(X = x_i, Y = y_j).$$

Inoltre, se X, Y sono stocasticamente indipendenti e $\phi_1, \phi_2 : \mathbb{R} \longrightarrow \mathbb{R}$, si ha che

$$\mathbf{P}(\phi_1(X)\phi_2(Y)) = \mathbf{P}(\phi_1(X))\mathbf{P}(\phi_2(Y)), \quad (2.3)$$

in quanto dalla (2.1) e dalla (2.2) segue che

$$\begin{aligned} \mathbf{P}(\phi_1(X)\phi_2(Y)) &= \sum_{(x_i, y_j) \in I(X, Y)} \phi_1(x_i)\phi_2(y_j)\mathbf{P}(X = x_i, Y = y_j) = \\ &= \sum_{(x_i, y_j) \in I(X) \times I(Y)} \phi_1(x_i)\phi_2(y_j)p_1(x_i)p_2(y_j) = \\ &= \sum_{x_i \in I(X)} \phi_1(x_i)p_1(x_i) \sum_{y_j \in I(Y)} \phi_2(y_j)p_2(y_j) = \\ &= \mathbf{P}(\phi_1(X))\mathbf{P}(\phi_2(Y)), \end{aligned}$$

se le sommatorie esistono.

2.12 La varianza nelle distribuzioni discrete

Si calcola ora la varianza per le distribuzioni discrete viste in precedenza.

1. *Varianza di un evento*

$$\sigma^2(E_i) = \mathbf{P}(E_i^2) - \mathbf{P}(E_i)^2 = p(1-p).$$

2. *Distribuzione binomiale*: Si utilizza la rappresentazione $X = E_1 + \dots + E_n$, dove gli E_i sono stocasticamente indipendenti. Si ottiene:

$$\sigma^2(E_1 + \dots + E_n) = \sum_{i=1}^n \sigma^2(E_i) = np(1-p).$$

3. *Distribuzione geometrica*: sapendo che $\sigma^2(X) = \mathbf{P}[X^2] - \mathbf{P}(X)^2$, basta calcolare

$$\mathbf{P}(X^2) = \sum_{i=1}^{+\infty} i^2 p(1-p)^{i-1}.$$

Si ottiene

$$\begin{aligned} \mathbf{P}(X^2) &= p \sum_{i=1}^{+\infty} i^2 (1-p)^{i-1} = \left(p \sum_{i=1}^{+\infty} i(i-1)(1-p)^{i-1} \right) + \mathbf{P}(X) \\ &= \left(p(1-p) \sum_{i=2}^{+\infty} i(i-1)(1-p)^{i-2} \right) + \mathbf{P}(X) \\ &= p(1-p) \frac{d}{dp} \left(- \sum_{i=1}^{+\infty} i(1-p)^{i-1} \right) + \mathbf{P}(X) \\ &= p(1-p) \frac{d^2}{d^2 p} \underbrace{\left(\sum_{i=1}^{+\infty} (1-p)^i \right)}_{\frac{1}{p}} + \frac{1}{p} \\ &= \frac{2(1-p)}{p^2} + \frac{1}{p} \\ &= \frac{2}{p^2} - \frac{1}{p}. \end{aligned}$$

Infine si ottiene che per la distribuzione geometrica la varianza è data da

$$\sigma^2(X) = \mathbf{P}[X^2] - \mathbf{P}(X)^2 = \frac{(1-p)}{p^2}.$$

4. *Distribuzione di Poisson*: Se X ha distribuzione di Poisson di parametro λ , si calcola

$$\begin{aligned}
P(X^2) &= \sum_{i=0}^{+\infty} i^2 \mathbf{P}(X = i) = \sum_{i=0}^{+\infty} i^2 \frac{\lambda^i}{i!} e^{-\lambda} = e^{-\lambda} \sum_{i=0}^{+\infty} [(i^2 - i) + i] \frac{\lambda^i}{i!} = \\
&= \lambda^2 e^{-\lambda} \sum_{i=2}^{+\infty} \frac{\lambda^{i-2}}{(i-2)!} + \lambda e^{-\lambda} \sum_{i=1}^{+\infty} \frac{\lambda^{i-1}}{(i-1)!} = \lambda e^{-\lambda} \sum_{k=0}^{+\infty} \frac{\lambda^k}{k!} + \lambda = \lambda^2 + \lambda.
\end{aligned}$$

Si ottiene che la varianza è data da

$$\sigma^2(X) = \mathbf{P}(X^2) - \mathbf{P}(X)^2 = \lambda^2 + \lambda - \lambda^2 = \lambda.$$

5. *Distribuzione ipergeometrica*: Nella stessa notazione della sezione 2.6, si utilizza la rappresentazione $X = E_1 + \dots + E_n$. Tali eventi non sono stocasticamente indipendenti e risultano correlati negativamente. Infatti, se $H < N$ per ogni scelta di $i \neq j \in \{1, \dots, n\}$ si ha

$$\mathbf{cov}(E_i, E_j) = \mathbf{P}(E_i E_j) - \mathbf{P}(E_i) \mathbf{P}(E_j) = \frac{H}{N^2} \frac{H-N}{N-1} < 0$$

in quanto

$$\mathbf{P}(E_i E_j) = \mathbf{P}(E_i | E_j) \mathbf{P}(E_j) = \frac{(H-1) D_{n-2}^{N-2}}{D_{n-1}^{N-1}} \frac{H}{N} = \frac{(H-1) H}{(N-1) N}.$$

La varianza di X si ottiene utilizzando la formula della varianza della somma di n numeri aleatori

$$\begin{aligned}
\sigma^2(X) &= \sum_{i=1}^n \sigma^2(E_i) + \sum_{\substack{i,j \\ i \neq j}} \mathbf{cov}(E_i, E_j) \\
&= n \frac{H}{N} \left(1 - \frac{H}{N}\right) + D_2^n \frac{H}{N^2} \frac{H-N}{N-1} = n \frac{N-n}{N-1} \frac{H}{N} \left(1 - \frac{H}{N}\right).
\end{aligned}$$

La disposizione $D_2^n = \frac{n!}{(n-2)!}$ conta il numero numero di elementi nella sommatoria $\sum_{\substack{i,j \\ i \neq j}} \mathbf{cov}(E_i, E_j)$ che corrisponde al numero di coppie ordinate di elementi distinti scelti su n .

2.13 Non correlazione ed indipendenza stocastica

Consideriamo due numeri aleatori X e Y con distribuzione congiunta discreta data da

$$p_{ij} = \mathbf{P}(X = x_i, Y = y_j)$$

e distribuzioni marginali rispettivamente date da

$$p_i = \mathbf{P}(X = x_i) \quad i = 1, \dots, m.$$

$$q_j = \mathbf{P}(Y = y_j) \quad j = 1, \dots, n.$$

X e Y sono non correlati se e solo se

$$\mathbf{P}(XY) = \mathbf{P}(X)\mathbf{P}(Y)$$

ovvero se e solo se

$$\sum_{i,j} x_i y_j p_{i,j} = \sum_i x_i p_i \sum_j y_j q_j.$$

Inoltre, devono valere le relazioni:

$$\begin{aligned} \sum_i p_i &= 1 & \sum_j p_{i,j} &= p_i \quad \forall i \\ \sum_j q_j &= 1 & \sum_i p_{i,j} &= q_j \quad \forall j, \end{aligned}$$

$$\sum_{i,j} p_{i,j} = 1.$$

Supponiamo di voler trovare valori di $p_{i,j}$ in modo che X e Y risultino non correlati e abbiano due distribuzioni marginali fissate $\{p_i\}$ e $\{q_j\}$. Osserviamo prima di tutto che i $p_{i,j}$ devono soddisfare la relazione $\sum_{i,j} p_{i,j} = 1$. Per fissare le due distribuzioni marginali dobbiamo imporre $m + n$ relazioni lineari sulle $p_{i,j}$. In realtà però basta imporne solo $(m - 1) + (n - 1)$ dato che $\sum_{i,j} p_{i,j} =$

1 , $\sum_i p_i = 1$, $\sum_j q_j = 1$. Infine per imporre la non correlazione dobbiamo imporre un'ulteriore condizione lineare

$$\sum_{i,j} x_i y_j p_{i,j} = ab,$$

dove $a = \sum_i x_i p_i$ e $b = \sum_j y_j q_j$. Quindi abbiamo un sistema di $1 + (m - 1) + (n - 1) + 1 = m + n$ equazioni lineari per mn incognite. Questo sistema ha sicuramente la soluzione $p_{i,j} = p_i q_j$, che è quella per cui X e Y sono stocasticamente indipendenti. Perché sia l'unica soluzione bisogna che il numero di equazioni linearmente indipendenti sia eguale al numero delle incognite, cioè $m + n = mn$, ovvero $mn - m - n = (m - 1)(n - 1) - 1 = 0$ e questo si ha solo se $m = n = 2$. Segue immediatamente che la non correlazione non implica in generale l'indipendenza stocastica. Se $m = n = 2$ allora il numero di relazioni è pari al numero di variabili; quindi, *due eventi sono non correlati se e solo se sono stocasticamente indipendenti*.

Si verifica che l'indipendenza stocastica implica invece la non correlazione. Se X, Y sono due numeri aleatori stocasticamente indipendenti con distribuzione discreta congiunta $p_{i,j} = \mathbf{P}(X = x_i, Y = y_j)$, $i = 1, \dots, m$, $j = 1, \dots, n$, la loro covarianza $\mathbf{cov}(X, Y)$ si calcola nel modo seguente:

$$\begin{aligned}
\text{cov}(X, Y) &= \mathbf{P}(XY) - \mathbf{P}(X)\mathbf{P}(Y) \\
&= \sum_{i,j} x_i y_j p_{i,j} - \left(\sum_i x_i p_i \right) \left(\sum_j x_j q_j \right) \\
&= \sum_{i,j} x_i y_j p_i q_j - \left(\sum_i x_i p_i \right) \left(\sum_j x_j q_j \right) \\
&= 0,
\end{aligned}$$

in quanto la distribuzione congiunta di due numeri aleatori stocasticamente indipendenti è data dal prodotto delle distribuzioni marginali p_i, q_j (proprietà (2.1)).

2.14 La funzione generatrice

Sia X un numero aleatorio con distribuzione discreta sugli interi non negativi. Dato $u \in \mathbb{C}$ tale che $|u| \leq 1$, si definisce *funzione generatrice* di X la funzione

$$\phi_X(u) := \mathbf{P}(u^X) = \sum_{k \in I(X)} u^k \mathbf{P}(X = k). \quad (2.4)$$

Anche nel caso in cui la sommatoria sia infinita, la condizione $|u| \leq 1$ garantisce che la serie (2.4) di potenze converge. Si osserva che

$$\phi_X(u) \Big|_{u=0} = \mathbf{P}(X = 0).$$

In generale calcolando la derivata n -esima della serie (2.4) in $u = 0$, si ottiene

$$\mathbf{P}(X = n) = \frac{1}{n!} \frac{d^n \phi_X(u)}{dx^n} \Big|_{u=0},$$

per ogni $n \in I(X)$. Dalla funzione generatrice si può quindi ottenere la distribuzione di probabilità di X . Inoltre dalla conoscenza di ϕ_X si possono calcolare anche la previsione e la varianza (se esistono).

Proposizione 2.14.1. *Se $\mathbf{P}(X) < \infty$ vale $\mathbf{P}(X) = \lim_{u \rightarrow 1^-} \phi'_X(u)$. Se invece $\lim_{u \rightarrow 1^-} \phi'_X(u) = \infty$, allora $\mathbf{P}(X) = \infty$.*

Vale in realtà un risultato più generale.

Proposizione 2.14.2. *Sia ha che $\mathbf{P}(X(X-1)\cdots(X-k+1)) < \infty$ se e solo se $\lim_{u \rightarrow 1^-} \phi_X^{(k)}(u) < \infty$, dove $\phi_X^{(k)}(u)$ indica la derivata k -esima della funzione generatrice, e vale che*

$$\mathbf{P}(X(X-1)\cdots(X-k+1)) = \lim_{u \rightarrow 1^-} \phi_X^{(k)}(u).$$

In particolare si può applicare questo risultato ponendo $k = 2$ per calcolare la varianza di X a partire dalla sua funzione generatrice, come segue:

$$\sigma^2(X) = \mathbf{P}(X^2) - \mathbf{P}(X)^2 = \phi_X''(u) + \phi_X'(u) - (\phi_X'(u))^2,$$

dove ϕ_X' e ϕ_X'' indicano rispettivamente la derivata prima e seconda di ϕ_X . Si possono calcolare facilmente le funzioni generatrici delle distribuzioni discrete più comuni:

1. Schema di Bernoulli:

$$\phi_X(u) = up + (1 - p).$$

2. Distribuzione binomiale $\mathcal{B}(n, p)$

$$\begin{aligned} \phi_X(u) &= \sum_{k=0}^n u^k \binom{n}{k} p^k (1-p)^{n-k} \\ &= \sum_{k=0}^n \binom{n}{k} (up)^k (1-p)^{n-k} = (up + (1-p))^n, \end{aligned}$$

dove si è applicato lo sviluppo del binomio di Newton.

3. Distribuzione geometrica

$$\begin{aligned} \phi_X(u) &= \sum_{k=1}^{\infty} u^k p (1-p)^{k-1} \\ &= up \sum_{k=1}^{\infty} [u(1-p)]^{k-1} = \frac{up}{1 - u(1-p)}, \end{aligned}$$

dove si è usata la formula della somma della serie geometrica.

4. Distribuzione di Poisson

$$\begin{aligned} \phi_X(u) &= \sum_{k=0}^{\infty} u^k \frac{\lambda^k}{k!} e^{-\lambda} \\ &= e^{-\lambda} \sum_{k=0}^{\infty} \frac{(u\lambda)^k}{k!} = e^{-\lambda(1-u)}, \end{aligned}$$

dove si è usato lo sviluppo in serie somma della funzione esponenziale.

Se X, Y sono due numeri aleatori stocasticamente indipendenti, allora si dimostra facilmente che

$$\phi_{X+Y}(u) = \phi_X(u)\phi_Y(u).$$

Infatti basta osservare che

$$\begin{aligned}\phi_{X+Y}(u) &= \mathbf{P}(u^{X+Y}) = \mathbf{P}(u^X u^Y) = \\ \sum_{i,j} u^{i+j} \mathbf{P}(X=i, Y=j) &= \sum_{i,j} u^i u^j \mathbf{P}(X=i) \mathbf{P}(Y=j) = \phi_X(u) \phi_Y(u),\end{aligned}$$

se X, Y sono stocasticamente indipendenti. Questa proprietà si può generalizzare al caso in cui si voglia calcolare la funzione generatrice della somma aleatoria

$$S_N = X_1 + \cdots + X_N,$$

dove N è un *numero aleatorio* con distribuzione discreta ed $(X_n)_{n \in \mathbb{N}}$ una successione di numeri aleatori stocasticamente indipendenti ed equidistribuiti. Supponiamo per semplicità che l'insieme dei valori possibili $I(N)$ di N sia finito. Usando la formula delle probabilità totali si ottiene che

$$\begin{aligned}\phi_{S_N}(u) &= \mathbf{P}(u^{X_1 + \cdots + X_N}) = \sum_{k \in I(N)} \mathbf{P}(u^{X_1 + \cdots + X_N} | N = k) \mathbf{P}(N = k) \\ &= \sum_{k \in I(N)} \mathbf{P}(u^{X_1 + \cdots + X_k}) \mathbf{P}(N = k) \\ &= \sum_{k \in I(N)} \mathbf{P}(u^{X_1}) \cdots \mathbf{P}(u^{X_k}) \mathbf{P}(N = k) \\ &= \sum_{k \in I(N)} \phi_{X_1}(u) \cdots \phi_{X_k}(u) \mathbf{P}(N = k) \\ &= \sum_{k \in I(N)} \phi_{X_1}(u)^k \mathbf{P}(N = k) \\ &= \phi_N(\phi_{X_1}(u)),\end{aligned}$$

dove ϕ_N è la funzione generatrice di N . La funzione generatrice di $S_N = X_1 + \cdots + X_N$ è data da quella di N calcolata nella funzione generatrice delle X_i .

Per una trattazione completa sulla funzione generatrice e la dimostrazione dei risultati citati, si rimanda a [FF].

Distribuzioni assolutamente continue unidimensionali

3.1 Introduzione

Negli esempi di numeri aleatori con distribuzione discreta, la distribuzione era specificata completamente dalla probabilità di assumere i singoli valori. Se si vogliono introdurre numeri aleatori che possano assumere valori, ad esempio in un intervallo reale $[a, b]$, per definire un valore “scelto a caso” in $[a, b]$, le probabilità dei singoli valori non sono chiaramente più sufficienti per descriverne la distribuzione. Nell'esempio citato la probabilità di ogni singolo valore deve essere uguale a 0, ma questo non ci dice nulla sulla probabilità che il numero aleatorio appartenga ad un intervallo contenuto in $[a, b]$. Vogliamo quindi trovare un modo per descrivere la distribuzione di un numero aleatorio generale.

3.2 Funzione di ripartizione

Dato un numero aleatorio X , si introduce la *funzione di ripartizione* (o *distribuzione*) F di X data da

$$F(x) := \mathbf{P}(X \leq x), \quad x \in \mathbb{R}.$$

Assegnare la distribuzione di probabilità di X significa specificare la sua *funzione di ripartizione*.

Esempio 3.2.1 (Caso discreto). Nel caso di numeri aleatori con distribuzione discreta si ha

$$F(x) = \mathbf{P}(X \leq x) = \sum_{x_i \leq x} \mathbf{P}(X = x_i).$$

Se si conosce F , si può calcolare la probabilità degli intervalli

$$\mathbf{P}(a < X \leq b) = F(b) - F(a).$$

La funzione di ripartizione è una funzione reale con le seguenti proprietà:

1. $0 \leq F(x) \leq 1$ (per definizione)
2. *monotonia*: $F(b) \geq F(a)$ se $b > a$ in quanto $F(b) - F(a) = \mathbf{P}(a < X \leq b) \geq 0$.

Le seguenti proprietà aggiuntive si suppongono in genere verificate. Si possono pensare come proprietà di regolarità perché affermano che la probabilità di un evento $P(E)$ è uguale al limite delle probabilità $P(E_n)$ dove E_n è una successione di eventi che converge ad E .

1. *continuità a destra*: $F(x) = \lim_{y \rightarrow x^+} F(y)$,
2. $\lim_{x \rightarrow +\infty} F(x) = 1$,
3. $\lim_{x \rightarrow -\infty} F(x) = 0$.

Le ultime tre condizioni sono quindi proprietà aggiuntive di regolarità che saranno sempre verificate nei casi che considereremo. È possibile considerare casi in cui non valgono. Poiché F è monotona e limitata, il limite a sinistra esiste ed è finito. Nei casi che considereremo, per le stesse ragioni di regolarità tale limite è dato da:

$$F(x^-) = \lim_{y \rightarrow x^-} F(y) = \lim_{y \rightarrow x^-} \mathbf{P}(X \leq y) = \mathbf{P}(X < x).$$

Da cui $\mathbf{P}(X = x) = F(x) - F(x^-)$.

3.3 Distribuzioni assolutamente continue

Sia X un numero aleatorio. Si dice che X ha *distribuzione assolutamente continua* se esiste una funzione $f: \mathbb{R} \rightarrow \mathbb{R}$ con le seguenti proprietà:

$$\forall x \in \mathbb{R}, f(x) \geq 0, \quad (3.1)$$

$$f \text{ integrabile}, \quad (3.2)$$

$$\int_{\mathbb{R}} f(s) ds = 1, \quad (3.3)$$

tale che la funzione di ripartizione di X si scrive come

$$F(x) = \int_{-\infty}^x f(t) dt.$$

Tale funzione si dice *densità di probabilità*. Si noti che la f non è unica. Infatti, ad esempio, se cambiamo la f in un insieme numerabile di punti, la nuova funzione è ancora una densità di X poiché il suo integrale non cambia. Per questo la funzione di densità associata ad una distribuzione assolutamente continua non è unica, ma possiede infiniti rappresentanti. Nel seguito le

proprietà elencate valgono a prescindere dal rappresentante scelto.

Date la funzione di ripartizione e la funzione di densità di probabilità, vale la seguente uguaglianza:

$$f(x) = \frac{dF}{dx}$$

nei punti in cui f è continua.

La condizione (3.3) deriva dal fatto che dalle ipotesi di regolarità sulla funzione di distribuzione si deve avere

$$1 = \lim_{x \rightarrow +\infty} F(x) = \int_{-\infty}^{+\infty} f(s) ds.$$

La proprietà (3.1) si giustifica nel caso in cui f sia continua nell'intervallo $[a, b]$ nel seguente modo. Dal teorema del valor medio, esiste $\xi \in (a, b)$

$$\mathbf{P}(a < X \leq b) = F(b) - F(a) = \int_a^b f(s) ds = f(\xi)(b - a),$$

da cui

$$f(\xi)(b - a) > 0 \Rightarrow f \geq 0.$$

Si noti che la probabilità degli intervalli si calcola come:

$$\begin{aligned} \mathbf{P}(a < X \leq b) &= F(b) - F(a) \\ &= \int_{-\infty}^b f(s) ds - \int_{-\infty}^a f(s) ds = \int_a^b f(s) ds. \end{aligned}$$

Per quanto riguarda la previsione, pensiamo dapprima al caso di un numero aleatorio X che assuma valori in un intervallo $[a, b]$ con una densità f continua che sarà dunque uguale a zero fuori di $[a, b]$. Possiamo suddividere $[a, b]$ in n intervallini I_i di lunghezza $\frac{b-a}{n}$ (l'inclusione a meno degli estremi negli intervalli non ha importanza in questo caso e possiamo quindi supporre che gli intervalli siano chiusi a destra ed aperti a sinistra, tranne il primo che supponiamo chiuso). Definiamo quindi due numeri aleatori discreti $X_-^{(n)}$ e $X_+^{(n)}$: se X assume valori in I_i , $X_-^{(n)}$ è uguale all'estremo sinistro di I_i e $X_+^{(n)}$ all'estremo destro di I_i . Dato che $X_-^{(n)}$ e $X_+^{(n)}$ hanno distribuzione discreta con un numero finito di valori, possiamo calcolare le loro previsioni che sono rispettivamente uguali a

$$\begin{aligned} \mathbf{P}(X_-^{(n)}) &= \sum_{j=0}^{n-1} \left(a + j \frac{b-a}{n} \right) \int_{a+j \frac{b-a}{n}}^{a+(j+1) \frac{b-a}{n}} f(x) dx, \\ \mathbf{P}(X_+^{(n)}) &= \sum_{j=0}^{n-1} \left(a + (j+1) \frac{b-a}{n} \right) \int_{a+j \frac{b-a}{n}}^{a+(j+1) \frac{b-a}{n}} f(x) dx. \end{aligned}$$

D'altra parte si ha $X_-^{(n)} \leq X \leq X_+^{(n)}$, quindi dobbiamo avere

$$\mathbf{P}(X_-^{(n)}) \leq \mathbf{P}(X) \leq \mathbf{P}(X_+^{(n)}).$$

Si vede facilmente usando la continuità di $f(x)$ che quando $n \rightarrow \infty$ sia $P(X_-^{(n)})$ che $P(X_+^{(n)})$ tendono a

$$\int_a^b xf(x)dx = \int_{\mathbb{R}} xf(x)dx,$$

che quindi è il valore di $P(X)$. Argomenti di approssimazione portano ad estendere questa formula al caso generale di una X con distribuzione assolutamente continua con densità $f(x)$ purché

$$\int_{\mathbb{R}} |x|f(x)dx < \infty, \quad (3.4)$$

ovvero si assume che se vale (3.4) la previsione nel caso assolutamente continuo sia data da

$$\mathbf{P}(X) = \int_{-\infty}^{+\infty} xf(x)dx.$$

Analogamente se $\psi : \mathbb{R} \rightarrow \mathbb{R}$ è una funzione reale tale che $\psi(x)f(x)$ sia integrabile, siamo portati ad assegnare a $\mathbf{P}(\psi(X))$ il valore

$$\mathbf{P}(\psi(X)) = \int_{-\infty}^{+\infty} \Psi(x)f(x)dx. \quad (3.5)$$

Ne segue che la varianza si ottiene come

$$\begin{aligned} \sigma^2(X) &= \mathbf{P}(X^2) - \mathbf{P}(X)^2 \\ &= \int_{-\infty}^{+\infty} x^2 f(x) dx - \left(\int_{-\infty}^{+\infty} xf(x) dx \right)^2. \end{aligned}$$

Nelle prossime sezioni introduciamo alcune fra le più note delle distribuzioni assolutamente continue unidimensionali.

3.4 Distribuzione uniforme in $[0, 1]$

Un numero aleatorio X ha *distribuzione uniforme* in $[0, 1]$ se la sua funzione di distribuzione è data da

$$F(x) = \begin{cases} 0 & x \leq 0 \\ x & 0 < x < 1 \\ 1 & x \geq 1. \end{cases}$$

La probabilità di un singolo punto è

$$\mathbf{P}(X = x) = F(x) - F(x^-) = 0$$

La densità di probabilità si definisce come

$$f(x) = F'(x) = \begin{cases} 0 & x \leq 0 \\ 1 & 0 < x < 1 \\ 0 & x \geq 1. \end{cases}$$

Come nei casi seguenti, il valore della densità nei punti di discontinuità può essere scelto in maniera arbitraria. La previsione della distribuzione uniforme è

$$\mathbf{P}(X) = \int_{\mathbb{R}} xf(x) dx = \int_0^1 x dx = \left[\frac{x^2}{2} \right]_0^1 = \frac{1}{2},$$

mentre la varianza è data da

$$\sigma^2(X) = \int_0^1 x^2 dx - \frac{1}{4} = \frac{1}{12}.$$

3.5 Distribuzione uniforme su un intervallo qualunque $[a, b]$

Un numero aleatorio X ha *distribuzione uniforme* in $[a, b]$ se la sua funzione di distribuzione è data da

$$F(x) = \begin{cases} 0 & x \leq a \\ c(x - a) & a < x < b \\ 1 & x \geq b \end{cases}$$

Per calcolare la costante c , basta imporre che $F(b) = 1$. Si ottiene

$$c = \frac{1}{b - a}.$$

La previsione della distribuzione uniforme è

$$\mathbf{P}(X) = \int_a^b \frac{x}{b - a} dx = \left[\frac{x^2}{2(b - a)} \right]_a^b = \frac{a + b}{2},$$

mentre la varianza è data da

$$\begin{aligned} \sigma^2(X) &= \mathbf{P}((X - \mathbf{P}(X))^2) = \int_a^b \frac{1}{b - a} \left(x - \frac{a + b}{2}\right)^2 dx \\ &= \frac{1}{b - a} \frac{1}{3} \left[\left(x - \frac{a + b}{2}\right)^3 \right]_a^b = \frac{(b - a)^2}{12}. \end{aligned}$$

3.6 Distribuzione esponenziale di parametro λ

Un numero aleatorio X ha *distribuzione esponenziale* di parametro λ se la sua funzione di distribuzione è data da

$$F(x) = \begin{cases} 1 - e^{-\lambda x} & x \geq 0 \\ 0 & x < 0. \end{cases}$$

La densità è data da:

$$f(x) = F'(x) = \begin{cases} \lambda e^{-\lambda x} & x \geq 0 \\ 0 & x < 0 \end{cases}$$

Se attribuiamo ad X il significato di un tempo aleatorio in cui si verifica un fatto (ad esempio l'istante di decadimento di un atomo), la distribuzione esponenziale ha la proprietà di *assenza di memoria*, ovvero dati $x, y \geq 0$ vale che

$$\mathbf{P}(X > x + y | X > y) = \mathbf{P}(X > x). \quad (3.6)$$

Usiamo la formula delle probabilità composte per dimostrare (3.6):

$$\begin{aligned} \mathbf{P}(X > x + y | X > y) &= \frac{\mathbf{P}(X > x + y, X > y)}{\mathbf{P}(X > y)} \\ &= \frac{\mathbf{P}(X > x + y)}{\mathbf{P}(X > y)} \\ &= \frac{e^{-\lambda(x+y)}}{e^{-\lambda y}} \\ &= e^{-\lambda x} \\ &= \mathbf{P}(X > x). \end{aligned}$$

Vedremo in seguito come la distribuzione esponenziale possa essere ottenuta come limite della distribuzione geometrica, che pure possiede la proprietà di assenza di memoria. La previsione della distribuzione esponenziale è

$$\mathbf{P}(X) = \int_0^{+\infty} \lambda x e^{-\lambda x} dx = [-x e^{-\lambda x}]_0^{+\infty} + \int_0^{+\infty} e^{-\lambda x} dx = \frac{1}{\lambda},$$

mentre la varianza è

$$\begin{aligned} \sigma^2(X) &= \mathbf{P}(X^2) - \mathbf{P}(X)^2 \\ &= \int_0^{+\infty} \lambda x^2 e^{-\lambda x} dx - \frac{1}{\lambda^2} \\ &= [-x^2 e^{-\lambda x}]_0^{+\infty} + 2 \int_0^{+\infty} x e^{-\lambda x} dx - \frac{1}{\lambda^2} \\ &= \frac{2}{\lambda^2} - \frac{1}{\lambda^2} \\ &= \frac{1}{\lambda^2}. \end{aligned}$$

3.7 Un'altra caratterizzazione della distribuzione esponenziale

Sia X un numero aleatorio con distribuzione esponenziale e siano assegnati $x, y > 0$. Per caratterizzare questa distribuzione possiamo anche usare il tasso di rischio.

Definiamo *tasso di rischio* al tempo x

$$h(x) = \lim_{h \rightarrow 0} \frac{\mathbf{P}(x < X < x + h | X > x)}{h}.$$

Si può esprimere $h(x)$ mediante la densità usando la formula delle probabilità subordinate. Supponiamo

$$F(x) = \mathbf{P}(X \leq x) = \int_{-\infty}^x f(y) dy.$$

Ora

$$\lim_{h \rightarrow 0} \frac{\mathbf{P}(x < X < x + h)}{h\mathbf{P}(X > x)} = \frac{f(x)}{1 - F(x)} = -\frac{d}{dx} \log(1 - F(x)).$$

Per la distribuzione esponenziale di parametro λ , si vede facilmente che il tasso di rischio è costante e pari a

$$h(x) = \frac{f(x)}{1 - F(x)} = \frac{\lambda e^{-\lambda x}}{e^{-\lambda x}} = \lambda.$$

L'altra caratteristica della distribuzione esponenziale che stavamo cercando è che ha *il tasso di rischio costante pari a λ* .

Fissata $h(x)$ possiamo determinare $F(x)$ supponendo $F(0) = 0$, risolvendo l'equazione

$$h(x) = -\frac{d}{dx} \log(1 - F(x)),$$

da cui

$$F(x) = 1 - e^{-\int_0^x h(y) dy}.$$

3.8 Distribuzione normale

Un numero aleatorio X ha *distribuzione normale standard* (si indica con la notazione $N(0, 1)$) se la sua funzione di densità è

$$f(x) = K e^{-\frac{x^2}{2}}, \quad x \in \mathbb{R}.$$

La costante K di normalizzazione si può calcolare nel seguente modo:

$$\begin{aligned} \left(\int_{\mathbb{R}} e^{-\frac{x^2}{2}} dx \right)^2 &= \int_{\mathbb{R}} \int_{\mathbb{R}} e^{-\frac{x^2}{2}} e^{-\frac{y^2}{2}} dx dy = \\ \int_{\mathbb{R}} e^{-\frac{x^2+y^2}{2}} dx dy &= \int_0^{2\pi} \int_0^{+\infty} e^{-\frac{\rho^2}{2}} \rho d\rho d\theta = \\ 2\pi \int_0^{+\infty} e^{-\frac{\rho^2}{2}} \rho d\rho &= 2\pi \left[-e^{-\frac{\rho^2}{2}} \right]_0^{+\infty} = 2\pi. \end{aligned}$$

dove si è effettuato il cambio di variabile $x = \rho \cos \theta$, $y = \rho \sin \theta$ ed il determinante jacobiano di tale sostituzione è pari a ρ (si veda l'appendice G).

Ne segue che $K^{-2} = 2\pi$, ovvero $K^{-1} = \sqrt{2\pi}$, quindi

$$K = \frac{1}{\sqrt{2\pi}}$$

La funzione di ripartizione si indica con

$$\mathcal{N}(x) := \int_{-\infty}^x n(t) dt,$$

dove si è definito $n(t) := \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}}$.

Per la simmetria, si ottiene che

$$\mathcal{N}(-x) = 1 - \mathcal{N}(x)$$

La previsione della distribuzione normale standard è

$$\mathbf{P}(X) = \int_{\mathbb{R}} x n(x) dx = 0$$

poiché la funzione $f(x) = x e^{-\frac{x^2}{2}}$ è dispari, mentre la varianza è data da

$$\sigma^2(X) = \int_{\mathbb{R}} \frac{x^2}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx = \underbrace{\left[-\frac{x}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \right]_{-\infty}^{+\infty}}_{\substack{\text{funzione dispari} \\ 0}} + \underbrace{\int_{\mathbb{R}} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx}_1 = 1.$$

Introduciamo la distribuzione *normale*, indicata con la notazione $N(\mu, \sigma^2)$. Sia $X \sim N(0, 1)$ e consideriamo $Y = \mu + \sigma X$, con $\sigma > 0$; la funzione di distribuzione di Y è

$$\begin{aligned} F_Y(y) &= \mathbf{P}(Y \leq y) \\ &= \mathbf{P}(\mu + \sigma X \leq y) \\ &= \mathbf{P}\left(X \leq \frac{y - \mu}{\sigma}\right) \\ &= \mathcal{N}\left(\frac{y - \mu}{\sigma}\right). \end{aligned}$$

La densità di Y è allora

$$f_Y(y) = \frac{d}{dy} \mathcal{N}\left(\frac{y-\mu}{\sigma}\right) = \frac{1}{\sigma} n\left(\frac{y-\mu}{\sigma}\right) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(y-\mu)^2}{2\sigma^2}}.$$

3.9 Stima delle code

Non esiste una formula in termini di funzioni elementari per $\mathcal{N}(x)$ e quindi la probabilità che $X \sim N(0, 1)$ sia più grande di un $x > 0$ fissato. Possiamo darne delle stime asintotiche dall'alto e dal basso.

Proposizione 3.9.1. *Sia X un numero aleatorio con distribuzione normale standard. Per ogni $x > 0$, vale che*

$$\frac{n(x)}{x} - \frac{n(x)}{x^3} < \mathbf{P}(X \geq x) < \frac{n(x)}{x},$$

dove $n(x) := \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$.

Il procedimento consiste nell'integrazione per parti della funzione di densità di probabilità $n(x)$. La prima integrazione per parti ci fornisce la maggiorazione:

$$\begin{aligned} \mathbf{P}(X > x) &= \int_x^{+\infty} n(t) dt = \int_x^{+\infty} t \frac{n(t)}{t} dt \\ &= \underbrace{\left[-\frac{n(t)}{t}\right]_x^{+\infty}}_{\frac{n(x)}{x}} - \int_x^{+\infty} \underbrace{\frac{n(t)}{t^2}}_{>0} dt < \frac{n(x)}{x}. \end{aligned}$$

Con un'ulteriore integrazione per parti si ottiene la minorazione:

$$\begin{aligned} \mathbf{P}(X > x) &= \frac{n(x)}{x} - \int_x^{+\infty} t \frac{n(t)}{t^3} dt \\ &= \frac{n(x)}{x} - \underbrace{\left[-\frac{n(t)}{t^3}\right]_x^{+\infty}}_{\frac{n(x)}{x^3}} + \int_x^{+\infty} \underbrace{\frac{3n(t)}{t^4}}_{>0} dt > \frac{n(x)}{x} - \frac{n(x)}{x^3}. \end{aligned}$$

3.10 Distribuzione gamma $\Gamma(\alpha, \lambda)$

Siano $\alpha, \lambda > 0$. Il numero aleatorio X ha *distribuzione gamma* di parametri α e λ se X è un numero aleatorio con distribuzione assolutamente continua di densità

$$g_{\alpha, \lambda}(x) = \begin{cases} K x^{\alpha-1} e^{-\lambda x} & x \geq 0 \\ 0 & x < 0. \end{cases}$$

Si noti che la distribuzione esponenziale è un caso particolare di distribuzione gamma corrispondente alla scelta del parametro $\alpha = 1$.

Per calcolare la costante di normalizzazione K , si considera la *funzione gamma di Eulero* definita nel modo seguente:

$$\Gamma(\alpha) = \int_0^{+\infty} x^{\alpha-1} e^{-x} dx.$$

Le proprietà di $\Gamma(\alpha)$ che ci servono per studiare questa distribuzione di probabilità sono le seguenti.

1. $\Gamma(\alpha + 1) = \alpha\Gamma(\alpha)$.

Dimostrazione. Si procede integrando per parti.

$$\begin{aligned} \Gamma(\alpha + 1) &= \int_0^{+\infty} x^{\alpha+1-1} e^{-x} dx \\ &= \int_0^{+\infty} x^{\alpha} e^{-x} dx \\ &= [-x^{\alpha} e^{-x}]_0^{+\infty} + \int_0^{+\infty} \alpha x^{\alpha-1} e^{-x} dx \\ &= 0 + \alpha \int_0^{+\infty} x^{\alpha-1} e^{-x} dx \\ &= \alpha \Gamma(\alpha). \end{aligned}$$

2. Se $\alpha = n$ allora $\Gamma(\alpha) = \Gamma(n) = (n - 1)!$.

Ne segue che la funzione Γ è un'estensione del fattoriale $n!$.

Per calcolare il valore della costante di normalizzazione si procede imponendo che l'integrale della funzione di densità di probabilità sia uguale a 1.

$$\int_{-\infty}^{+\infty} g_{\alpha,\lambda}(x) dx = \int_0^{+\infty} K x^{\alpha-1} e^{-\lambda x} dx = 1.$$

Perciò si ottiene che

$$K = \frac{1}{\int_0^{+\infty} x^{\alpha-1} e^{-\lambda x} dx}$$

Calcoliamo l'integrale al denominatore effettuando il cambio di variabile $y = \lambda x$:

$$\begin{aligned} \int_0^{+\infty} x^{\alpha-1} e^{-\lambda x} dx &= \int_0^{+\infty} \frac{y^{\alpha-1}}{\lambda^{\alpha-1}} e^{-y} \frac{dy}{\lambda} = \\ \frac{1}{\lambda^{\alpha}} \int_0^{+\infty} y^{\alpha-1} e^{-y} dy &= \frac{\Gamma(\alpha)}{\lambda^{\alpha}}. \end{aligned}$$

Ne segue che la costante di normalizzazione c è data da

$$K = \frac{\lambda^\alpha}{\Gamma(\alpha)}.$$

La previsione di questa distribuzione si calcola usando ancora una volta le proprietà della funzione $\Gamma(\alpha)$ come segue:

$$\begin{aligned} \mathbf{P}(X) &= \int_{\mathbb{R}} x g_{\alpha, \lambda}(x) \, dx \\ &= \int_0^{+\infty} x \frac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\lambda x} \, dx \\ &= \frac{\lambda^\alpha}{\Gamma(\alpha)} \int_0^{+\infty} x^\alpha e^{-\lambda x} \, dx \\ &= \frac{\lambda^\alpha}{\Gamma(\alpha)} \frac{\Gamma(\alpha + 1)}{\lambda^{\alpha+1}} \\ &= \frac{\lambda^\alpha}{\Gamma(\alpha)} \frac{\alpha \Gamma(\alpha)}{\lambda^{\alpha+1}}. \\ &= \frac{\alpha}{\lambda}. \end{aligned}$$

Analogamente per la varianza si ottiene

$$\sigma^2(X) = \mathbf{P}(X^2) - \frac{\alpha^2}{\lambda^2},$$

da cui

$$\begin{aligned} \mathbf{P}(X^2) &= \int_0^{+\infty} x^2 \frac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\lambda x} \, dx \\ &= \frac{\lambda^\alpha}{\Gamma(\alpha)} \int_0^{+\infty} x^{\alpha+1} e^{-\lambda x} \, dx \\ &= \frac{\lambda^\alpha}{\Gamma(\alpha)} \frac{\Gamma(\alpha + 2)}{\lambda^{\alpha+2}} \\ &= \frac{\lambda^\alpha}{\Gamma(\alpha)} \frac{(\alpha + 1) \alpha \Gamma(\alpha)}{\lambda^{\alpha+2}} \\ &= \frac{\alpha(\alpha + 1)}{\lambda^2}. \end{aligned}$$

Si può concludere che

$$\sigma^2(X) = \frac{\alpha(\alpha + 1)}{\lambda^2} - \frac{\alpha^2}{\lambda^2} = \frac{\alpha}{\lambda^2}.$$

Si può facilmente notare che la distribuzione esponenziale di parametro λ è un caso particolare di distribuzione Γ che si ottiene ponendo $\alpha = 1$.

3.11 Distribuzione χ^2

Dalla distribuzione normale si può ricavare un'altra distribuzione assolutamente continua utile in statistica. Si tratta della *distribuzione chi-quadro* χ^2 di parametro 1, che come vedremo rappresenta un caso particolare della distribuzione Gamma. Nel capitolo 4 introdurremo la distribuzione χ^2 con parametro $\nu \in \mathbb{N}$.

Sia X un numero aleatorio con distribuzione gaussiana standard $N(0, 1)$ e si consideri $Y = X^2$. Si calcola la funzione di ripartizione. Si osserva subito che

$$F_Y(y) = \mathbf{P}(Y \leq y) = 0 \quad \text{se } y < 0$$

in quanto Y è sempre non negativo. Supponiamo $y \geq 0$. Si ottiene

$$\begin{aligned} F_Y(y) &= \mathbf{P}(Y \leq y) = \mathbf{P}(X^2 \leq y) \\ &= \mathbf{P}(-\sqrt{y} \leq X \leq \sqrt{y}) \\ &= \mathcal{N}(\sqrt{y}) - \mathcal{N}(-\sqrt{y}) \\ &= \mathcal{N}(\sqrt{y}) - (1 - \mathcal{N}(\sqrt{y})) \\ &= 2\mathcal{N}(\sqrt{y}) - 1. \end{aligned}$$

In conclusione

$$\mathbf{P}(Y \leq y) = \begin{cases} 0 & \text{se } y < 0 \\ 2\mathcal{N}(\sqrt{y}) - 1 & \text{se } y \geq 0. \end{cases}$$

Calcoliamo ora la densità di Y (supponiamo $y \geq 0$):

$$\begin{aligned} f_Y(y) &= \frac{d}{dy}(\mathbf{P}(Y \leq y)) \\ &= 2n(y) \frac{1}{\sqrt{y}} \\ &= \frac{1}{\sqrt{2\pi}} \frac{1}{\sqrt{y}} e^{-\frac{y}{2}} \\ &= \frac{1}{\sqrt{2\pi}} y^{\frac{1}{2}-1} e^{-\frac{1}{2}y}. \end{aligned}$$

Otteniamo quindi una distribuzione gamma di parametri $\alpha = \frac{1}{2}$ e $\lambda = \frac{1}{2}$. Da questo segue anche che

$$\frac{1}{\sqrt{2}} \frac{1}{\sqrt{\pi}} = \left(\frac{1}{2}\right)^{\frac{1}{2}} \frac{1}{\Gamma\left(\frac{1}{2}\right)},$$

ovvero

$$\Gamma\left(\frac{1}{2}\right) = \sqrt{\pi}.$$

Si ricava inoltre una formula ricorsiva per calcolare $\Gamma\left(\frac{2k+1}{2}\right)$, $k \in \mathbb{N}$, data da

$$\Gamma\left(\frac{2k+1}{2}\right) = \frac{(2k-1)(2k-3)\cdots 1}{2^k} \Gamma\left(\frac{1}{2}\right) = \frac{(2k-1)(2k-3)\cdots 1}{2^k} \frac{\sqrt{\pi}}{2}.$$

3.12 La distribuzione di Cauchy

Vediamo ora un esempio di una distribuzione di probabilità assolutamente continua che non possiede previsione. Si tratta della *distribuzione di Cauchy* che si può costruire partendo da un numero aleatorio Θ con distribuzione uniforme su $[-\frac{\pi}{2}, \frac{\pi}{2}]$. Il numero aleatorio $Y = \tan \Theta$ ha funzione di ripartizione

$$F_Y(y) = \mathbf{P}(Y \leq y) = \mathbf{P}(X \leq \arctan y) = \int_{-\frac{\pi}{2}}^{\arctan y} \frac{1}{\pi} dx = \frac{1}{\pi} \left(\arctan y + \frac{\pi}{2} \right), \quad y \in \mathbb{R}.$$

Derivando, si ottiene che la densità di Y è dunque

$$f_Y(y) = \frac{1}{\pi(1+y^2)}, \quad y \in \mathbb{R}.$$

Se si vuole calcolare la previsione di Y ci si accorge che non è definita. Bisognerebbe infatti calcolare l'integrale improprio

$$\mathbf{P}(Y) = \int_{\mathbb{R}} y f_Y(y) dy = \int_{\mathbb{R}} y \frac{1}{\pi(1+y^2)} dy = \lim_{n_1, n_2 \rightarrow \infty} \int_{n_1}^{n_2} y \frac{1}{\pi(1+y^2)} dy,$$

ma si nota che si ottengono limiti diversi a seconda di quale sottosuccessione per n_1, n_2 si scelga. Per esempio, se $n_1 = n, n_2 = -n$

$$\lim_{n \rightarrow \infty} \int_{-n}^n y \frac{1}{\pi(1+y^2)} dy = \left[\frac{1}{2\pi} \log(1+y^2) \right]_{-n}^n = 0,$$

mentre se $n_1 = n^2, n_2 = -n$ si ottiene invece

$$\lim_{n \rightarrow \infty} \int_{-n}^{n^2} y \frac{1}{\pi(1+y^2)} dy = \left[\frac{1}{2\pi} \log(1+y^2) \right]_{-n}^{n^2} = +\infty.$$

Quindi il limite non esiste e di conseguenza nemmeno la previsione di Y .

3.13 Funzioni di ripartizione miste

Oltre alle funzioni di ripartizione discrete ed assolutamente continue, vi sono anche funzioni di ripartizione continue ma non assolutamente continue che

però non considereremo in questo libro. Si possono poi considerare funzioni di ripartizione miste che si ottengono come combinazioni lineari convesse di funzioni di ripartizione di tipo diverso. Limitandoci a considerare il tipo discreto e quello assolutamente continuo, dato p , $0 < p < 1$, F_1 discreta ed F_2 assolutamente continua possiamo costruire la *funzione di ripartizione mista* F

$$F(x) := pF_1(x) + (1 - p)F_2(x).$$

Se X ha funzione di ripartizione F ed X_i funzione di ripartizione F_i per $i = 1, 2$, si vede facilmente che per calcolare la previsione di un numero aleatorio $\phi(X)$ funzione di X si ottiene

$$\mathbf{P}(\phi(X)) = p\mathbf{P}(\phi(X_1)) + (1 - p)\mathbf{P}(\phi(X_2)),$$

se i termini a destra hanno senso e quindi è la corrispondente combinazione convessa di due termini: il primo è una somma od una serie, il secondo è un integrale.

Un esempio di un numero aleatorio con funzione di ripartizione mista è il tempo di funzionamento T di un dispositivo, ad esempio di una lampadina, che ha probabilità non nulla p di essere non funzionante sin dall'inizio ed altrimenti un tempo di funzionamento con distribuzione assolutamente continua, ad esempio esponenziale di parametro λ . La funzione di ripartizione di T è quindi data da

$$F(x) = \begin{cases} 0 & \text{per } x < 0 \\ p + (1 - p)(1 - e^{-\lambda x}) & \text{per } x \geq 0, \end{cases}$$

e si ha che $\mathbf{P}(T) = \frac{1 - p}{\lambda}$.

Distribuzioni assolutamente continue n -dimensionali

4.1 Distribuzioni bidimensionali

Consideriamo il vettore aleatorio (X, Y) . La *funzione di ripartizione congiunta* di (X, Y) è definita

$$F(x, y) = \mathbf{P}(X \leq x, Y \leq y).$$

La funzione di ripartizione è quindi una funzione da \mathbb{R}^2 in $[0, 1]$:

$$F : \mathbb{R}^2 \longrightarrow [0, 1].$$

La probabilità degli intervalli per $a_1 < b_1, a_2 < b_2$ è data da

$$\begin{aligned} & \mathbf{P}(a_1 < X \leq b_1, a_2 < Y \leq b_2) = \\ & \mathbf{P}[(X \leq b_1) - (X \leq a_1)]((Y \leq b_2) - (Y \leq a_2)) = \\ & \mathbf{P}(X \leq b_1, Y \leq b_2) - \mathbf{P}(X \leq a_1, Y \leq b_2) - \\ & \mathbf{P}(X \leq b_1, Y \leq a_2) + \mathbf{P}(X \leq a_1, Y \leq a_2) = \\ & F(b_1, b_2) - F(a_1, b_2) - F(b_1, a_2) + F(a_1, a_2). \end{aligned} \quad (4.1)$$

Supporremo sempre verificate le seguenti *ipotesi di continuità*:

1. $\lim_{\substack{x \rightarrow +\infty \\ y \rightarrow +\infty}} F(x, y) = 1,$
2. $\lim_{x \rightarrow -\infty} F(x, y) = \lim_{y \rightarrow -\infty} F(x, y) = 0,$
3. $\lim_{\substack{x \rightarrow x_0^+ \\ y \rightarrow y_0^+}} F(x, y) = F(x_0, y_0),$
4. $\mathbf{P}(X = x_0, Y = y_0) = F(x_0, y_0) - F(x_0^-, y_0) - F(x_0, y_0^-) + F(x_0^-, y_0^-),$

ed altre simili che citeremo quando eventualmente saranno utilizzate.

4.2 Funzioni di ripartizione marginali

Dati X, Y due numeri aleatori con funzione di ripartizione congiunta $F(x, y)$, si dicono *funzioni di ripartizione marginali* le funzioni di ripartizione F_1, F_2 di X e Y .

La funzione di ripartizione marginale per il numero aleatorio X si determina nel modo seguente

$$F_1(x) = \mathbf{P}(X \leq x) = \lim_{y \rightarrow +\infty} F(x, y),$$

dove anche questa è un'ipotesi di continuità. Analogamente, si determina la funzione di ripartizione marginale per Y

$$F_2(y) = \mathbf{P}(Y \leq y) = \lim_{x \rightarrow +\infty} F(x, y).$$

Due numeri aleatori si dicono *stocasticamente indipendenti* se:

$$F(x, y) = F_1(x)F_2(y)$$

per ogni coppia $(x, y) \in \mathbb{R}^2$.

Proposizione 4.2.1. *Due numeri aleatori X, Y sono stocasticamente indipendenti se e solo per ogni quadrupla (a, b, c, d) di numeri reali vale che*

$$\mathbf{P}(a < X \leq b, c < Y \leq d) = \mathbf{P}(a < X \leq b)\mathbf{P}(c < Y \leq d).$$

Dimostrazione. Si usa il fatto che vale (4.1) ottenendo

$$\begin{aligned} \mathbf{P}(a < X \leq b, c < Y \leq d) &= F(b, d) - F(b, c) - F(a, d) + F(a, c) = \\ &= F_1(b)F_2(d) - F_1(b)F_2(c) - F_1(a)F_2(d) + F_1(a)F_2(c) = \\ &= (F_1(b) - F_1(a))(F_2(d) - F_2(c)) = \\ &= \mathbf{P}(a < X \leq b)\mathbf{P}(c < Y \leq d). \end{aligned}$$

4.3 Caso assolutamente continuo

Il vettore aleatorio (X, Y) ha distribuzione assolutamente continua se esiste

$$f : \mathbb{R}^2 \longrightarrow \mathbb{R}$$

tale che

1. f sia non negativa e integrabile,
2. $\int_{\mathbb{R}^2} f(x, y) \, dx dy = 1$

e vale che

$$F(x, y) = \mathbf{P}(X \leq x, Y \leq y) = \int_{-\infty}^x \int_{-\infty}^y f(s, t) \, ds dt.$$

Tale $f(x, y)$ si dice *densità congiunta*. Applicando la formula (4.1) per la probabilità degli intervalli, si ottiene

$$\begin{aligned} \mathbf{P}(a < X \leq b, c < Y \leq d) &= \\ F(b, d) - F(a, d) - F(c, b) + F(a, c) &= \\ \int_{-\infty}^b \int_{-\infty}^d f(s, t) \, ds dt - \int_{-\infty}^a \int_{-\infty}^d f(s, t) \, ds dt - \\ \int_{-\infty}^b \int_{-\infty}^c f(s, t) \, ds dt + \int_{-\infty}^a \int_{-\infty}^c f(s, t) \, ds dt &= \\ \int_a^b \int_c^d f(s, t) \, ds dt. \end{aligned}$$

In generale, la probabilità che il vettore aleatorio (X, Y) appartenga ad una regione A del piano \mathbb{R}^2 è data dall'integrale della densità congiunta su A

$$\mathbf{P}((X, Y) \in A) = \int \int_A f(s, t) \, ds dt.$$

Inoltre, se $\psi : \mathbb{R}^2 \rightarrow \mathbb{R}$ è una funzione tale che ψf è integrabile, posto $Z = \psi(X, Y)$, come nel caso unidimensionale si ha che

$$\mathbf{P}(Z) = \int \int_{\mathbb{R}^2} \psi(s, t) f(s, t) \, ds dt. \quad (4.2)$$

Per esempio, se $Z = XY$ si ottiene che

$$\mathbf{P}(XY) = \int \int_{\mathbb{R}^2} st f(s, t) \, ds dt,$$

se $stf(s, t)$ è integrabile. Per calcolare le *densità marginali* di X e Y si usa il fatto che la funzione di ripartizione marginale si calcola come

$$F_X(x) = \mathbf{P}(X \leq x) = \int_{-\infty}^{+\infty} \int_{-\infty}^x f(s, t) \, ds dt = \int_{-\infty}^x \left(\int_{\mathbb{R}} f(s, t) \, dt \right) ds,$$

dove si è potuto scambiare l'ordine di integrazione grazie al teorema di Fubini-Tonelli. Ne segue che la densità marginale è data da

$$f_X(x) = \int_{-\infty}^{+\infty} f(s, t) \, dt.$$

Analogamente,

$$f_Y(y) = \int_{-\infty}^{+\infty} f(s, t) ds.$$

Dalla definizione di indipendenza stocastica, segue che X ed Y sono stocasticamente indipendenti *se e solo se*

$$f(x, y) = f_X(x)f_Y(y), \quad (4.3)$$

ovvero la densità congiunta è uguale al prodotto delle densità marginali. Come nel caso delle distribuzioni discrete, si ha che se X, Y sono stocasticamente indipendenti e $\phi_1, \phi_2 : \mathbb{R} \rightarrow \mathbb{R}$, allora

$$\mathbf{P}(\phi_1(X)\phi_2(Y)) = \mathbf{P}(\phi_1(X))\mathbf{P}(\phi_2(Y)),$$

se $\phi_1(x)\phi_2(y)f_X(x)f_Y(y)$ è integrabile. La dimostrazione è analoga al caso discreto e segue da (4.2) e (4.3).

4.4 La densità di $Z = X + Y$

Siano X ed Y due numeri aleatori con densità congiunta $f(x, y)$. Si vuole calcolare la densità di

$$Z = X + Y.$$

Si calcola la funzione di ripartizione

$$\begin{aligned} F_Z(z) &= \mathbf{P}(Z \leq z) \\ &= \mathbf{P}(X + Y \leq z) \\ &= \int_{-\infty}^{+\infty} \int_{-\infty}^{z-x} f(x, y) dy dx \\ &= \int_{-\infty}^{+\infty} dx \int_{-\infty}^z f(x, t-x) dt \\ &= \int_{-\infty}^z dt \int_{-\infty}^{+\infty} f(x, t-x) dx, \end{aligned}$$

dove si è effettuato il cambio coordinate $x = x$ e $t = x + y$ con corrispondente determinante jacobiano $|\det J| = 1$. Derivando si ottiene che la densità di Z è

$$f_Z(z) = F'_Z(z) = \int_{\mathbb{R}} f(x, z-x) dx.$$

In particolare, se $f(x, y) = f_X(x)f_Y(y)$, allora

$$f_Z(z) = \int_{\mathbb{R}} f_X(x)f_Y(z-x) dx.$$

Come esempio particolare di questa formula, si considerino due numeri aleatori X ed Y stocasticamente indipendenti e, rispettivamente, di densità $\Gamma(\alpha, \lambda)$ e

$\Gamma(\beta, \lambda)$ (ovvero densità gamma di parametri α, λ e β, λ). La densità congiunta di (X, Y) è data da

$$f(x, y) = f_X(x)f_Y(y).$$

Utilizzando la formula precedente, si calcola la densità di $Z = X + Y$.

$$\begin{aligned} f_Z(z) &= \int_{-\infty}^{+\infty} f(x, z-x) dx \\ &= \int_{-\infty}^{+\infty} f_X(x)f_Y(z-x) dx \\ &= \int_{-\infty}^{+\infty} \frac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\lambda x} I_{\{x>0\}} \frac{\lambda^\beta}{\Gamma(\beta)} (z-x)^{\beta-1} e^{-\lambda(z-x)} I_{\{(z-x)>0\}} dx \\ &= \frac{\lambda^{\alpha+\beta}}{\Gamma(\alpha)\Gamma(\beta)} e^{-\lambda z} \int_{-\infty}^{+\infty} x^{\alpha-1} (z-x)^{\beta-1} I_{\{x>0, (z-x)>0\}} dx. \end{aligned}$$

Dato $A \subset \mathbb{R}^n$, si ricordi che la funzione $I_A(x)$ si chiama *funzione indicatrice* ed è definita nel seguente modo

$$I_A(x) = \begin{cases} 0 & x \notin A \\ 1 & x \in A. \end{cases}$$

Si ottiene dunque

- Se $z \leq 0$, allora $f_Z(z) = 0$;
- Se $z > 0$ allora $0 < x < z$ e

$$\begin{aligned} f_Z(z) &= \frac{\lambda^{\alpha+\beta}}{\Gamma(\alpha)\Gamma(\beta)} e^{-\lambda z} \int_0^z x^{\alpha-1} (z-x)^{\beta-1} dx \\ &= \frac{\lambda^{\alpha+\beta}}{\Gamma(\alpha)\Gamma(\beta)} e^{-\lambda z} \int_0^1 (zt)^{\alpha-1} (z-zt)^{\beta-1} z dt \\ &= \frac{\lambda^{\alpha+\beta}}{\Gamma(\alpha)\Gamma(\beta)} z^{\alpha+\beta-1} e^{-\lambda z} \int_0^1 t^{\alpha-1} (1-t)^{\beta-1} dt \\ &= \frac{\lambda^{\alpha+\beta}}{\Gamma(\alpha)\Gamma(\beta)} \int_0^1 t^{\alpha-1} (1-t)^{\beta-1} dt z^{\alpha+\beta-1} e^{-\lambda z} \\ &= K z^{\alpha+\beta-1} e^{-\lambda z}. \end{aligned}$$

dove si è effettuato il cambio di variabili $x = zt$, $dx = zdt$ e si è posto

$$K = \frac{\lambda^{\alpha+\beta}}{\Gamma(\alpha)\Gamma(\beta)} \int_0^1 t^{\alpha-1} (1-t)^{\beta-1} dt. \quad (4.4)$$

Ne segue che Z ha distribuzione $\Gamma(\alpha+\beta, \lambda)$. In particolare, se X_1, \dots, X_n sono stocasticamente indipendenti con distribuzione esponenziale di parametro λ , allora

$$Z = X_1 + \cdots + X_n \sim \Gamma(1, \lambda) + \cdots + \Gamma(1, \lambda) = \Gamma(n, \lambda).$$

Inoltre, se X_1, \dots, X_n sono stocasticamente indipendenti con distribuzione chi-quadro di parametro 1, ovvero $X_i \sim \Gamma(\frac{1}{2}, \frac{1}{2})$ per ogni $i = 1, \dots, n$, allora

$$Z_n = X_1 + \cdots + X_n \sim \Gamma(\frac{n}{2}, \frac{1}{2})$$

si dice avere *distribuzione chi-quadro di parametro n* , che si indica con $\chi^2(n)$.

Osservazione 4.4.1. Da questa dimostrazione si ottiene inoltre il valore dell'integrale $\int_0^1 t^{\alpha-1} (1-t)^{\beta-1} dt$. Infatti poiché la costante di normalizzazione della distribuzione Gamma $\Gamma(\alpha + \beta, \lambda)$ deve essere

$$K = \frac{\lambda^{\alpha+\beta}}{\Gamma(\alpha + \beta)},$$

da (4.4) si ottiene che

$$\int_0^1 t^{\alpha-1} (1-t)^{\beta-1} dt = \frac{\Gamma(\alpha) \Gamma(\beta)}{\Gamma(\alpha + \beta)}.$$

4.5 La distribuzione beta $\mathcal{B}(\alpha, \beta)$

Siano $\alpha, \beta > 0$. Un numero aleatorio X ha *distribuzione beta* $\mathcal{B}(\alpha, \beta)$ se ha densità

$$f(x) = \begin{cases} K x^{\alpha-1} (1-x)^{\beta-1} & x \in [0, 1] \\ 0 & \text{altrimenti.} \end{cases}$$

Dai calcoli svolti nella sezione precedente, si ottiene immediatamente la costante di normalizzazione

$$K = \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha) \Gamma(\beta)} = \frac{1}{\int_0^1 x^{\alpha-1} (1-x)^{\beta-1} dx}.$$

Per la previsione si usano le proprietà della funzione Gamma di Eulero:

$$\begin{aligned} \mathbf{P}(X) &= \int_0^1 x f(x) dx \\ &= \int_0^1 \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha) \Gamma(\beta)} x^\alpha (1-x)^{\beta-1} dx \\ &= \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha) \Gamma(\beta)} \frac{\Gamma(\alpha + 1) \Gamma(\beta)}{\Gamma(\alpha + \beta + 1)} \\ &= \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha) \Gamma(\beta)} \frac{\alpha \Gamma(\alpha) \Gamma(\beta)}{(\alpha + \beta) \Gamma(\alpha + \beta)} \\ &= \frac{\alpha}{(\alpha + \beta)}. \end{aligned}$$

Per calcolare la varianza bisogna prima fare il calcolo della previsione quadratica $\mathbf{P}(X^2)$ come segue

$$\begin{aligned}\mathbf{P}(X^2) &= \int_0^1 x^2 f(x) dx \\ &= \int_0^1 \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha) \Gamma(\beta)} x^{\alpha+1} (1-x)^{\beta-1} dx \\ &= \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha) \Gamma(\beta)} \frac{\Gamma(\alpha + 2) \Gamma(\beta)}{\Gamma(\alpha + \beta + 2)} \\ &= \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha) \Gamma(\beta)} \frac{(\alpha + 1) \alpha \Gamma(\alpha) \Gamma(\beta)}{(\alpha + \beta + 1) (\alpha + \beta) \Gamma(\alpha + \beta)} \\ &= \frac{(\alpha + 1) \alpha}{(\alpha + \beta + 1) (\alpha + \beta)}.\end{aligned}$$

Si ottiene

$$\begin{aligned}\sigma^2(X) &= \mathbf{P}(X^2) - \mathbf{P}(X)^2 \\ &= \frac{(\alpha + 1) \alpha}{(\alpha + \beta + 1) (\alpha + \beta)} - \frac{\alpha^2}{(\alpha + \beta)^2} \\ &= \frac{\alpha \beta}{(\alpha + \beta)^2 (\alpha + \beta + 1)}.\end{aligned}$$

4.6 La distribuzione di Student

Vediamo ora un'altra distribuzione utile in statistica, la *distribuzione di Student* di parametro ν . Consideriamo due numeri aleatori stocasticamente indipendenti Z ed U , dove Z ha distribuzione normale standard ed U distribuzione gamma di parametri $\frac{\nu}{2}, \frac{1}{2}$ (ovvero χ^2 di parametro ν con $\nu \in \mathbb{N}$). Sia $\nu > 0$

e consideriamo il numero aleatorio $T = \frac{Z}{\sqrt{\frac{U}{\nu}}}$. Per ottenere la densità di T ,

calcoliamo la funzione di ripartizione di T utilizzando la densità congiunta $f(z, u)$ di Z, U :

$$F_T(t) = \mathbf{P}(T \leq t) = \mathbf{P}(Z \leq t \sqrt{\frac{U}{\nu}}) = \int_0^\infty \int_{-\infty}^{\sqrt{\frac{u}{\nu}} t} f(z, u) dz du,$$

dove

$$f(z, u) = \frac{1}{2^{\frac{\nu}{2}} \sqrt{2\pi} \Gamma(\frac{\nu}{2})} e^{-\frac{z^2}{2}} u^{\frac{\nu}{2}-1} e^{-\frac{u}{2}}, \quad u > 0, z \in \mathbb{R},$$

0 altrimenti. Derivando $F_T(t)$, dal teorema fondamentale del calcolo integrale si ha che la densità della distribuzione di Student è pari a

$$\begin{aligned}
 f_T(t) &= \int_0^\infty f\left(t\sqrt{\frac{u}{\nu}}, u\right) \sqrt{\frac{u}{\nu}} du = \\
 &= \frac{1}{2^{\frac{\nu}{2}} \sqrt{2\pi\nu} \Gamma\left(\frac{\nu}{2}\right)} \int_0^\infty u^{\frac{\nu+1}{2}-1} e^{-\frac{u}{2}\left(1+\frac{t^2}{\nu}\right)} du = \\
 &= \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\pi\nu} \Gamma\left(\frac{\nu}{2}\right)} \frac{1}{\left(1+\frac{t^2}{\nu}\right)^{\frac{\nu+1}{2}}}.
 \end{aligned}$$

Si osserva che per $\nu = 1$ si ottiene la distribuzione di Cauchy.

Vediamo ora per quali valori del parametro ν esiste la previsione della corrispondente distribuzione di Student. Poiché l'integrale

$$\int_{\mathbb{R}} \frac{t}{(1+t^2)^{\frac{\nu+1}{2}}} dt$$

deve essere finito affinché $\mathbf{P}(T) < \infty$, si ha che la previsione di T esiste finita per $\nu > 1$. In tal caso

$$\mathbf{P}(T) = \mathbf{P}\left(\frac{Z}{\sqrt{\frac{U}{\nu}}}\right) = \mathbf{P}(Z) \mathbf{P}\left(\frac{1}{\sqrt{\frac{U}{\nu}}}\right) = 0,$$

in quanto Z, U sono stocasticamente indipendenti e $\mathbf{P}(Z) = 0$. Per quanto riguarda la varianza, basta calcolare la previsione quadratica

$$\begin{aligned}
 \sigma(T) &= \mathbf{P}(T^2) = \mathbf{P}\left(\frac{\nu Z^2}{U}\right) = \\
 &= \nu \mathbf{P}(Z^2) \mathbf{P}\left(\frac{1}{U}\right) = \nu \mathbf{P}\left(\frac{1}{U}\right) = \\
 &= \frac{\nu}{2^{\frac{\nu}{2}} \Gamma\left(\frac{\nu}{2}\right)} \int_0^\infty \frac{1}{u} u^{\frac{\nu}{2}-1} e^{-\frac{u}{2}} du = \\
 &= \frac{\nu}{2^{\frac{\nu}{2}} \Gamma\left(\frac{\nu}{2}\right)} \int_0^\infty u^{\frac{\nu-2}{2}-1} e^{-\frac{u}{2}} du = \\
 &= \frac{\nu}{2^{\frac{\nu}{2}} \Gamma\left(\frac{\nu}{2}\right)} 2^{\frac{\nu-2}{2}} \Gamma\left(\frac{\nu-2}{2}\right) = \\
 &= \frac{\nu}{\nu-2}.
 \end{aligned}$$

Si osserva quindi che la varianza esiste finita se $\nu > 2$.

4.7 Distribuzioni n -dimensionali

Sia (X_1, X_2, \dots, X_n) un vettore aleatorio di dimensione n . La funzione

$$F : \mathbb{R}^n \longrightarrow [0, 1]$$

definita come

$$F(x_1, x_2, \dots, x_n) = \mathbf{P}(X_1 \leq x_1, X_2 \leq x_2, \dots, X_n \leq x_n)$$

si dice *funzione di ripartizione congiunta* di (X_1, X_2, \dots, X_n) . Si assumono come nel caso bidimensionale le seguenti ipotesi di continuità.

1. $\lim_{x_1 \rightarrow +\infty} F(x_1, x_2, \dots, x_n) = 1,$
- \vdots
2. $\lim_{x_i \rightarrow -\infty} F(x_1, x_2, \dots, x_n) = 0$

ed analoghe.

4.8 Distribuzioni assolutamente continue n -dimensionali

Il vettore aleatorio (X_1, X_2, \dots, X_n) ha distribuzione assolutamente continua se esiste una funzione

$$f : \mathbb{R}^n \longrightarrow \mathbb{R}$$

tale che

1. f sia *non negativa e integrabile*
2. $\int_{\mathbb{R}^n} f(x_1, \dots, x_n) dx_1 \dots dx_n = 1$

e vale che la funzione di ripartizione congiunta di (X_1, X_2, \dots, X_n) è data da

$$F(x_1, x_2, \dots, x_n) = \int_{-\infty}^{x_1} \int_{-\infty}^{x_2} \dots \int_{-\infty}^{x_n} f(t_1, t_2, \dots, t_n) dt_1 dt_2 \dots dt_n .$$

La funzione di densità marginale di $(X_{i_1}, X_{i_2}, \dots, X_{i_n})$ si calcola nel seguente modo

$$f_{X_{i_1}, \dots, X_{i_r}}(x_{i_1}, \dots, x_{i_r}) = \int_{\mathbb{R}^{n-r}} f(x_{i_1}, \dots, x_{i_r}, t_{i_{r+1}}, \dots, t_{i_n}) dt_{i_{r+1}} \dots dt_{i_n}$$

per ogni scelta di indici i_1, \dots, i_r in $\{1, \dots, n\}$.

4.9 Distribuzione gaussiana n -dimensionale

Un vettore aleatorio (X_1, X_2, \dots, X_n) ha *distribuzione gaussiana n -dimensionale* se ha densità

$$f(x_1, x_2, \dots, x_n) = K e^{-\frac{1}{2}Ax + b \cdot x}$$

dove $x = (x_1, x_2, \dots, x_n)^t \in \mathbb{R}^n$, A è una matrice

- *simmetrica*: $A^t = A$, ovvero $a_{ij} = a_{ji}$,

• *definita positiva*: $Ax \cdot x \geq 0, \forall x \in \mathbb{R}^n$, e $Ax \cdot x = 0$ implica che $x = 0$

e $b = (b_1, b_2, \dots, b_n)^t$ è un vettore in \mathbb{R}^n . Il simbolo A^t indica la *matrice trasposta*, ovvero di elementi

$$[A^t]_{i,j} = [A]_{j,i}.$$

Si ricordi inoltre che $b \cdot x$ indica il *prodotto scalare* fra il vettore b ed il vettore x , ovvero

$$b \cdot x = \sum_{i=1}^n b_i x_i,$$

mentre Ax è il vettore che si ottiene come prodotto della matrice A per il vettore x , le cui componenti sono

$$[Ax]_i = \sum_j a_{ij} x_j.$$

L'espressione $Ax \cdot x$ è una forma quadratica del tipo

$$Ax \cdot x = \sum_{i,j} a_{ij} x_i x_j.$$

Viceversa, se si parte da una forma quadratica

$$\sum_{i,j} \alpha_{ij} x_i x_j.$$

ci si può sempre ricondurre ad una rappresentazione matriciale associata ad una matrice simmetrica di componenti:

$$a_{ij} = \begin{cases} \alpha_{ii} & i = j \\ (\alpha_{ij} + \alpha_{ji})/2 & i \neq j \end{cases}.$$

Caso 1: A diagonale e $b = 0$

Siano

$$A = \begin{pmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & \lambda_n \end{pmatrix}$$

e $b = 0$. Si ottiene

$$f(x_1, x_2, \dots, x_n) = K \exp \left(- \left(\lambda_1 \frac{x_1^2}{2} + \lambda_2 \frac{x_2^2}{2} + \dots + \lambda_n \frac{x_n^2}{2} \right) \right),$$

ovvero

$$f(x_1, x_2, \dots, x_n) = f_{X_1}(x_1) f_{X_2}(x_2) \cdots f_{X_n}(x_n)$$

dove

$$f_{X_i}(x_i) = \sqrt{\frac{\lambda_i}{2\pi}} e^{-\frac{\lambda_i x_i^2}{2}}$$

è la densità marginale di X_i . Ne segue che:

1. X_1, \dots, X_n sono stocasticamente indipendenti.
2. Ogni X_i ha densità gaussiana $N\left(0, \frac{1}{\lambda_i}\right)$.
3. La costante di normalizzazione è data da

$$K = \sqrt{\frac{\lambda_1}{2\pi}} \sqrt{\frac{\lambda_2}{2\pi}} \cdots \sqrt{\frac{\lambda_n}{2\pi}} = \sqrt{\frac{\det A}{(2\pi)^n}}.$$

Il *vettore delle previsioni* di (X_1, X_2, \dots, X_n) è allora

$$(\mathbf{P}(X_1), \mathbf{P}(X_2), \dots, \mathbf{P}(X_n)) = (0, 0, \dots, 0)$$

e la *matrice di covarianza*

$$\begin{aligned}
 C &= \begin{pmatrix} \sigma^2(X_1) & \mathbf{cov}(X_1, X_2) & \cdots & \mathbf{cov}(X_1, X_n) \\ \mathbf{cov}(X_2, X_1) & \sigma^2(X_2) & \ddots & \vdots \\ \vdots & \ddots & \ddots & \mathbf{cov}(X_{n-1}, X_n) \\ \mathbf{cov}(X_n, X_1) & \cdots & \mathbf{cov}(X_n, X_{n-1}) & \sigma^2(X_n) \end{pmatrix} \\
 &= \begin{pmatrix} \frac{1}{\lambda_1} & 0 & \cdots & 0 \\ 0 & \frac{1}{\lambda_2} & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & \frac{1}{\lambda_n} \end{pmatrix} \\
 &= A^{-1}.
 \end{aligned}$$

Caso 2: Caso $b \neq 0$

Per ricondursi al caso $b = 0$ si utilizza la *traslazione* $X = U + c$ di componenti $[X]_i = [U]_i + [c]_i$, dove c è un vettore di \mathbb{R}^n . Si osserva che la funzione di ripartizione di U è data da quella di X calcolata nel punto $u + c$:

$$F_U(u) = \mathbf{P}(U \leq u) = \mathbf{P}(X - c \leq u) = \mathbf{P}(X \leq u + c) = F_X(u + c),$$

da cui derivando si ottiene che la densità congiunta del vettore aleatorio U è data dalla densità congiunta del vettore aleatorio X calcolata nel punto $u + c$. Si ottiene

$$\begin{aligned} f_U(u_1, u_2, \dots, u_n) &= f_X(u_1 + c_1, u_2 + c_2, \dots, u_n + c_n) \\ &= K' \exp \left[-\frac{1}{2} A(u + c) \cdot (u + c) + b \cdot (u + c) \right] \\ &= K' \exp \left(-\frac{1}{2} Au \cdot u - \frac{1}{2} Au \cdot c - \frac{1}{2} Ac \cdot u - \frac{1}{2} Ac \cdot c + b \cdot u + b \cdot c \right) \\ &= K' \underbrace{\exp \left(-\frac{1}{2} Ac \cdot c + b \cdot c \right)}_{\text{costante}} \exp \left(-\frac{1}{2} Au \cdot u + (b - Ac) \cdot u \right), \end{aligned}$$

in quanto

$$Ac \cdot u = Au \cdot c$$

perché A è simmetrica. Per ricondursi al caso precedente, bisogna annullare la parte di primo grado in U , quindi si sceglie

$$b - Ac = 0 \quad \Rightarrow \quad c = A^{-1}b$$

(A è invertibile in quanto definita positiva). Per tale scelta di c la densità $f_U(u_1, u_2, \dots, u_n)$ è data da

$$\begin{aligned} f_U(u_1, u_2, \dots, u_n) &= f_X(u_1 + c_1, u_2 + c_2, \dots, u_n + c_n) \\ &= K' \exp \left(A^{-1}b \cdot b - \frac{A(A^{-1}b) \cdot A^{-1}b}{2} \right) \exp \left(-\frac{1}{2} Au \cdot u \right) \\ &= K' \underbrace{\exp \left(\frac{1}{2} A^{-1}b \cdot b \right)}_{K''} \exp \left(-\frac{1}{2} Au \cdot u \right) \\ &= K'' \exp \left(-\frac{1}{2} Au \cdot u \right). \end{aligned}$$

Si vede facilmente che $\mathbf{P}(U_i) = 0$ per ogni i dato che la densità di $-U$ è uguale a quella di U . Usando i risultati precedenti, si ottiene

1. la previsione $\mathbf{P}(X_i) = \mathbf{P}(U_i + c_i) = P(U_i) + c_i = 0 + c_i = (A^{-1}b)_i$, ovvero in notazione vettoriale:

$$\mathbf{P}(X) = A^{-1}b.$$

2. La costante di normalizzazione è

$$K' = \sqrt{\frac{\det A}{(2\pi)^n}} e^{-\frac{1}{2} A^{-1}b \cdot b},$$

in quanto $K'' = \sqrt{\frac{\det A}{(2\pi)^n}}$.

3. La matrice di covarianza di X è la stessa di U in quanto una traslazione lascia invariate le covarianze.

$$\begin{aligned}
 C &= \begin{pmatrix} \sigma^2(X_1) & \mathbf{cov}(X_1, X_2) & \cdots & \mathbf{cov}(X_1, X_n) \\ \mathbf{cov}(X_2, X_1) & \sigma^2(X_2) & \ddots & \vdots \\ \vdots & \ddots & \ddots & \mathbf{cov}(X_{n-1}, X_n) \\ \mathbf{cov}(X_n, X_1) & \cdots & \mathbf{cov}(X_n, X_{n-1}) & \sigma^2(X_n) \end{pmatrix} \\
 &= \begin{pmatrix} \sigma^2(U_1) & \mathbf{cov}(U_1, U_2) & \cdots & \mathbf{cov}(U_1, U_n) \\ \mathbf{cov}(U_2, U_1) & \sigma^2(U_2) & \ddots & \vdots \\ \vdots & \ddots & \ddots & \mathbf{cov}(U_{n-1}, U_n) \\ \mathbf{cov}(U_n, U_1) & \cdots & \mathbf{cov}(U_n, U_{n-1}) & \sigma^2(U_n) \end{pmatrix}.
 \end{aligned}$$

Ne segue che

$$C = A^{-1}.$$

Caso 3: $b \neq 0$

Se A non è diagonale, le X_i non sono più stocasticamente indipendenti perciò la densità congiunta non è data dal prodotto delle densità marginali. A meno di fare una traslazione, si può pensare di avere $b = 0$. Poiché A è simmetrica, esiste una trasformazione ortogonale O che diagonalizza A , ovvero tale $O^t A O = D$, dove D è diagonale e $O^t = O^{-1}$.

Se si considera la trasformazione $X = OU$ si ottiene

$$\begin{aligned}
 f(u_1, \dots, u_n) &= K \exp\left(-\frac{1}{2} A O u \cdot O u\right) \\
 &= K \exp\left(-\frac{1}{2} O^t A O u \cdot u\right) \\
 &= K \exp\left(-\frac{1}{2} D u \cdot u\right).
 \end{aligned}$$

Ci siamo ricondotti al primo caso, in cui si aveva A diagonale e $b = 0$. La matrice di covarianza di X , in notazione multidimensionale, è data da

$$\begin{aligned}
 C &= \mathbf{P}(X X^t) = \mathbf{P}(O U (O U)^t) \\
 &= O \mathbf{P}(U U^t) O^t = O D^{-1} O^t = A^{-1}.
 \end{aligned}$$

Si è usata la proprietà, facilmente verificabile, che, se Z è una matrice aleatoria ed A, B sono matrici costanti, si ha che $\mathbf{P}(AZB) = \mathbf{A}\mathbf{P}(Z)\mathbf{B}$ (ovviamente si suppone che sia possibile effettuare il prodotto).

Riassumendo nel caso in A sia non diagonale e $b \neq 0$, si ottiene:

1. *Costante di normalizzazione*

$$K = \sqrt{\frac{\det A}{(2\pi)^n}} e^{-\frac{1}{2}A^{-1}b \cdot b}.$$

2. *previsione*

$$\mathbf{P}(X) = A^{-1}b.$$

3. *matrice di covarianza*

$$C = A^{-1}.$$

Osservazione 4.9.1. Anche in questo caso, le distribuzioni marginali delle X_i o di sottosinsiemi delle X_i sono gaussiane $N(\mathbf{P}(X_i), \sigma^2(X_i))$ nel caso unidimensionale ed analoghe nel caso a più dimensioni. Per la distribuzione gaussiana, $\mathbf{cov}(X_i, X_j) = 0$ implica l'indipendenza stocastica di X_i e X_j .

Osservazione 4.9.2. Nel caso particolare in cui $n = 2$, la matrice di covarianza è data da:

$$C = \begin{pmatrix} \sigma_1^2 & \rho \sigma_1 \sigma_2 \\ \rho \sigma_1 \sigma_2 & \sigma_2^2 \end{pmatrix}$$

dove con ρ si indica il coefficiente di correlazione. Si può quindi ricavare la matrice A dalla matrice di covarianza C nel seguente modo:

$$\begin{aligned} A = C^{-1} &= \frac{1}{\det C} \begin{pmatrix} \sigma_2^2 & -\rho \sigma_1 \sigma_2 \\ -\rho \sigma_1 \sigma_2 & \sigma_1^2 \end{pmatrix} \\ &= \frac{1}{\sigma_1^2 \sigma_2^2 - \rho^2 \sigma_1^2 \sigma_2^2} \begin{pmatrix} \sigma_2^2 & -\rho \sigma_1 \sigma_2 \\ -\rho \sigma_1 \sigma_2 & \sigma_1^2 \end{pmatrix} \\ &= \frac{1}{1 - \rho^2} \begin{pmatrix} \frac{1}{\sigma_1^2} & -\frac{\rho}{\sigma_1 \sigma_2} \\ -\frac{\rho}{\sigma_1 \sigma_2} & \frac{1}{\sigma_2^2} \end{pmatrix}. \end{aligned}$$

Quindi, la densità gaussiana bidimensionale con parametri $m_1, m_2, \sigma_1, \sigma_2, \rho$ è data da:

$$f(x, y) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \cdot \exp\left(-\frac{1}{2(1-\rho^2)}\left(\frac{(x-m_1)^2}{\sigma_1^2} - 2\frac{\rho(x-m_1)(y-m_2)}{\sigma_1\sigma_2} + \frac{(y-m_2)^2}{\sigma_2^2}\right)\right).$$

Convergenza di distribuzioni

5.1 Convergenza di funzioni di ripartizione

È naturale introdurre una nozione di convergenza per successioni di funzioni di ripartizione. Si vuole quindi capire quando ed in che senso $F_n \rightarrow F$. Definire questa convergenza come convergenza puntuale, se cioè $F_n(x) \rightarrow F(x)$ è troppo restrittivo. Consideriamo ad esempio la successione $F_n(x)$ definita come

$$F_n(x) = \begin{cases} 1 & \text{per } x \geq \frac{1}{n} \\ 0 & \text{per } x < 0. \end{cases}$$

Se X_n ha come la funzione di ripartizione F_n si ha che $\mathbf{P}(X_n = \frac{1}{n}) = 1$. È naturale che per una ragionevole nozione di convergenza si abbia $F_n \rightarrow F$, dove

$$F(x) = \begin{cases} 1 & \text{per } x \geq 0 \\ 0 & \text{per } x < 0. \end{cases}$$

Si vede subito però che non è vero che $F_n(x) \rightarrow F(x)$ per ogni $x \in \mathbb{R}$. Infatti $F_n(0) = 0$ per ogni n , mentre $F(0) = 1$. Si considera allora una definizione più debole.

Definizione 5.1.1. *Diremo che $F_n \rightarrow F$ se per ogni $\epsilon > 0$ esiste N tale che per $n \geq N$ si ha che*

$$F(x - \epsilon) - \epsilon < F_n(x) < F(x + \epsilon) + \epsilon.$$

Se x è un punto di continuità per F , questa definizione implica

$$\lim_{n \rightarrow \infty} F_n(x) = F(x).$$

Viceversa se per ogni punto x di continuità per F si ha che $\lim_{n \rightarrow \infty} F_n(x) = F(x)$, allora si ha che $F_n \rightarrow F$. Per una funzione di ripartizione infatti i

punti di continuità sono ovunque densi dato che i punti di discontinuità sono numerabili (non si possono avere più di n punti di discontinuità con salto maggiore od uguale a $\frac{1}{n}$). Sia quindi $x \in \mathbb{R}$ ed $\epsilon > 0$. Esistono quindi due punti di continuità x_0, x_1 di F tali che $x - \epsilon < x_0 < x < x_1 < x + \epsilon$. Abbiamo quindi

$$F(x - \epsilon) \leq \lim_{n \rightarrow \infty} F_n(x_0) = F(x_0)$$

ed anche

$$\lim_{n \rightarrow \infty} F_n(x_1) = F(x_1) \leq F(x + \epsilon).$$

D'altra parte per ogni n si ha che

$$F_n(x_0) \leq F_n(x) \leq F_n(x_1).$$

Quindi per n abbastanza grande avremo che

$$F(x - \epsilon) - \epsilon < F_n(x) < F(x + \epsilon) + \epsilon.$$

È facile costruire funzioni di ripartizione assolutamente continue F_n che convergono ad una funzione di ripartizione discreta. Ad esempio se $F_n(x) = \mathcal{N}(\sqrt{n}x)$, funzione di ripartizione della distribuzione normale con previsione 0 e varianza n , si ha che $F_n \rightarrow F$ con

$$F(x) = \begin{cases} 1 & \text{per } x \geq 0 \\ 0 & \text{per } x < 0. \end{cases}$$

Viceversa si possono costruire esempi di successioni di funzioni di ripartizione discrete che convergono ad una funzione di ripartizione assolutamente continua. Ad esempio se

$$F_n(x) = \begin{cases} 0 & \text{per } x \leq 0 \\ \frac{[nx]}{n} & \text{per } 0 < x \leq 1 \\ 1 & \text{per } x > 1, \end{cases}$$

si ha che $F_n \rightarrow F$ con F la funzione di ripartizione della distribuzione uniforme in $[0, 1]$

$$F(x) = \begin{cases} 0 & \text{per } x \leq 0 \\ x & \text{per } 0 < x \leq 1 \\ 1 & \text{per } x > 1. \end{cases}$$

5.2 Convergenza della distribuzione geometrica a quella esponenziale

In questa sezione vedremo in particolare la convergenza della distribuzione geometrica a quella esponenziale. Le due distribuzioni di probabilità condividono la stessa proprietà di assenza di memoria. Consideriamo una successione $(X_n)_{n \in \mathbb{N}}$ di numeri aleatori con distribuzione geometrica ciascuno di parametro p_n

$$\mathbf{P}(X_n = k) = p_n(1 - p_n)^{k-1}, \quad \forall k \geq 1,$$

e tale che $np_n \rightarrow \lambda$, $\lambda > 0$, se $n \rightarrow \infty$. Posto $Y_n = \frac{X_n}{n}$, si ottiene che

$$F_{Y_n} \rightarrow F,$$

dove F è la funzione di ripartizione della distribuzione esponenziale di parametro λ , ovvero

$$F(x) = \begin{cases} 0 & \text{per } x < 0 \\ 1 - e^{-\lambda x} & \text{per } x \geq 0. \end{cases}$$

Basta calcolare la funzione $F_{Y_n}(x) = \mathbf{P}(Y_n \leq x)$ di ripartizione di Y_n . Per $x < 0$, si ha subito che $F_{Y_n} \equiv 0$. Per $x \geq 0$

$$\begin{aligned} F_{Y_n}(x) &= \mathbf{P}(Y_n \leq x) = \mathbf{P}(X_n \leq nx) = \\ &1 - \sum_{k=[nx]+1}^{\infty} p_n(1 - p_n)^{k-1} = 1 - p_n(1 - p_n)^{[nx]} \sum_{i=0}^{\infty} (1 - p_n)^i = \\ &1 - p_n(1 - p_n)^{[nx]} \frac{1}{1 - (1 - p_n)} = 1 - (1 - p_n)^{[nx]}, \end{aligned}$$

dove si è usata la somma della serie geometrica. Si osserva che ogni numero reale si può scrivere come la somma della sua parte intera e di un numero compreso fra 0 e 1, ovvero

$$nx = [nx] + \delta_n, \quad 0 \leq \delta_n < 1.$$

Quindi si ottiene che

$$F_{Y_n}(x) = 1 - (1 - p_n)^{nx - \delta_n},$$

che tende a $1 - e^{-\lambda x}$ per $x \geq 0$, in quanto

$$\log(1 - p_n)^{nx} = nx \log(1 - p_n) = -xnp_n + o(np_n)$$

tende a $-\lambda$ per $n \rightarrow \infty$.

5.3 Convergenza della distribuzione binomiale a quella di Poisson

Vediamo ora come si può approssimare la distribuzione binomiale quando si considera il numero di successi su un numero molto grande di eventi. Consideriamo una successione $(X_n)_{n \in \mathbb{N}}$ di numeri aleatori con distribuzione binomiale di parametri n, p_n tali che $np_n \rightarrow \lambda$ con $\lambda > 0$. Si può rappresentare X_n come il numero di successi in n prove di Bernoulli di parametro p_n . A

crescere del numero delle prove si manda qui a zero la probabilità di successo in ogni prova. Si ha per $0 \leq k \leq n$

$$\begin{aligned} \mathbf{P}(X_n = k) &= \binom{n}{k} p_n^k (1 - p_n)^{n-k} = \\ &= \frac{n!}{k!(n-k)!} p_n^k (1 - p_n)^{n-k} \frac{n^k}{n^k} = \\ &\quad \underbrace{\frac{n!}{k!(n-k)!} p_n^k (1 - p_n)^{n-k}}_{\text{si moltiplica e divide per } n^k} = \\ &= \frac{1}{k!} \left(1 - \frac{1}{n}\right) \cdots \left(1 - \frac{k-1}{n}\right) (np_n)^k (1 - p_n)^{n-k}. \end{aligned}$$

Basta quindi osservare che

- $\left(1 - \frac{1}{n}\right) \cdots \left(1 - \frac{k-1}{n}\right)$ tende a 1 per $n \rightarrow \infty$;
- $(np_n)^k$ tende a λ^k per $n \rightarrow \infty$;
- $(1 - p_n)^{-k}$ tende a 1 per $n \rightarrow \infty$;
- $(1 - p_n)^n$ tende a $e^{-\lambda}$ per $n \rightarrow \infty$ in quanto

$$\log(1 - p_n)^n = n \log(1 - p_n) = -np_n + o(np_n)$$

tende a $-\lambda$.

Si conclude quindi che per ogni $k \in \mathbb{N}$

$$\mathbf{P}(X_n = k) \xrightarrow{n \rightarrow \infty} \frac{\lambda^k}{k!} e^{-\lambda},$$

da cui segue immediatamente che la successione di distribuzioni binomiali di parametri n, p_n tende alla distribuzione di Poisson di parametro λ .

5.4 Il teorema di De Moivre-Laplace

Vediamo ora un altro tipo di convergenza per successioni di distribuzioni binomiali. Questa volta mandiamo il numero delle prove n all'infinito mantenendo costante la probabilità di successo in ogni singola prova. Operiamo però un riscalamento lineare per ottenere un limite.

Theorem 5.4.1. *Sia $(X_i)_{i \in \mathbb{N}}$ una successione di numeri aleatori con distribuzione binomiale di parametri rispettivamente $\mathcal{B}(n, p)$. Dati i numeri aleatori standardizzati*

$$X_n^* = \frac{X_n - \mathbf{P}(X_n)}{\sigma(X_n)} = \frac{X_n - np}{\sqrt{np\tilde{p}}}$$

dove si è posto $\tilde{p} = 1 - p$, vale che

$$\mathbf{P}(X_n^* = x) = \frac{h_n}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} e^{E_n(x)},$$

dove $h_n = \frac{1}{\sqrt{np\tilde{p}}}$ ed $E_n(x)$ è l'errore che tende a zero uniformemente se x è limitato.

Dimostrazione. Se l'insieme dei valori possibili di X_n è $I(X_n) = \{0, 1, \dots, n\}$, si ottiene che

$$I(X_n^*) = \{h_n(-np), h_n(1 - np), \dots, h_n(n - np)\}$$

dove $h_n = \frac{1}{\sqrt{np\tilde{p}}}$ è la spaziatura dei valori di X_n^* .

Definiamo $\phi_n(x) = \log \mathbf{P}(X_n^* = x)$ e consideriamo il suo rapporto incrementale:

$$\frac{\phi_n(x + h_n) - \phi_n(x)}{h_n} = \frac{1}{h_n} \log \frac{\mathbf{P}(X_n^* = x + h_n)}{\mathbf{P}(X_n^* = x)}.$$

Posto $k = np + x\sqrt{np\tilde{p}}$, si ottiene

$$\begin{aligned} \frac{1}{h_n} \log \frac{\mathbf{P}(X_n^* = x + h_n)}{\mathbf{P}(X_n^* = x)} &= \\ \frac{1}{h_n} \log \frac{(n - k)p}{k + 1} \frac{p}{\tilde{p}} &= \\ \sqrt{np\tilde{p}} \log \frac{n\tilde{p} - x\sqrt{np\tilde{p}}}{np + 1 + x\sqrt{np\tilde{p}}} \frac{p}{\tilde{p}} &= \\ \sqrt{np\tilde{p}} \log \frac{1 - x\sqrt{\frac{p}{n\tilde{p}}}}{1 + \frac{1}{np} + x\sqrt{\frac{\tilde{p}}{np}}} &. \end{aligned}$$

Se n tende all'infinito, si può usare l'approssimazione $\log(1 + x) = x + O(x^2)$. Si ottiene

$$\begin{aligned} & \sqrt{np\tilde{p}} \log \frac{1 - x\sqrt{\frac{p}{n\tilde{p}}}}{1 + \frac{1}{np} + x\sqrt{\frac{\tilde{p}}{np}}} = \\ & \sqrt{np\tilde{p}} \left[-x\sqrt{\frac{p}{n\tilde{p}}} + O\left(\frac{x^2}{n}\right) - x\sqrt{\frac{p}{n\tilde{p}}} + O\left(\frac{x^2+1}{n}\right) \right] = \\ & -xp - x\tilde{p} + O\left(\frac{x^2+1}{\sqrt{n}}\right) = \\ & -x + O\left(\frac{x^2+1}{\sqrt{n}}\right). \end{aligned}$$

La funzione $\phi_n(x)$ non è definita ovunque, ma solo per i valori di x tali che $\mathbf{P}(X_n^* = x) \neq 0$. La possiamo estendere per interpolazione lineare ai valori intermedi. Possiamo quindi scrivere:

$$\phi_n(x) = \phi_n(0) + \int_0^x \phi'_n(y) dy.$$

Se $x \leq y \leq x + h_n$, $\phi'_n(y) = \Delta_{h_n} \phi_n(x) = -x + O\left(\frac{x^2+1}{\sqrt{n}}\right) = -y + O\left(\frac{x^2+1}{\sqrt{n}}\right)$

da cui:

$$\begin{aligned} \phi_n(x) &= \phi_n(0) + \int_0^x \phi'_n(y) dy = \\ & \phi_n(0) + \int_0^x (-y) dy + O\left(\frac{x^3+x}{\sqrt{n}}\right) = \\ & \phi_n(0) - \frac{x^2}{2} + O\left(\frac{x^3+x}{\sqrt{n}}\right). \end{aligned}$$

Poiché $\phi_n(x) = \log \mathbf{P}(X_n^* = x)$, si ottiene

$$\log \mathbf{P}(X_n^* = x) = e^{\phi_n(0)} e^{-\frac{x^2}{2}} e^{E_n(x)}$$

dove $E_n(x) = O\left(\frac{x^3+x}{\sqrt{n}}\right)$.

Stimiamo $e^{\phi_n(0)}$ nel seguente modo. X_n^* è un numero aleatorio standardizzato, quindi con $\mathbf{P}(X_n^* = 0) = 0$ e $\sigma^2(X_n^*) = 1$. Dalla disuguaglianza di Chebychev, si ha che:

$$\mathbf{P}(|X_n^*| \geq K) \leq \frac{1}{K^2}.$$

Si può scegliere K in modo da rendere questa probabilità arbitrariamente piccola, ovvero per ogni $\epsilon > 0$ esiste K tale che

$$1 - \frac{1}{K^2} \leq \mathbf{P}(|X_n^*| < K) \leq 1$$

o anche

$$1 - \epsilon \leq \mathbf{P}(|X_n^*| < K) \leq 1.$$

Poiché $\mathbf{P}(|X_n^*| < K) = \sum_{x, |x| < K} \mathbf{P}(|X_n^*| = x)$, ne segue

$$1 - \epsilon \leq \sum_{x, |x| < K} \mathbf{P}(|X_n^*| = x) \leq 1.$$

$\mathbf{P}(|X_n^*| < K) = \sum_{x, |x| < K} \mathbf{P}(|X_n^*| = x) = \sum_{x, |x| < K} h_n e^{-\frac{x^2}{2}}$ è la somma di Riemann relativa alla funzione $e^{-\frac{x^2}{2}}$, quindi tende a $\int_{-K}^K e^{-\frac{x^2}{2}} dx$ per $n \rightarrow \infty$. Ne segue che

$$1 - \epsilon \leq \frac{e^{\phi_n(0)}}{h_n} \int_{-K}^K e^{-\frac{x^2}{2}} dx \leq 1.$$

Facendo tendere K all'infinito, si ottiene

$$1 - \epsilon \leq \frac{e^{\phi_n(0)}}{h_n} \sqrt{2\pi} \leq 1$$

ovvero

$$\frac{e^{\phi_n(0)}}{h_n} \sqrt{2\pi} \xrightarrow{n \rightarrow \infty} 1.$$

Ne segue che

$$\mathbf{P}(|X_n^*| = x) = \frac{h_n}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} e^{E_n(x)},$$

dove $E_n(x)$ è l'errore che tende a zero uniformemente se x è limitato.

Come applicazione del teorema, si ottiene un'approssimazione della funzione di ripartizione della distribuzione binomiale. Dati $a, b, 0 < a < b$, si calcola:

$$\mathbf{P}(a \leq X_n^* \leq b) = \sum_{a \leq x \leq b} \mathbf{P}(|X_n^*| = x) = \sum_{a \leq x \leq b} \frac{h_n}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} e^{E_n(x)}.$$

A meno del resto, $\sum_{a \leq x \leq b} \frac{h_n}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} e^{E_n(x)}$ è la somma di Riemann di $e^{-\frac{x^2}{2}}$, quindi converge a

$$\int_a^b e^{-\frac{x^2}{2}} dx = \mathcal{N}(b) - \mathcal{N}(a),$$

dove $\mathcal{N}(x)$ è la funzione di ripartizione della gaussiana standard. Si può quindi studiare la convergenza della funzione di ripartizione $F_n(x)$ di X_n^* nel seguente modo:

$$\begin{aligned} F_n(x) &= \mathbf{P}(X_n^* \leq x) = \mathbf{P}(-k < X_n^* \leq x) + \mathbf{P}(X_n^* \leq -k) \\ &= \mathcal{N}(x) - \mathcal{N}(-k) + \mathbf{P}(X_n^* \leq -k) + E'_n(x) \end{aligned}$$

con $\lim_{n \rightarrow \infty} E'_n(x) = 0$. Per la disuguaglianza di Chebychev, si può rendere $\mathbf{P}(X_n^* \leq -k)$ piccola a piacere. Inoltre anche $\mathcal{N}(-k)$ tende a zero per $k \rightarrow \infty$, quindi la funzione di ripartizione della binomiale tende alla funzione di ripartizione della gaussiana standard.

Catene di Markov con tempo discreto

6.1 Catene di Markov omogenee con un numero finito di stati e tempo discreto

Definiamo una *catena di Markov omogenea* con spazio degli stati S , dove S è un insieme finito, come una successione di numeri aleatori $(X_i)_{i \in \mathbb{N}}$ con $I(X_i)_{i \in \mathbb{N}} = S$ tale che

$$\mathbf{P}(X_0 = s_0, X_1 = s_1, \dots, X_n = s_n) = \rho_{s_0} p_{s_0, s_1} p_{s_1, s_2} \cdots p_{s_{n-1}, s_n},$$

dove

1. ρ_{s_i} , $i = 1, \dots, n$, è detta la *distribuzione iniziale*

$$\rho_{s_i} = \mathbf{P}(X_0 = s_i) \quad i = 1, \dots, n$$

2. P tale che $[P]_{ij} = p_{ij}$ è la *matrice di transizione* e possiede le seguenti proprietà:
 - P è una matrice quadrata di ordine n ,
 - $0 \leq p_{ij} \leq 1$,
 - $\sum_{j=1}^n p_{ij} = 1 \quad \forall i = 1, \dots, n$.

L'elemento p_{ij} di P è la probabilità subordinata di passare dallo stato i allo stato j . La catena $(X_i)_{i \in \mathbb{N}}$ si può interpretare come un sistema che si evolve nel tempo, passando da uno stato all'altro di S in maniera aleatoria.

Si interpreta quindi l'elemento p_{ij} di P come la probabilità subordinata di passare dallo stato i allo stato j .

Si ha inoltre che

$$\begin{aligned} \mathbf{P}(X_r = s_r | X_{r-1} = s_{r-1}, \dots, X_0 = s_0) &= \\ \frac{\mathbf{P}(X_r = s_r, X_{r-1} = s_{r-1}, \dots, X_0 = s_0)}{\mathbf{P}(X_{r-1} = s_{r-1}, \dots, X_0 = s_0)} &= \\ \frac{\rho_{s_0} p_{s_0, s_1} p_{s_1, s_2} \cdots p_{s_{r-1}, s_r}}{\rho_{s_0} p_{s_0, s_1} p_{s_1, s_2} \cdots p_{s_{r-2}, s_{r-1}}} &= \\ p_{s_{r-1}, s_r}, \end{aligned}$$

se il denominatore è positivo. La probabilità di trovarsi al tempo r nello stato s_r sapendo tutti gli stati precedenti dipende solo dallo stato immediatamente precedente (*proprietà di Markov*).

Esempio 6.1.1 (Passeggiata aleatoria). Una *passeggiata aleatoria* in un intervallo di interi $[a, b] \subset \mathbb{Z}$ con $a < b$, con condizioni al bordo assorbenti è rappresentata da una catena di Markov con matrice di transizione

$$P = \begin{pmatrix} 1 & 0 & 0 & \cdots & \cdots & 0 \\ 1-p & 0 & p & \ddots & & \vdots \\ 0 & \ddots & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & 0 \\ \vdots & & & \ddots & 1-p & 0 & p \\ 0 & \cdots & \cdots & 0 & 0 & 0 & 1 \end{pmatrix}.$$

Le condizioni al bordo sono determinate dalle probabilità di transizione dallo stato iniziale a a quello finale b . Si possono considerare differenti condizioni al bordo ad esempio riflettenti o miste definendo diversamente queste probabilità di transizione.

Nel caso di $p = \frac{1}{2}$ si parla di *passeggiata aleatoria simmetrica*.

Esempio 6.1.2. Consideriamo due urne A e B , ciascuna con N palline. Fra tutte le palline ve ne sono N bianche ed N nere. Si scelgono una pallina dall'urna A e una pallina dall'urna B e si scambiano.

Il numero aleatorio X_i rappresenta il numero di palline bianche nell'urna A al tempo i . L'insieme degli stati è quindi

$$S = \{0, 1, \dots, N\},$$

mentre per la probabilità p_{kl} di passare dallo stato k allo stato l si ottiene

$$\begin{aligned} p_{k,k} &= \mathbf{P}(\text{si estraggono 2 palline bianche oppure 2 nere}) \quad (6.1) \\ &= \frac{k}{N} \frac{N-k}{N} + \frac{N-k}{N} \frac{k}{N} = 2 \frac{k}{N} \frac{N-k}{N}; \end{aligned}$$

$$\begin{aligned}
 p_{k,k+1} &= \mathbf{P}(\text{si estraggono 1 pallina nera da } A \text{ e 1 bianca da } B) & (6.2) \\
 &= \frac{N-k}{N} \frac{N-k}{N} = \frac{(N-k)^2}{N^2};
 \end{aligned}$$

$$\begin{aligned}
 p_{k,k-1} &= \text{estraggo 1 bianca da } A \text{ e 1 nera da } B & (6.3) \\
 &= \frac{k}{N} \frac{k}{N} = \frac{k^2}{N^2}.
 \end{aligned}$$

Queste formule valgono anche per $k = 0$ e $k = N$. Si ottiene quindi la matrice di transizione

$$P = \begin{pmatrix} 0 & 1 & 0 & 0 & \cdots & 0 \\ \frac{1}{N^2} & \frac{2(N-1)}{N^2} & \left(\frac{N-1}{N}\right)^2 & 0 & \cdots & 0 \\ 0 & \frac{4}{N^2} & \frac{4(N-2)}{N^2} & \left(\frac{N-2}{N}\right)^2 & \cdots & 0 \\ \vdots & & & & \ddots & \\ 0 & \cdots & 0 & \cdots & 1 & 0 \end{pmatrix}.$$

6.2 La probabilità di transizione in n passi

Usando la formula delle probabilità composte possiamo calcolare la probabilità di passare dallo stato s allo stato s' in n passi. Sia $s_0, s_1, \dots, s_{m-1}, s$ una successione tale che $\rho_0 p_{s_0, s_1} p_{s_1, s_2} \cdots p_{s_{m-1}, s}$ siano strettamente positivi. Si ha

$$\begin{aligned}
 &\mathbf{P}(X_{m+n} = s' | X_m = s) = \\
 &\mathbf{P}(X_{m+n} = s' | X_m = s, X_{m-1} = s_{m-1}, \dots, X_0 = s_0) = \\
 &\frac{\mathbf{P}(X_{m+n} = s', X_m = s, X_{m-1} = s_{m-1}, \dots, X_0 = s_0)}{\mathbf{P}(X_m = s, X_{m-1} = s_{m-1}, \dots, X_0 = s_0)} = \\
 &\frac{\sum_{s_{m+1}, \dots, s_{m+n-1}} \mathbf{P}(X_{m+n} = s', X_{m+n-1} = s_{m+n-1}, \dots, X_0 = s_0)}{\mathbf{P}(X_m = s, X_{m-1} = s_{m-1}, \dots, X_0 = s_0)} = \\
 &\frac{\sum_{s_{m+1}, \dots, s_{m+n-1}} \rho_0 p_{s_0, s_1} \cdots p_{s_{m-1}, s} p_{s, s_{m+1}} \cdots p_{s_{m+n-1}, s'}}{\rho_0 p_{s_0, s_1} p_{s_1, s_2} \cdots p_{s_{m-1}, s}} =
 \end{aligned}$$

$$\sum_{s_{m+1}, \dots, s_{m+n-1}} p_{s, s_{m+1}} \cdots p_{s_{m+n-1}, s'} =$$

$$[P^n]_{s, s'} = \mathbf{P}(X_{m+n} = s' | (X_m = s)).$$

Tale probabilità *non* dipende da m , ma solo da n , ovvero dal numero di passi intermedi, e si ottiene come l'elemento di coordinate s, s' dell' n -esima potenza della matrice di transizione P ; questo, da ora in avanti, sarà indicato nel modo seguente:

$$p_{s, s'}^{(n)} := \mathbf{P}(X_{m+n} = s' | X_m = s) = [P^n]_{s, s'}.$$

Per convenzione, si definisce

$$p_{s, s'}^{(0)} := \rho_{s, s'} = \begin{cases} 1 & \text{se } s = s' \\ 0 & \text{altrimenti} \end{cases}.$$

Per esempio

$$p_{s, s'}^{(2)} = \mathbf{P}(X_{m+2} = s' | X_m = s) = \sum_{s_1 \in S} p_{s, s_1} p_{s_1, s'} = [P^2]_{s, s'}.$$

6.3 Classi di equivalenza

Sia $(X_i)_{i \in \mathbb{N}}$ una catena di Markov con numero finito di stati ed omogenea. Uno stato s *comunica* con s' se

$$\exists n \mid p_{s, s'}^{(n)} > 0$$

ovvero se esiste un percorso di lunghezza tale che si passa con probabilità positiva dallo stato s allo stato s' . Questa proprietà si indica con il simbolo: $s \prec s'$.

Due stati, s e s' , si dicono *equivalenti* se $s \prec s'$ e $s' \prec s$. Tale relazione è una relazione di equivalenza in quanto è riflessiva, simmetrica e transitiva. Basta verificare la proprietà transitiva. Supponiamo che $s \prec s'$ e che $s' \prec s''$. Allora

$$s \prec s' \Rightarrow \exists n_1 \mid p_{s, s'}^{(n_1)} > 0,$$

$$s' \prec s'' \Rightarrow \exists n_2 \mid p_{s', s''}^{(n_2)} > 0.$$

Si può concludere che $s \prec s''$ in quanto, se si sceglie $n = n_1 + n_2$, si ottiene

$$p_{s, s''}^{(n_1+n_2)} = [P^{n_1+n_2}]_{s, s''} = \sum_{s_1} p_{s, s_1}^{(n_1)} p_{s_1, s''}^{(n_2)} \geq \underbrace{p_{s, s'}^{(n_1)}}_{>0} \underbrace{p_{s', s''}^{(n_2)}}_{>0} > 0.$$

La relazione \prec di comunicazione fra stati si può estendere senza ambiguità alle classi di equivalenza. Indichiamo con $[s]$ la classe di equivalenza di uno stato s ,

ovvero l'insieme degli stati s' equivalenti ad s secondo la relazione introdotta sopra. Quando $s \prec s'$ diciamo che s' segue s . Diciamo che $[s]$ comunica con $[s']$ e scriviamo $[s] \prec [s']$ se $s \prec s'$. È facile vedere che questa definizione non è ambigua, ovvero non dipende dalla scelta degli elementi nelle classi di equivalenza.

Una classe di equivalenza si dice *massimale* se non è seguita da nessun'altra classe nella relazione che abbiamo definito. Se la catena di Markov si trova in uno stato di una classe massimale, agli istanti successivi, con probabilità 1, essa si trova in uno stato che appartiene alla stessa classe.

Si vuol definire il *periodo* di uno stato. Consideriamo:

$$A_s^+ = \{n \mid p_{s,s}^{(n)} > 0\}.$$

Il periodo di uno stato s è dato dal massimo comun divisore (MCD) degli elementi di A_s^+ . Se il periodo è 1, si dice che lo stato è *aperiodico*. Per esempio, nella passeggiata aleatoria gli stati hanno periodo 2. Se l'insieme $A_s^+ = \{0\}$, il periodo si può considerare ∞ o non definito, a seconda delle convenzioni.

Tutti gli stati di una classe di equivalenza hanno lo stesso periodo, per cui si può parlare del periodo di una classe di equivalenza.

Dimostrazione. Consideriamo due stati equivalenti $s \sim s'$, siano q periodo di s e q' periodo di s' . Basta dimostrare che essi hanno lo stesso periodo è sufficiente provare che q' divide ogni $n \in A_s^+$. Per l'equivalenza, si ha che esiste n_1 tale che $p_{s,s'}^{(n_1)} > 0$ ed esiste n_2 tale che $p_{s',s}^{(n_2)} > 0$. Allora $(n_1 + n_2) \in A_s^+$ perché

$$p_{s,s}^{(n_1+n_2)} = \sum_{s_1} p_{s,s_1}^{(n_1)} p_{s_1,s}^{(n_2)} \geq p_{s,s'}^{(n_1)} p_{s',s}^{(n_2)} > 0.$$

Ma $(n_1 + n_2)$ appartiene anche ad $A_{s'}^+$; quindi q e q' dividono $(n_1 + n_2)$. Inoltre, per ogni $n \in A_s^+$, $(n + n_1 + n_2)$ sta in $A_{s'}^+$ perché

$$p_{s',s'}^{(n+n_1+n_2)} \geq p_{s',s}^{(n_2)} p_{s,s}^{(n)} p_{s,s'}^{(n_1)} > 0.$$

Quindi, q e q' dividono $(n + n_1 + n_2)$ per ogni $n \in A_s^+$ e per ogni $n \in A_{s'}^+$, ovvero q e q' dividono n per ogni $n \in A_s^+$ e per ogni $n \in A_{s'}^+$. Questo implica che

$$q = q'.$$

Una classe C di equivalenza di periodo $q < \infty$ si può decomporre in q sottoinsiemi:

$$C = C_0 \cup C_1 \cup \dots \cup C_{q-1}$$

con la proprietà che se $s \in C_i$, $s' \in C_j$ e $p_{s,s'}^{(n)} > 0$ allora

$$n \equiv (j - i) \pmod{q}.$$

Se una classe di equivalenza massimale ha periodo q , i q sottoinsiemi in cui è decomposta sono percorsi ciclicamente dalla catena di Markov: ovvero se $X_0 \in C_i$, allora $X_1 \in C_{[i+1]_q}$, $X_2 \in C_{[i+2]_q}$ dove indichiamo con la notazione $[k]_q$ l'elemento dell'insieme $\{0, \dots, q-1\}$ equivalente a k modulo q .

6.4 Teorema ergodico

Si vuole studiare il comportamento della catena al passare del tempo. Una catena di Markov con una sola classe di equivalenza aperiodica e con spazio degli stati finito ha la proprietà di convergenza ad una distribuzione invariante sugli stati e indipendente dallo stato iniziale. Questa proprietà è il risultato del seguente teorema detto *ergodico*.

Theorem 6.4.1 (Teorema ergodico). *Sia $(X_i)_{i \in \mathbb{N}}$ una catena di Markov omogenea con insieme finito di stati. Se la catena è irriducibile (ovvero ha una sola classe di equivalenza) aperiodica, allora esiste una distribuzione di probabilità $\Pi = \{\pi_1, \dots, \pi_n\}$ sullo spazio degli stati, e delle costanti C e $0 \leq \delta < 1$ tali che $\forall s \in S, \forall n$ si ha:*

$$|p_{s',s}^{(n)} - \pi_s| \leq C\delta^n.$$

In altre parole esistono π_s tali che:

1. $0 \leq \pi_s \leq 1$
2. $\sum_{s \in S} \pi_s = 1$

e vale che

$$\lim_{n \rightarrow +\infty} p_{s',s}^{(n)} = \pi_s \quad \forall s' \in S$$

con velocità esponenziale.

Questo teorema può essere utilizzato anche nel caso di periodo q maggiore di 1, considerando la catena di Markov associata alla matrice di transizione P^q . Infatti, la restrizione di tale catena ad ognuno dei sottoinsiemi C_0, C_1, \dots, C_{q-1} soddisfa le ipotesi del teorema ergodico.

La distribuzione di probabilità Π che appare nel teorema ergodico ha la proprietà di essere una distribuzione *invariante*: se la poniamo come distribuzione iniziale, se cioè $\mathbf{P}(X_0 = s) = \pi_s$ valesse per ogni $s \in S$, allora per ogni $s \in S$ e per ogni $n \geq 0$

$$\mathbf{P}(X_n = s) = \pi_s.$$

Questa proprietà ci permette di calcolare π_s come soluzione di un sistema di equazioni lineari. Infatti

$$\begin{aligned} \pi_s &= \mathbf{P}(X_1 = s) \\ &= \sum_{s' \in S} \mathbf{P}(X_0 = s') p_{s',s} \\ &= \sum_{s' \in S} \pi_{s'} p_{s',s}. \end{aligned}$$

Inoltre, dato che π_s è una distribuzione di probabilità, abbiamo

$$\sum_{s \in S} \pi_s = 1.$$

$(\pi_s)_{s \in S}$ si dice *distribuzione stazionaria* per la catena. Si può dimostrare che, sotto le ipotesi del teorema ergodico, esiste una e una sola soluzione di questo sistema di $|S| + 1$ equazioni in $|S|$ incognite; una delle equazioni, in questo caso una delle prime $|S|$ equazioni, è funzione lineare delle altre e quindi può non essere considerata nella soluzione del sistema:

$$\begin{cases} \Pi = \Pi^t P \\ \sum \pi_s = 1 \end{cases} \quad (6.4)$$

dove si è posto

$$\Pi = \begin{pmatrix} \pi_1 \\ \pi_2 \\ \vdots \\ \pi_n \end{pmatrix}.$$

Il teorema ergodico ci dice che la catena dimentica lo stato di partenza all'aumentare del tempo. Si dimostra ora l'*unicità*.

Dimostrazione. Supponiamo che $(\mu_s)_{s \in S}$ sia un'altra distribuzione che soddisfa il sistema 6.4. Si ottiene

$$\begin{cases} \mu = \mu^t P \\ \sum \mu_s = 1. \end{cases}$$

Si ha che

$$\mu = \mu^t P \Rightarrow \mu = \mu^t P = \mu^t P^2 = \dots = \mu^t P^n$$

se n cresce all'infinito, P^n converge alla matrice

$$\begin{pmatrix} \pi_1 & \pi_2 & \cdots & \pi_n \\ \pi_1 & \pi_2 & \cdots & \pi_n \\ \vdots & \vdots & & \vdots \\ \pi_1 & \pi_2 & \cdots & \pi_n \end{pmatrix}$$

ovvero

$$\mu_j = \sum_i \mu_i p_{ij} = \sum_i \mu_i p_{ij}^{(n)}.$$

Prendendo il limite $\lim_{n \rightarrow +\infty} p_{ij}^{(n)} = \pi_j$, si ottiene

$$\mu_j = \sum_i \mu_i \pi_j = \pi_j \underbrace{\sum_i \mu_i}_1.$$

Ne segue che $\mu_j = \pi_j$, ovvero l'unicità.

Catene di Markov con tempo continuo

7.1 Introduzione

In questo capitolo introdurremo alcuni semplici sistemi a coda. Un *sistema a coda* può essere descritto in termini di sportelli e di un flusso di clienti che accedono agli sportelli e vengono serviti secondo regole prestabilite e possono o rimanere od uscire dal sistema.

Il caso più semplice è quello in cui abbiamo un flusso di clienti che accedono ad un certo numero di sportelli. Se vi è uno sportello libero, il cliente viene servito subito e dopo il servizio, esce dal sistema. Se invece tutti gli sportelli sono occupati si dispone in coda ed aspetta il suo turno. Dopo esser stato servito esce dal sistema.

Si fa l'ipotesi che i tempi di servizio agli sportelli per i vari clienti siano un sistema di numeri aleatori stocasticamente indipendenti ed identicamente distribuiti e si vuole studiare qual è la distribuzione di probabilità che nel sistema siano presenti n clienti. Chiaramente, per fare ciò non basta considerare un solo numero aleatorio perché il numero di clienti si considera a vari istanti nel tempo. Si introduce quindi un nuovo concetto, quello di *processo stocastico*.

Definizione 7.1.1. *Un processo stocastico $(X_t)_{t \in I}$, con I intervallo di \mathbb{R} , è una famiglia di numeri aleatori indicizzati su un intervallo I di \mathbb{R} .*

Esempio 7.1.2. Una catena di Markov $(X_i)_{i \in \mathbb{N}}$ è un esempio di processo stocastico a *tempi discreti* in quanto gli indici variano fra i numeri naturali.

Si suppone quindi che il flusso dei clienti sia descritto da un processo stocastico $N(t)$ stocasticamente indipendente dai tempi di servizio. Fissato t , $N(t)$ rappresenta il numero aleatorio dei clienti nel sistema al tempo t , mentre al variare di t il processo $N(t)$ ci dà il numero di clienti nel sistema al passar del tempo.

Per caratterizzare un sistema come quello che abbiamo descritto bisogna specificare

1. il processo stocastico che regola il flusso dei clienti;

2. la distribuzione del tempo di servizio;
3. il numero degli sportelli.

Si usa indicare le ipotesi che si fanno su un sistema a code di questo tipo con la seguente notazione:

1. M denota il flusso di Poisson o la distribuzione esponenziale per i tempi di servizio;
2. E_r indica la distribuzione di Erlang di parametri r per i tempi fra gli arrivi (che si suppongono i.i.d.) o per i tempi di servizio;
3. D indica i tempi fra gli arrivi o tempi di servizio (deterministici);
4. G indica che non si fa nessuna ipotesi sulla distribuzione dei tempi fra gli arrivi o dei tempi di servizio (che però si suppongono sempre i.i.d.).

Un processo a coda del tipo considerato sarà indicato con tre simboli separati da due sbarre. Il primo simbolo si riferisce alla distribuzione dei tempi fra gli arrivi (sempre supposti i.i.d.). Il secondo indica la distribuzione dei tempi di servizio (sempre supposti i.i.d. e stocasticamente indipendenti dal flusso degli arrivi). Infine l'ultimo simbolo è un numero e denota il numero degli sportelli. Può assumere eventualmente il valore ∞ .

Considereremo tre esempi di sistemi a coda e precisamente i sistemi $M/M/1$, $M/M/n$ con $n > 1$ e $M/M/\infty$. Prima di far ciò, studiamo le catene di Markov con tempo continuo con insieme di stati numerabile ed il processo $N(t)$ che rappresenta il numero dei clienti entrati nel sistema prima del tempo t .

7.2 Catene di Markov con tempo continuo ed insieme di stati numerabile

Una catena di Markov omogenea in tempo continuo con insieme di stati numerabile è un processo stocastico $(X_t)_{t \geq 0}$ tale che $I(X_t) = \mathbb{N}$ per ogni $t \geq 0$. Nel caso di catene di Markov omogenee con tempo continuo, bisogna considerare per ogni intervallo di tempo t una matrice di transizione $p_{s,s'}(t) = [\Pi(t)]_{ss'}$ non essendovi un intervallo di tempo minimale come nel caso del tempo discreto. Le matrici di transizione sono collegate fra loro dalle *equazioni di Chapman-Kolmogorov* che si possono scrivere in maniera sintetica

$$\Pi(t + t') = \Pi(t) \Pi(t') \quad \forall t, t' \geq 0$$

o esplicitamente

$$p_{s,s'}(t + t') = \sum_{s''} p_{s,s''}(t) p_{s'',s'}(t').$$

Per poter trattare esempi interessanti, per esempio la teoria delle code, dobbiamo considerare il caso in cui lo spazio degli stati sia numerabile. In questo caso $\Pi(t)$ è una “matrice” con infinite righe e infinite colonne con elementi non negativi tale che la somma delle serie degli elementi di ogni riga è uguale

ad 1.

Il prodotto riga per colonna di due matrici di questo tipo si può definire come nel caso di matrici quadrate con una serie al posto della somma: il risultato è ancora una matrice con la stessa proprietà, come è facile verificare.

Come nel caso del tempo discreto si possono determinare le probabilità di transizione in più passi da quelle in un passo, così nel caso di tempo continuo è possibile determinare le probabilità di transizione a partire dal loro comportamento per un intervallo di tempo infinitamente piccolo. Il caso più semplice si ha nel modello denominato “*processo di Poisson*”.

7.3 Processi di Poisson

Il *processo di Poisson* è una catena di Markov a tempo continuo con spazio degli stati $S = \mathbb{N}$. Nel seguito denoteremo con $N(t)$ un processo di Poisson e lo useremo per modellizzare il numero totale dei clienti presenti in un sistema a coda, assumendo che non abbia importanza sapere in che ordine i clienti vengano serviti. Le probabilità di transizione $p_{s,s'}(t)$ di $N(t)$ verificano le seguenti proprietà per intervalli di tempo piccolo:

1. $p_{s,s}(h) = 1 - \lambda h + o(h)$
2. $p_{s,s+1}(h) = \lambda h + o(h)$
3. $p_{s,s'}(h) = o(h)$ per $s' \notin \{s, s+1\}$,

dove λ è un parametro strettamente positivo e $o(h)$ denota infinitesimi di ordine maggiore di h , che supponiamo essere uniformemente infinitesimi.

A partire da queste ipotesi possiamo ottenere un sistema di infinite equazioni differenziali dette *equazioni di Kolmogorov in avanti* per le probabilità di transizione nei vari stati a partire da uno stato fissato (\bar{s}). Fissiamo per esempio $\bar{s} = 0$ e supponiamo che $\mathbf{P}(N(0) = 0) = 1$. Intuitivamente, questo significa che al tempo 0 non c'è nessun cliente nel sistema. Si pone

$$\mu_s(t) = p_{0,s}(t) \quad \text{per } s \in S$$

e denotiamo con μ'_s la derivata prima di μ_s . Le μ_s verificano il sistema di equazioni:

$$\begin{cases} \mu'_0(t) = -\lambda\mu_0(t) \\ \mu'_s(t) = -\lambda\mu_s(t) + \lambda\mu_{s-1}(t) \end{cases} \quad \text{per } s \geq 1 \quad (7.1)$$

Questo si può calcolare come segue. Consideriamo il rapporto incrementale $\frac{\mu_s(t+h) - \mu_s(t)}{h}$ per $s > 0$.

$$\begin{aligned}
\frac{\mu_s(t+h) - \mu_s(t)}{h} &= \frac{p_{0,s}(t+h) - p_{0,s}(t)}{h} = \\
&= \frac{\sum_{j \in S} p_{0,j}(t) p_{j,s}(h) - p_{0,s}(t)}{h} = \\
\frac{1}{h} &((1 - \lambda h + o(h))p_{0,s}(t) + (\lambda h + o(h))p_{0,s-1}(t) + \\
&\frac{1}{h} \left(\sum_{\substack{j \in S \\ j \neq s, j \neq s-1}} p_{0,j}(t) p_{j,s}(h) - p_{0,s}(t) \right) = \\
-\lambda p_{0,s}(t) + \lambda p_{0,s-1}(t) + \frac{o(h)}{h} &= -\lambda \mu_s(t) + \lambda \mu_{s-1}(t) + \frac{o(h)}{h}.
\end{aligned}$$

Per h tendente a zero, tale rapporto converge a

$$\mu'_s(t) = -\lambda \mu_s(t) + \lambda \mu_{s-1}(t) \quad \text{per } s \geq 1.$$

Per $s = 0$, si ottiene

$$\begin{aligned}
\frac{\mu_0(t+h) - \mu_0(t)}{h} &= \frac{p_{0,0}(t+h) - p_{0,0}(t)}{h} = \\
&= \frac{\sum_{j \in S} p_{0,j}(t) p_{j,0}(h) - p_{0,0}(t)}{h} = \\
\frac{(1 - \lambda h + o(h))p_{0,0}(t) + \sum_{j \in S, j \neq 0} p_{0,j}(t) p_{j,0}(h) - p_{0,0}(t)}{h} &= \\
-\lambda p_{0,0}(t) + \frac{o(h)}{h} &= -\lambda \mu_0(t) + \frac{o(h)}{h}
\end{aligned}$$

che converge a

$$\mu'_0(t) = -\lambda \mu_0(t)$$

per h che tende a zero. Possiamo risolvere il sistema di equazioni ottenendo

$$p_{0,s}(t) = \frac{(\lambda t)^s}{s!} e^{-\lambda t}$$

La distribuzione del processo al tempo $t > 0$ è di Poisson con parametro λt . Se invece si assume che il sistema parta a tempo 0 da uno stato \bar{s} arbitrario, le probabilità di transizione sono

$$\begin{cases} p_{\bar{s},s}(t) = 0 & \text{per } s < \bar{s} \\ p_{\bar{s},s}(t) = \frac{(\lambda t)^{s-\bar{s}}}{(s-\bar{s})!} e^{-\lambda t} & \text{per } s \geq \bar{s}. \end{cases} \quad (7.2)$$

Dimostriamo che il sistema (7.2) fornisce la soluzione delle equazioni differenziali (7.1) con stato di partenza \bar{s} . Consideriamo la funzione generatrice

$$\Phi(z, t) = \sum_s p_{\bar{s}, s}(t) z^s.$$

Possiamo derivare $\Phi(z, t)$ rispetto a t e si vede facilmente che possiamo passare la derivata nei termini della serie. Otteniamo quindi, applicando il sistema di equazioni per $\mu_s(t) = p_{\bar{s}, s}(t)$, che

$$\begin{aligned} \frac{\partial}{\partial t} \Phi(z, t) &= \sum_s \mu'_s(t) z^s = \\ -\lambda \sum_{s=0}^{\infty} \mu_s(t) z^s + \lambda \sum_{s \geq 1}^{\infty} \mu_{s-1}(t) z^s &= \lambda(z-1)\Phi(z, t). \end{aligned}$$

Quindi

$$\frac{1}{\Phi(z, t)} \frac{\partial}{\partial t} \Phi(z, t) = \frac{\partial}{\partial t} \log \Phi(z, t) = \lambda(z-1).$$

Da cui

$$\log \Phi(z, t) = \lambda(z-1)t + K,$$

ossia $\Phi(z, t) = K e^{\lambda(z-1)t}$. Dato che $\mu_{\bar{s}}(0) = 1$ e $\mu_s(0) = 0$ se $s \neq \bar{s}$, si ha $\Phi(z, 0) = z^{\bar{s}}$. Abbiamo quindi che

$$\begin{aligned} \Phi(z, t) &= z^{\bar{s}} e^{\lambda(z-1)t} = \\ e^{-\lambda t} z^{\bar{s}} e^{\lambda z t} &= e^{-\lambda t} \sum_k z^{\bar{s}+k} \frac{(\lambda t)^k}{k!}. \end{aligned}$$

Da cui $p_{\bar{s}, s}(t) = 0$ per $s < \bar{s}$, $p_{\bar{s}, s}(t) = \frac{(\lambda t)^{s-\bar{s}}}{(s-\bar{s})!} e^{-\lambda t}$ per $s \geq \bar{s}$. Il processo di Poisson è non decrescente con probabilità 1. Possiamo rappresentare graficamente il processo di Poisson come in figura 7.1, dove la freccia che collega due stati sovrascritta da λ indica che l'intensità di transizione fra i due stati è λ . Osserviamo che da ogni punto s parte una freccia e che in ogni punto $s' \geq 1$ arriva una freccia. Queste due frecce, una uscente ed una entrante, corrispondono ai due termini nella parte destra dell'equazione differenziale. Per $s = 0$ si ha solo una freccia uscente che corrisponde all'unico termine dell'equazione differenziale.

Se indichiamo con $P_s(t) = \mathbf{P}(N(t) = s)$ la probabilità che il processo di Poisson si trovi nello stato s al tempo t , abbiamo

$$P_s(t) = \sum_{\bar{s} \in S} \rho_{\bar{s}} p_{\bar{s}, s}(t),$$

dove ρ_s è la distribuzione iniziale. Da qui si vede che per ogni distribuzione iniziale le $P_s(t)$ soddisfano lo stesso sistema di equazioni

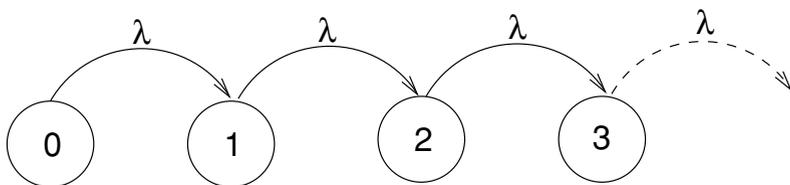


Figura 7.1. Schema di un processo di Poisson di parametro λ .

$$\begin{cases} P'_0(t) = -\lambda P_0(t) \\ P'_s(t) = -\lambda P_s(t) + \lambda P_{s-1}(t) \quad \text{per } s \geq 1 \end{cases}$$

Le $p_{\bar{s},s}(t)$ si possono considerare casi particolari in cui $\rho_{\bar{s}} = 1$ e $\rho_s = 0$ per $s \neq \bar{s}$.

7.4 Processi a coda

Consideriamo ora alcuni esempi di catene di Markov in tempo continuo che servono come modelli per i processi a coda. Nella teoria delle code vi è una notazione simbolica per denotare i tipi di processi. Negli esempi che considereremo il flusso di arrivi segue un processo di Poisson con parametro λ . I clienti che trovano uno sportello libero iniziano un tempo di servizio allo sportello, gli altri si mettono in coda. Quando uno sportello si libera, uno dei clienti in coda inizia il suo servizio.

Per quello che ci interessa non ha importanza in quale ordine i clienti accedono al servizio; poniamo, per esempio, che l'ordine sia casuale. Facciamo l'ipotesi che i tempi di servizio siano indipendenti e identicamente distribuiti e indipendenti dal processo di Poisson che regola il flusso degli arrivi. Supponiamo che la distribuzione del tempo di servizio sia esponenziale di parametro μ .

Un processo di questo tipo viene denotato col simbolo $M/M/n$. Il primo M indica che il flusso degli arrivi è di Poisson, il secondo indica che il tempo di servizio è esponenziale ed n indica il numero degli sportelli e può variare da 1 a ∞ (anche il valore ∞ è ammissibile).

7.5 Code $M/M/\infty$

Si considera una situazione idealizzata in cui vi sono infiniti sportelli. Il processo degli arrivi è dato da un processo di Poisson di parametro λ ed il tempo di servizio T ha distribuzione esponenziale di parametro μ .

Sia $(X_t)_{t \geq 0}$ il processo che indica il numero di persone presenti nel sistema. Come distribuzione iniziale si assume che

$$\begin{cases} \mathbf{P}(X_0 = 0) = 1 \\ \mathbf{P}(X_0 = i) = 0 & i > 0 \end{cases}$$

ovvero che nel sistema non vi sia presente nessuno al tempo 0. Inoltre, il tempo di servizio ed il processo degli arrivi sono stocasticamente indipendenti. Per conoscere l'intensità di servizio si calcola la probabilità che un cliente venga servito nel tempo $(t + h)$ se non è stato servito fino al tempo t .

$$\begin{aligned} \mathbf{P}(T < t + h | T > t) &= \frac{\mathbf{P}(t < T < t + h)}{\mathbf{P}(T > t)} \\ &= \frac{e^{-\mu t} - e^{-\mu(t+h)}}{e^{-\mu t}} \\ &= 1 - e^{-\mu h} \\ &= 1 - (1 - \mu h + o(h)) \\ &= \mu h + o(h), \end{aligned}$$

dove si è usata l'approssimazione $e^{-\mu h} = (1 - \mu h + o(h))$ per h piccolo. Supponiamo che nel sistema vi siano n clienti. Se nessuno di essi è stato servito fino al tempo t , la probabilità che almeno uno di essi sia servito nel tempo $(t + h)$ per h piccolo è allora

$$\begin{aligned} 1 - \mathbf{P}(T_1 > t + h, \dots, T_n > t + h | T_1 > t, \dots, T_n > t) &= \\ 1 - \mathbf{P}(T > t + h | T > t)^n &= \\ 1 - e^{-n\mu h} &= \\ n\mu h + o(h). & \end{aligned}$$

L'intensità con cui un cliente esce dal sistema è quindi proporzionale al numero di clienti che si trovano presenti nel sistema stesso, ovvero allo stato in cui si trova la catena. Si ottiene che il processo può essere rappresentato graficamente come in figura 7.2.

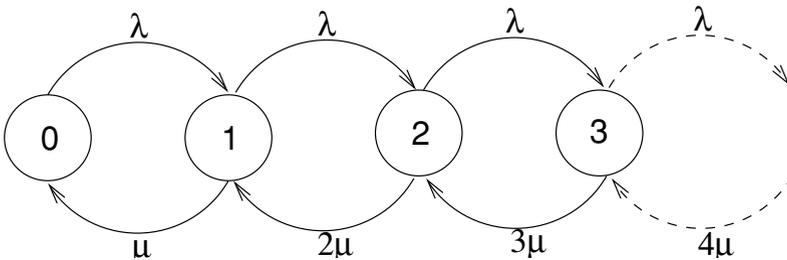


Figura 7.2. Rappresentazione grafica di una coda $M/M/\infty$.

Ponendo $p_{0,s}(t) = \mu_s(t)$, si possono scrivere le equazioni di Kolmogorov utilizzando la regola descritta precedentemente:

$$\begin{cases} \mu'_0(t) = \mu\mu_1(t) - \lambda\mu_0(t) \\ \mu'_i(t) = -(\lambda + i\mu)\mu_i(t) + \lambda\mu_{i-1}(t) + (i+1)\mu\mu_{i+1}(t) \end{cases}$$

dove $\mu'_i(t)$ indica la derivata prima.

Abbiamo visto che per le catene di Markov con tempo discreto che soddisfano le ipotesi del teorema ergodico le probabilità di transizione in n passi convergono alla distribuzione stazionari. Risultati analoghi valgono nel caso di tempo continuo. Cerchiamo quindi una soluzione stazionaria $(p_i)_{i \geq 0}$ di tale sistema, ovvero una distribuzione che *non* dipende dal tempo. Si impone ovvero che $\mu'_i(t) = 0$ e quindi $\mu_i(t) = p_i$ ottenendo

$$\begin{cases} 0 = \mu p_1 - \lambda p_0 \\ 0 = -(\lambda + i\mu)p_i + \lambda p_{i-1} + (i+1)\mu p_{i+1} \\ \sum_{i=0}^{+\infty} p_i = 1. \end{cases}$$

Sommando le equazioni fino alla i -esima si ottiene la formula ricorsiva

$$p_i = \frac{\lambda}{i\mu} p_{i-1} = \frac{1}{i!} \left(\frac{\lambda}{\mu} \right)^i p_0$$

Se si impone la condizione $\sum_{i=0}^{+\infty} p_i = 1$, si ottiene

$$\sum_{i=0}^{+\infty} \frac{1}{i!} \left(\frac{\lambda}{\mu} \right)^i p_0 = 1.$$

La serie $\sum_{i=0}^{+\infty} \frac{1}{i!} \left(\frac{\lambda}{\mu} \right)^i$ è la serie esponenziale con argomento $e^{\frac{\lambda}{\mu}}$, quindi

$$p_i = \frac{\lambda}{i\mu} p_{i-1} = \frac{1}{i!} \left(\frac{\lambda}{\mu} \right)^i e^{-\frac{\lambda}{\mu}}, \quad i \in \mathbb{N}.$$

In questo caso la distribuzione stazionaria esiste sempre.

7.6 Code $M/M/1$

Anche in questo caso, si suppone che il tempo di servizio T abbia distribuzione esponenziale di parametro μ , che sia indipendente dal processo degli arrivi e che questi siano regolati da un processo di Poisson di parametro λ .

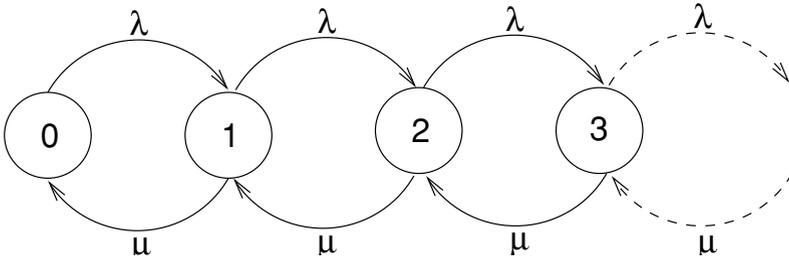


Figura 7.3. Schema di una coda $M/M/1$ con stato iniziale in 0.

Poiché si ha un solo sportello, l'intensità con cui un cliente esce dal sistema è costante e pari a μ . Il grafico del processo è mostrato in figura 7.3. Il sistema di equazioni differenziali, posto $\mu_s(t) = p_{\bar{s},s}(t)$, è dato da

$$\begin{cases} \mu'_0(t) = \mu\mu_1(t) - \lambda\mu_0(t) \\ \mu'_i(t) = -(\lambda + \mu)\mu_i(t) + \lambda\mu_{i-1}(t) + \mu\mu_{i+1}(t) \end{cases}$$

Anche in questo caso si cerca la *soluzione stazionaria*, ovvero tale che $\mu'_i(t) = 0$ per ogni $i \in \mathbb{N}$ con $\mu_i(t) = p_i$.

$$\begin{cases} 0 = \mu p_1 - \lambda p_0 \\ 0 = -(\lambda + \mu)p_i + \lambda p_{i-1} + \mu p_{i+1} \\ \sum_{i=0}^{+\infty} p_i = 1. \end{cases}$$

Dal sistema precedente si ottiene la relazione ricorsiva

$$p_i = \frac{\lambda}{\mu} p_{i-1} = \left(\frac{\lambda}{\mu}\right)^i p_0.$$

Imponendo la condizione

$$\sum_{i=0}^{+\infty} p_i = 1,$$

si ottiene

$$\left(\sum_{i \geq 0} \left(\frac{\lambda}{\mu}\right)^i\right) p_0 = 1,$$

la cui soluzione è

$$p_0 = 1 - \frac{\lambda}{\mu},$$

se $\frac{\lambda}{\mu} < 1$. La distribuzione di probabilità stazionaria è quindi

$$p_i = \left(\frac{\lambda}{\mu}\right)^i \left(1 - \frac{\lambda}{\mu}\right), \quad i \in \mathbb{N},$$

ed esiste se e solo se $\lambda < \mu$, ovvero l'intensità di arrivo deve essere strettamente minore del parametro della distribuzione esponenziale del tempo di servizio.

7.7 Code $M/M/n$

Consideriamo infine le code $M/M/n$ con $n \geq 2$. Dalle considerazioni svolte per i casi precedenti otteniamo per le probabilità di transizione $\mu_s(t) = p_{\bar{s},s}(t)$ per \bar{s} fissato il sistema di equazioni

$$\left\{ \begin{array}{l} \mu'_0(t) = \mu\mu_1(t) - \lambda\mu_0(t) \\ \mu'_1(t) = -(\lambda + \mu)\mu_1(t) + \lambda\mu_0(t) + 2\mu\mu_2(t) \\ \dots \\ \mu'_{n-1}(t) = -(\lambda + (n-1)\mu)\mu_{n-1}(t) + \lambda\mu_{n-2}(t) + n\mu\mu_n(t) \\ \mu'_n(t) = -(\lambda + n\mu)\mu_n(t) + \lambda\mu_{n-1}(t) + n\mu\mu_{n+1}(t) \\ \mu'_{n+1}(t) = -(\lambda + n\mu)\mu_{n+1}(t) + \lambda\mu_n(t) + n\mu\mu_{n+2}(t) \\ \dots, \end{array} \right.$$

dove λ e μ sono, come nei due casi precedenti, rispettivamente il parametro del processo di Poisson che regola gli arrivi dei clienti e quello della distribuzione esponenziale dei tempi di servizio. Il sistema corrisponde al grafico 7.4.

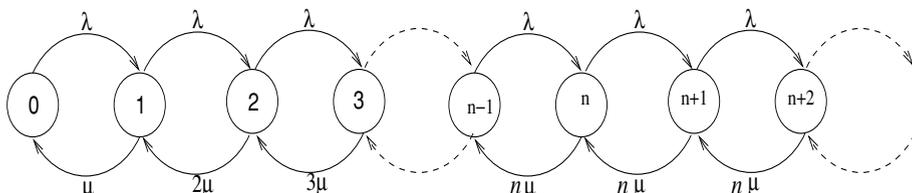


Figura 7.4. Schema di una coda $M/M/n$ con stato iniziale in 0.

Cerchiamo ora la distribuzione stazionaria, ovvero tale che $\mu'_i(t) = 0$ per ogni $i \in \mathbb{N}$. Si ottiene $\mu_i(t) = p_i$ ed il sistema di equazioni lineari

$$\left\{ \begin{array}{l} 0 = \mu p_1 - \lambda p_0 \\ 0 = 2\mu p_2 - \lambda p_1 \\ \dots \\ 0 = (n-1)\mu p_{n-1} - \lambda p_{n-2} \\ 0 = n\mu p_n - \lambda p_{n-1} \\ 0 = n\mu p_{n+1} - \lambda p_n \\ \dots \\ \sum_{i=0}^{+\infty} p_i = 1. \end{array} \right.$$

Si ricavano le seguenti relazioni ricorsive

$$p_i = \frac{\lambda}{i\mu} p_{i-1} \text{ per } i = 1, \dots, n$$

$$p_i = \frac{\lambda}{n\mu} p_{i-1} \text{ per } i \geq n+1.$$

Abbiamo quindi

$$p_i = \left(\frac{\lambda}{\mu}\right)^i \frac{1}{i!} p_0 \text{ per } i = 0, \dots, n$$

$$p_i = \left(\frac{\lambda}{\mu}\right)^i \frac{1}{n!n^{i-n}} p_0 \text{ per } i \geq n+1.$$

Perché esista una soluzione del sistema si deve avere

$$\sum_{i=0}^{n-1} \left(\frac{\lambda}{\mu}\right)^i \frac{1}{i!} + \sum_{i=n}^{\infty} \left(\frac{\lambda}{\mu}\right)^i \frac{1}{n!n^{i-n}} < +\infty.$$

Il primo termine a sinistra è una somma finita e quindi non entra nella discussione sulla convergenza. La serie del secondo termine, ponendo $j = i - n$, si può scrivere

$$\frac{1}{n!} \left(\frac{\lambda}{\mu}\right)^n \sum_{j=0}^{\infty} \left(\frac{\lambda}{n\mu}\right)^j.$$

La condizione per la convergenza è dunque $\frac{\lambda}{n\mu} < 1$, ovvero $\lambda < n\mu$. Questo risultato risponde al problema di quanti sportelli bisogna inserire in un sistema

a coda con flusso di arrivi fissato perché la coda si stabilizzi (per cui quindi esista la distribuzione stazionaria). Per $\lambda < n\mu$ abbiamo quindi

$$p_0 = \left(\sum_{i=0}^{n-1} \left(\frac{\lambda}{\mu}\right)^i \frac{1}{i!} + \frac{1}{n!} \left(\frac{\lambda}{\mu}\right)^n \frac{1}{1 - \frac{\lambda}{n\mu}} \right)^{-1}$$

e

$$p_i = \left(\frac{\lambda}{\mu}\right)^i \frac{1}{i!} p_0 \text{ per } i = 0, \dots, n$$

$$p_i = \left(\frac{\lambda}{\mu}\right)^i \frac{1}{n!n^{i-n}} p_0 \text{ per } i \geq n + 1.$$

7.8 Sistemi a coda in regime stazionario e relazioni di Little

Per i sistemi a coda Markoviani che abbiamo introdotto, l'esistenza di una distribuzione invariante permette di considerare un regime stazionario per $N(t)$, il processo del numero di clienti nel sistema ovvero un processo a coda del tipo assegnato le cui caratteristiche probabilistiche non cambiano nel tempo. Basta considerare il processo con distribuzione iniziale uguale a quella stazionaria. Inoltre si può dimostrare che in tale situazione un sistema a coda si evolve verso il regime stazionario e le medie temporali di quantità di interesse calcolate su intervalli temporali tendono, quando la lunghezza degli intervalli tende all'infinito, alla previsione calcolata nel regime stazionario. Tutto quello che abbiamo detto andrebbe precisato e supportato con dimostrazioni. Ci limitiamo invece ad accettarli a livello intuitivo e passiamo ad esaminare alcune quantità interessanti per lo studio dei sistemi a coda dal punto di vista del loro rendimento e relazioni che le collegano. D'ora in avanti ci riferiremo sempre a sistemi che hanno un regime stazionario.

Per valutare l'efficienza di un sistema a coda si introduce il *fattore di utilizzazione* ρ . Questa quantità è data dal tasso medio di arrivo dei clienti λ per la previsione \bar{x} del tempo di servizio diviso il numero m degli sportelli. Si può dimostrare che il fattore di utilizzazione è uguale alla percentuale media di utilizzazione degli sportelli. Per un sistema non deterministico in regime stazionario si verifica che $\rho < 1$, cioè non è possibile avere una utilizzazione piena degli sportelli (nel regime stazionario ci sarà una probabilità positiva per uno sportello di essere libero). Altre quantità interessanti sono:

1. il numero medio L di clienti nel sistema;
2. il numero medio L_c di clienti in attesa nelle code;

3. il tempo medio W che un cliente trascorre nel sistema;
4. il tempo medio W_c che un cliente trascorre in attesa nelle code.

Le ultime due quantità sono collegate dalla relazione

$$W = W_c + \bar{x},$$

dove \bar{x} è la previsione del tempo di servizio.

Immaginiamo che ogni cliente paghi un importo uguale al tempo che rimane nel sistema. In un intervallo di tempo T la previsione dell'importo pagato, a meno di quantità di ordine inferiore a T , sarà dato da λT (previsione del numero dei clienti arrivati nell'intervallo) moltiplicato per W (previsione del tempo di permanenza nel sistema). Alternativamente la stessa quantità sarà data da LT . Da cui otteniamo la prima *relazione di Little* $L = \lambda W$.

Analogamente se supponiamo che ogni cliente paghi un importo uguale al tempo trascorso in coda, con un argomento analogo otteniamo la seconda *relazione di Little* $L_c = \lambda W_c$.

Le due relazioni di Little sono valide per un'ampia classe di sistemi a coda con regime stazionario.

Consideriamo il caso di un sistema a coda $M/M/1$. Come abbiamo visto, questo sistema ha una distribuzione invariante e quindi un regime stazionario se e solo se $\lambda < \mu$. In questo caso la distribuzione stazionaria per il numero di clienti nel sistema è data da

$$\rho_k = \left(\frac{\lambda}{\mu}\right)^k \left(1 - \frac{\lambda}{\mu}\right).$$

Abbiamo quindi

$$L = \sum_{k=1}^{\infty} k \rho_k = \sum_{k=1}^{\infty} k \left(\frac{\lambda}{\mu}\right)^k \left(1 - \frac{\lambda}{\mu}\right) = \frac{\lambda}{\mu - \lambda}$$

e

$$L_c = \sum_{k=2}^{\infty} (k-1) \rho_k = \sum_{k=1}^{\infty} (k-1) \left(\frac{\lambda}{\mu}\right)^k \left(1 - \frac{\lambda}{\mu}\right) = \frac{\lambda^2}{\mu(\mu - \lambda)}.$$

Quindi usando le relazioni di Little abbiamo

$$W = \frac{1}{\mu - \lambda}, \quad W_c = \frac{\lambda}{\mu(\mu - \lambda)},$$

che soddisfano la relazione $W = W_c + \bar{x}$, dove $\bar{x} = \frac{1}{\mu}$ (previsione della distribuzione esponenziale di parametro μ).

In questo sistema il fattore di utilizzazione ρ è dato da $\frac{\lambda}{\mu}$. Osserviamo che, quando ρ tende a 1, il numero medio dei clienti nel sistema ed in coda ed il tempo medio di un cliente nel sistema ed in coda tendono tutti all'infinito.

Questa è una caratteristica generale dei sistemi a coda stocastici. Se si cerca di aumentare il fattore di utilizzazione si deve pagare il prezzo di far crescere il numero dei clienti in coda ed i loro tempi d'attesa. Il valore 1 non è raggiungibile da un sistema a coda stocastico con regime stazionario; può essere ottenuto da un sistema deterministico ad esempio con uno sportello in cui i clienti arrivano ad intervalli di tempi regolari fissi uguali al loro tempo di servizio.

Statistica

8.1 Densità subordinata di due numeri aleatori

Prima di affrontare lo studio della statistica, si introduce la *densità subordinata* di un numero aleatorio Y rispetto ad un numero aleatorio X . Sia $f(x, y)$ la densità congiunta di (X, Y) . Consideriamo un intervallo $(x, x + h)$ tale che la probabilità

$$P(x < X < x + h) = \int_x^{x+h} \int_{-\infty}^{\infty} f(s, t) dt ds$$

sia strettamente positiva. La probabilità subordinata dell'evento $(Y \leq y)$ dato che $x < X < x + h$ è allora data da

$$\begin{aligned} P(Y \leq y | x < X < x + h) &= \frac{P(x < X < x + h, Y \leq y)}{P(x < X < x + h)} \\ &= \frac{\int_x^{x+h} \int_{-\infty}^y f(s, t) dt ds}{\int_x^{x+h} \int_{-\infty}^{\infty} f(s, t) dt ds}. \end{aligned}$$

Vogliamo far tendere h a 0. Supponiamo che $f(x, y)$ soddisfi le seguenti ipotesi di continuità:

1. $f(x, y)$ è continua rispetto a x tranne che per al più un numero finito di y ;
2. per s appartenente ad un intorno di x , $f(s, y) \leq G(y)$ con $\int_{-\infty}^{\infty} G(y) dy < +\infty$.

Allora

$$\lim_{h \rightarrow 0} P(Y \leq y | x < X < x + h) = \frac{\int_{-\infty}^y f(s, t) dt}{f_X(x)},$$

dove $f_X(x)$ è la densità marginale di X . Se $f(x, y)$ è continua in y , possiamo derivare ed otteniamo $\frac{f(x, y)}{f_X(x)}$. In tal caso, si definisce *densità subordinata di Y rispetto ad X* :

$$f_{Y|X}(y|x) = \frac{f(x, y)}{f_X(x)}$$

se $f_X(x) > 0$.

Si ottiene subito una *formula di Bayes per le densità*. Poiché vale

$$f(x, y) = f_{Y|X}(y|x) f_X(x)$$

ed anche

$$f(x, y) = f_{X|Y}(x|y) f_Y(y),$$

si ottiene

$$f_{Y|X}(y|x) = \frac{f_{X|Y}(x|y) f_Y(y)}{f_X(x)}.$$

8.2 Statistica Bayesiana

8.3 Induzione statistica sulla Binomiale

Si consideri una successione di eventi $(E_i)_{i \in \mathbb{N}}$ stocasticamente indipendenti subordinatamente alla conoscenza di un parametro Θ , ovvero tali che

$$\mathbf{P}(E_i = 1 | \Theta = \theta) = \theta$$

dove $0 < \theta < 1$.

Gli eventi E_i possono essere pensati come il risultato di un esperimento; il fatto che essi siano stocasticamente indipendenti rispetto alla conoscenza di Θ significa che

$$\mathbf{P}(E_1 = \epsilon_1, \dots, E_n = \epsilon_n | \Theta = \theta) = \prod_{i=1}^n \mathbf{P}(E_i = \epsilon_i | \Theta = \theta)$$

dove $\epsilon_i \in \{0, 1\}$.

Si assume che Θ abbia una *distribuzione a priori* $\pi_0(\theta)$ e si vuole vedere come cambia la distribuzione di Θ dopo aver effettuato n esperimenti E_1, \dots, E_n . Supponiamo quindi di aver effettuato gli n esperimenti con esito $E_1 = \epsilon_1, \dots, E_n = \epsilon_n$. La *densità a posteriori* di Θ è data da

$$\pi_n(\theta | E_1 = \epsilon_1, \dots, E_n = \epsilon_n).$$

Usando la formula di Bayes per la densità subordinata di due numeri aleatori si ottiene

$$\begin{aligned} \pi_n(\theta | E_1 = \epsilon_1, \dots, E_n = \epsilon_n) &= \frac{\mathbf{P}(E_1 = \epsilon_1, \dots, E_n = \epsilon_n | \Theta = \theta) \pi_0(\theta)}{\mathbf{P}(E_1 = \epsilon_1, \dots, E_n = \epsilon_n)} \\ &= \frac{1}{c} \pi_0(\theta) \prod_{i=1}^n \mathbf{P}(E_i = \epsilon_i | \Theta = \theta). \end{aligned}$$

dove $c = P(E_1 = \epsilon_1, \dots, E_n = \epsilon_n)$.

In particolare, se la distribuzione a priori $\pi_0(\theta)$ è una distribuzione beta $\mathcal{B}(\alpha, \beta)$ di parametri α e β , anche la distribuzione a posteriori sarà dello stesso tipo, di parametri

$$\alpha' = \alpha + \sum_{i=1}^n \epsilon_i \text{ e } \beta' = \beta + n - \sum_{i=1}^n \epsilon_i$$

dove $\sum_{i=1}^n \epsilon_i$ e $n - \sum_{i=1}^n \epsilon_i$ contano, rispettivamente, il numero di eventi verificati e il numero degli eventi che non si sono verificati. Infatti

$$\begin{aligned} \pi_n(\theta|E_1 = \epsilon_1, \dots, E_n = \epsilon_n) &= k \pi_0(\theta) \prod_{i=1}^n \mathbf{P}(E_i = \epsilon_i | \Theta = \theta) \\ &= k \theta^{\alpha-1} (1-\theta)^{\beta-1} \theta^{\sum_i \epsilon_i} (1-\theta)^{n-\sum_i \epsilon_i} \\ &= k \theta^{\alpha-1+\sum_i \epsilon_i} (1-\theta)^{\beta-1+n-\sum_i \epsilon_i}, \end{aligned}$$

dove k è un'opportuna costante di normalizzazione. Si ricorda la densità della distribuzione beta

$$\pi_0(\theta) = \begin{cases} K \theta^{\alpha-1} (1-\theta)^{\beta-1} & \theta \in [0, 1] \\ 0 & \text{altrimenti.} \end{cases}$$

dove K si calcola come

$$K = \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha) \Gamma(\beta)}$$

Ne segue che la densità a posteriori è una distribuzione beta di parametri $\alpha' = \alpha + \sum_i \epsilon_i$ e $\beta' = \beta + n - \sum_i \epsilon_i$, ovvero

$$\pi_n(\theta|E_1 = \epsilon_1, \dots, E_n = \epsilon_n) = \begin{cases} \frac{\Gamma(\alpha'+\beta')}{\Gamma(\alpha') \Gamma(\beta')} \theta^{\alpha'-1} (1-\theta)^{\beta'-1} & \theta \in [0, 1] \\ 0 & \text{altrimenti.} \end{cases}$$

8.4 Induzione statistica sulla media della distribuzione normale

Sia $(X_i)_{i \in \mathbb{N}}$ una successione di numeri aleatori stocasticamente indipendenti subordinatamente alla conoscenza di un parametro Θ con densità subordinata

$$f(x_i|\theta) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x_i - \theta)^2}{2\sigma^2}}$$

con $\sigma \in \mathbb{R}^+$.

Se la distribuzione a priori $\pi_0(\theta)$ è una gaussiana $N(\mu_0, \sigma_0^2)$, dopo aver

ottenuto i risultati dei primi n esperimenti la distribuzione a posteriori è allora

$$\begin{aligned}\pi_n(\theta|X_1, \dots, X_n) &= k\pi_0(\theta) f(x_1, \dots, x_n|\theta) \\ &= \pi_0(\theta) \prod_{i=1}^n f(x_i|\theta) \\ &= ke^{-\frac{(\theta - \mu_0)^2}{2\sigma_0^2}} e^{-\sum_{i=1}^n \frac{(x_i - \theta)^2}{2\sigma^2}} \\ &= k \exp \left[-\frac{1}{2} \left(\frac{1}{\sigma_0^2} + \frac{n}{\sigma^2} \right) \theta^2 - 2 \left(\frac{\mu_0}{\sigma_0^2} + \frac{\sum x_i}{\sigma^2} \right) \theta \right] \\ &= k \exp \left[-\frac{1}{2} \frac{(\theta - m_n)^2}{\sigma_n^2} \right],\end{aligned}$$

dove $m_n = \frac{\frac{\mu_0}{\sigma_0^2} + \frac{\sum x_i}{\sigma^2}}{\frac{1}{\sigma_0^2} + \frac{n}{\sigma^2}}$ e $\sigma_n^2 = \left(\frac{1}{\sigma_0^2} + \frac{n}{\sigma^2} \right)^{-1}$ e k è un'opportuna costante di normalizzazione. La densità a posteriori è quindi una gaussiana

$$N \left(\frac{\frac{\mu_0}{\sigma_0^2} + \frac{\sum x_i}{\sigma^2}}{\frac{1}{\sigma_0^2} + \frac{n}{\sigma^2}}, \frac{1}{\frac{1}{\sigma_0^2} + \frac{n}{\sigma^2}} \right).$$

8.5 Induzione statistica sulla varianza della distribuzione normale

Consideriamo ora il caso di induzione statistica sulla varianza della distribuzione normale. Si introduce il concetto di *precisione*, definita come l'inverso della varianza, ovvero

$$\Phi = \frac{1}{\sigma^2}.$$

È conveniente utilizzare come parametro l'inverso della varianza invece della varianza stessa, ma è comunque del tutto equivalente. La precisione Φ ha valori positivi. Il termine precisione è legato agli errori di strumenti di misurazione. Sia $(X_n)_{n \in \mathbb{N}}$ una successione di numeri aleatori stocasticamente indipendenti subordinatamente alla conoscenza della precisione, ovvero tali che

$$f(x_1, \dots, x_n|\phi) = f(x_1|\theta) \cdots f(x_n|\theta),$$

dove $\mu \in \mathbb{R}$ e

$$f(x_i|\phi) = k\phi^{\frac{1}{2}} \exp\left(-\frac{\phi}{2}(x_i - \mu)^2\right)$$

per ogni $i = 1, \dots, n$. Il fattore di verosimiglianza è dunque pari a

$$\begin{aligned} f(x_1, \dots, x_n|\phi) &= k\phi^{\frac{n}{2}} \exp\left(-\frac{\phi}{2} \sum_{i=1}^n (x_i - \mu)^2\right) \\ &= k\phi^{\frac{n}{2}} \exp\left(-\frac{nS^2\phi}{2}\right), \end{aligned}$$

dove

$$S^2 := \frac{\sum_{i=1}^n (x_i - \mu)^2}{n}$$

è la *media degli scarti quadratici*. Se si assume che la distribuzione a priori di Φ sia una Gamma $\Gamma(\alpha_0, \lambda_0)$, allora dopo aver effettuato n esperimenti con risultato $X_1 = x_1, \dots, X_n = x_n$ si ottiene che la densità a posteriori di Φ è data da

$$f(\phi|x_1, \dots, x_n) = k\phi^{\frac{n}{2} + \alpha_0 - 1} \exp\left(-\phi\left(\lambda_0 + \frac{nS^2}{2}\right)\right)$$

per $\phi \geq 0$, 0 altrimenti.

8.6 L'uso delle distribuzioni improprie

Consideriamo l'induzione sulla previsione della distribuzione normale. Vogliamo descrivere uno stato iniziale di informazione "vago". Un modo per descrivere un tale stato di informazione è scegliere come distribuzione a priori una distribuzione normale con varianza grande. Si può fare tendere la varianza all'infinito. Al limite non si ottiene una distribuzione di probabilità. Osserviamo tuttavia che le distribuzioni a posteriori convergono. Il limite che si ottiene può essere alternativamente ricavato con l'introduzione di una distribuzione a priori cosiddetta *uniforme impropria* $\pi_0(\theta) = K$. Questa π_0 non rappresenta una distribuzione di probabilità, ma deve essere interpretata nel senso del limite che abbiamo descritto.

8.7 Induzione statistica sulla media e la varianza della distribuzione normale

Consideriamo ora il caso di induzione statistica su media e varianza della distribuzione normale contemporaneamente. Supponiamo di trovarci nel caso di informazione a priori vaga, che, come abbiamo detto, può essere descritta con distribuzioni a priori improprie. Si hanno dunque due parametri incogniti: Θ e Φ , rispettivamente la media e la precisione, ovvero l'inverso della varianza,

della distribuzione normale. Si introduce quindi una distribuzione impropria uniforme per Θ e $\log \Phi$. Questo corrisponde per Θ e Φ alla densità impropria

$$\pi_0(\theta, \phi) = k\phi^{-1}, \quad (\phi > 0).$$

Supponiamo quindi di avere una successione di numeri aleatori che subordinatamente a $\Theta = \theta, \Phi = \phi$ abbiano densità

$$f(x|\theta, \phi) = \frac{1}{\sqrt{2\pi}}\phi^{\frac{1}{2}} \exp\left(-\frac{\phi}{2}(x - \theta)^2\right).$$

Il fattore di verosimiglianza date n osservazioni

$$X_1 = x_1, \dots, X_n = x_n$$

è quindi

$$\begin{aligned} f(x_1, \dots, x_n|\theta, \phi) &= k\phi^{\frac{n}{2}} \exp\left(-\frac{\phi}{2} \sum_{i=1}^n (x_i - \theta)^2\right) \\ &= K\phi^{\frac{n}{2}} \exp\left(-\frac{\phi}{2} ((\bar{x} - \theta)^2 + \nu s^2)\right), \end{aligned}$$

dove $\bar{x} = \frac{\sum_{i=1}^n x_i}{n}$, $s^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{\nu}$. La densità congiunta a posteriori di Θ, Φ è data da

$$\pi_n(\theta, \phi|x_1, \dots, x_n) = K\phi^{\frac{n}{2}-1} \exp\left(-\frac{\phi}{2} ((\bar{x} - \theta)^2 + \nu s^2)\right).$$

Dalla densità congiunta possiamo ottenere la densità marginale di Θ e Φ integrando sull'altra variabile. L'integrale in ϕ si riconduce alla funzione gamma e, dopo aver raccolto nella costante k i fattori che non dipendono da θ , otteniamo

$$\pi_n(\theta|x_1, \dots, x_n) = \int_0^{+\infty} \pi_n(\theta, \phi|x_1, \dots, x_n) d\phi = \frac{k}{((\bar{x} - \theta)^2 + \nu s^2)}.$$

Da questo segue che il numero aleatorio $T = \frac{\bar{x} - \theta}{s\sqrt{\nu}}$ ha densità data dalla t di Student con ν gradi di libertà

$$f_T(t) = \frac{K}{\left(1 + \frac{t^2}{\nu}\right)^{\frac{\nu+1}{2}}}.$$

In modo analogo possiamo ottenere la densità a posteriori di Φ integrando rispetto a Θ . Si tratta di un integrale gaussiano, che a meno di costanti, dà un fattore $\phi^{-\frac{1}{2}}$. Otteniamo quindi che la densità marginale a posteriori di Φ è allora

$$\pi_n(\phi|x_1, \dots, x_n) = \int_0^{+\infty} \pi_n(\theta, \phi|x_1, \dots, x_n) d\theta = k\phi^{\frac{\nu}{2}-1} \exp\left(-\frac{\nu s^2 \phi}{2}\right),$$

dove si è posto $\nu = n-1$. Facendo un cambiamento di variabile lineare vediamo che $U = \nu s^2 \phi$ ha una distribuzione a posteriori con densità

$$k u^{\frac{\nu}{2}-1} \exp\left(-\frac{u}{2}\right),$$

ovvero una chi-quadro con ν gradi di libertà. La costante di normalizzazione è quindi data da $k = \frac{1}{2^{\frac{\nu}{2}} \Gamma(\frac{\nu}{2})}$.

8.8 Test bayesiani di ipotesi. Intervalli di fiducia

Un modo per descrivere in maniera sintetica la distribuzione a posteriori è introdurre intervalli di fiducia (o confidenza) o, nel caso multidimensionale, regioni di fiducia. Sia $\alpha \in (0, 1)$. Si dice *intervallo o regione di fiducia a livello* α un intervallo o una regione che abbia probabilità a posteriori α . È chiaro che vi è arbitrarietà nella scelta di un tale intervallo o regione. Nelle situazioni concrete la scelta può essere fatta seguendo criteri di simmetria se la densità a posteriori è simmetrica rispetto ad un punto o altrimenti minimizzando il volume nello spazio dei parametri.

Nell'impostazione bayesiana i test di ipotesi possono essere collegati con la definizione di intervalli o regioni di fiducia. L'ipotesi che i parametri abbiano un certo valore viene rifiutata ad un livello α se l'intervallo o la regione di fiducia contengono tale valore. Si deve sottolineare che tale procedura è arbitraria in quanto l'intervallo o la regione possono essere scelti in modo arbitrario. Tuttavia dato che in molte situazioni vi è una scelta preferenziale l'uso dei test di ipotesi nell'impostazione bayesiana può essere accettata come una forma più abbreviata ed imprecisa rispetto ad un'analisi completa con la distribuzione a posteriori.

8.9 Confronto fra le medie della distribuzione normale

Supponiamo di avere due campioni di numerosità rispettivamente n_1 e n_2 che, subordinatamente alla conoscenza che i parametri Θ_1 e Θ_2 sono uguali rispettivamente a θ_1 e θ_2 , sono campioni indipendenti con distribuzione gaussiana con densità rispettivamente $N(\theta_1, \sigma_1^2)$ e $N(\theta_2, \sigma_2^2)$. Se la densità a priori di Θ_1 e Θ_2 è uniforme impropria, allora nella distribuzione a posteriori Θ_1 e Θ_2 sono stocasticamente indipendenti con densità rispettivamente $N(\bar{x}_1, \frac{\sigma_1^2}{n_1})$ e $N(\bar{x}_2, \frac{\sigma_2^2}{n_2})$, dove \bar{x}_1, \bar{x}_2 sono le due medie campionarie. Infatti per l'indipendenza stocastica dei campioni e delle densità a priori possiamo applicare

separatamente ai due campioni i risultati sull'induzione sulla media della distribuzione normale nel caso di distribuzione a priori uniforme impropria. Se definiamo $\Theta = \Theta_2 - \Theta_1$, abbiamo che la distribuzione a posteriori di D è una distribuzione normale con previsione $\bar{x}_2 - \bar{x}_1$ e varianza $\frac{\sigma_2^2}{n_2} + \frac{\sigma_1^2}{n_1}$.

Consideriamo ora il caso in cui la varianza delle distribuzioni normali sono incognite ma eguali. Possiamo prendere equivalentemente come parametro l'inverso della varianza Φ detta anche precisione. Abbiamo quindi due campioni di numerosità rispettivamente n_1 e n_2 tali che se $\Theta_1 = \theta_1$, $\Theta_2 = \theta_2$, $\Phi = \phi$ sono campioni indipendenti con densità rispettivamente

$$f_1(x_1|\theta_1, \theta_2, \phi) = \frac{1}{\sqrt{2\pi}} \phi^{\frac{1}{2}} \exp\left(-\frac{\phi}{2}(x_1 - \theta_1)^2\right),$$

$$f_2(x_2|\theta_1, \theta_2, \phi) = \frac{1}{\sqrt{2\pi}} \phi^{\frac{1}{2}} \exp\left(-\frac{\phi}{2}(x_2 - \theta_2)^2\right).$$

Anche qui consideriamo il caso in cui la distribuzione a priori è impropria e precisamente $\Theta_1, \Theta_2, \log \Phi$ sono stocasticamente indipendenti con distribuzione impropria uniforme su \mathbb{R} . Questo corrisponde per Θ_1, Θ_2, Φ a una densità impropria

$$\pi_0(\theta_1, \theta_2, \phi) = K\phi^{-1}, \quad (\phi > 0).$$

Consideriamo dapprima l'induzione statistica per Φ . Qui possiamo applicare senza sostanziale modifica quanto abbiamo visto per l'induzione sulla normale con i due parametri incogniti ed otteniamo quindi che la densità a posteriori di Φ è data da

$$K\phi^{\frac{\nu_1 + \nu_2}{2} - 1} \exp\left(-\frac{S^2\phi}{2}\right),$$

dove $S^2 = \nu_1 s_1^2 + \nu_2 s_2^2$ con $\nu_i = n_i - 1$, $i = 1, 2$ e

$$s_i^2 = \frac{\sum_{j=1}^{n_i} (x_{i,j} - \bar{x}_i)^2}{\nu_i}.$$

Combinando insieme i due risultati precedenti possiamo ottenere la densità a posteriori di $\Theta = \Theta_2 - \Theta_1$ nel caso in cui Φ è incognita; infatti basta integrare la densità per Φ noto moltiplicata per la densità a posteriori di Φ . Otteniamo

$$\pi(\theta|x_1, x_2) = K \int_{\mathbb{R}^+} \phi^{\frac{1}{2}} \exp\left(-\frac{\phi}{2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)} (\theta - (\bar{x}_1 - \bar{x}_2))^2\right) \phi^{\frac{\nu_1 + \nu_2}{2} - 1} \exp\left(-\frac{S^2\phi}{2}\right) d\phi.$$

Questo integrale può essere ridotto ad un integrale di tipo Γ ottenendo

$$\begin{aligned} \pi(\theta|x_1, x_2) &= K \left[\frac{(\theta - (\bar{x}_1 - \bar{x}_2))^2}{\left(\frac{1}{n_1} + \frac{1}{n_2}\right)^{-1} + S^2} \right]^{-\frac{\nu_1 + \nu_2 + 1}{2}} \\ &= K \left[\frac{(\theta - (\bar{x}_1 - \bar{x}_2))^2}{\nu S^2 \left(\frac{1}{n_1} + \frac{1}{n_2}\right)^{-1} + 1} \right]^{-\frac{\nu_1 + \nu_2 + 1}{2}}. \end{aligned}$$

Se definiamo

$$T = \frac{\theta - (\bar{x}_1 - \bar{x}_2)}{S \left(\frac{1}{n_1} + \frac{1}{n_2}\right)^{-\frac{1}{2}}},$$

vediamo che la distribuzione a posteriori di T è t di Student con $\nu = \nu_1 + \nu_2$ gradi di libertà. Questo ci permette di utilizzare le tavole della distribuzione di Student per ottenere intervalli di fiducia per D .

Parte II

Esercizi

Calcolo Combinatorio

Esercizio 9.1. Il bridge si gioca fra quattro giocatori con un mazzo di 52 carte. Calcolare:

1. Il numero di modi diversi in cui un giocatore può ricevere una mano di 13 carte;
2. Il numero di modi diversi in cui si possono distribuire le carte fra i 4 giocatori;
3. Il numero di modi in cui un giocatore può ricevere una mano di carte fra loro tutte diverse in valore. Quanti sono invece i modi in cui tutti e 4 i giocatori ricevono carte tutte differenti in valore?
4. Il numero di modi in cui un giocatore può ricevere una scala dello stesso seme. In quanti modi può invece ricevere almeno due carte uguali in valore?

Soluzione 9.1. 1. Il numero di modi diversi in cui un giocatore può ricevere una mano è dato dalle combinazioni semplici

$$\binom{52}{13}.$$

Infatti, si tratta di scegliere 13 elementi su 52 senza ripetizione e senza tener conto dell'ordine.

2. Nel caso di quattro giocatori, per il primo si è già calcolato il numero di modi al primo punto. Per il secondo giocatore bisogna scegliere 13 carte sulle restanti $52 - 13 = 39$ carte. Si procede in modo analogo per il terzo giocatore. Il quarto giocatore riceve le ultime 13 carte. Il numero di modi in cui i quattro giocatori possono ricevere le carte è allora:

$$\binom{52}{13} \binom{39}{13} \binom{26}{13} \binom{13}{13} = \frac{52!}{(13!)^4}.$$

Il coefficiente multinomiale fornisce infatti il numero di modi di formare 4 gruppi di 13 elementi ciascuno a partire da un insieme di 52 carte.

3. Poiché le carte del giocatore sono tutte diverse in valore, posso pensarle ordinate in scala. Per la prima carta, il giocatore potrà aver ricevuto uno dei quattro assi. Per la seconda carta, uno dei quattro due; e così via fino alla tredicesima carta. Il numero di modi di ricevere una mano di carte tutte diverse in valore è allora:

$$\underbrace{4 \cdot 4 \cdot \dots \cdot 4}_{13 \text{ volte}} = 4^{13}.$$

Se si considerano tutti e quattro i giocatori, si ha che per il secondo giocatore le scelte per ogni carta si riducono a tre; egli può ricevere una mano di carte differenti in

$$\underbrace{3 \cdot 3 \cdot \dots \cdot 3}_{13 \text{ volte}} = 3^{13}$$

modi diversi. I quattro giocatori ricevono carte in valore tutte differenti in

$$4^{13} \cdot 3^{13} \cdot 2^{13} \cdot 1^{13} = (4!)^{13}$$

modi.

4. Un giocatore può ricevere una scala in quattro modi diversi. Esistono scale infatti di quattro semi. Se invece consideriamo quattro giocatori, allora il numero di modi di assegnare a ciascuno una scala è pari al numero di modi di permutare i quattro semi, ovvero

$$4 \cdot 3 \cdot 2 \cdot 1 = 4!.$$

5. È più semplice calcolare il numero dei modi possibili in cui un giocatore può ricevere almeno due carte uguali. Infatti, tale numero è pari a

$$\binom{52}{13} - 4^{13}$$

ovvero, il numero di modi possibili cui è stato tolto il numero dei modi di avere una mano contenente carte tutte diverse.

□

Esercizio 9.2. La biglietteria di un teatro dispone di 100 biglietti numerati da 1 a 100. I biglietti vengono distribuiti a caso agli acquirenti. Quattro amici A, B, C, D acquistano separatamente un biglietto per ciascuno.

1. Qual è la probabilità che essi abbiano ricevuto i biglietti numero 31, 32, 33, 34?
2. Qual è la probabilità che ai quattro amici siano toccati i biglietti numero 31, 32, 33, 34 in questo preciso ordine?
3. Qual è la probabilità che essi ricevano i biglietti con quattro numeri consecutivi?
4. Qual è la probabilità che A, B, C ricevano biglietti con numero maggiore di 50?

Soluzione 9.2. 1. La probabilità si calcola utilizzando la formula

$$\frac{\# \text{ casi favorevoli}}{\# \text{ casi possibili}}$$

I casi possibili sono tutti i modi di scegliere 4 numeri su 100, ovvero

$$\binom{100}{4}.$$

I casi favorevoli sono uno soltanto, ovvero quello di scegliere i numeri 31, 32, 33, 34. Quindi, la probabilità che i tre amici abbiano ricevuto i biglietti 31, 32, 33, 34 è pari a

$$p = \frac{1}{\binom{100}{4}}.$$

2. In questo caso conta l'ordine quindi se il numero dei casi possibili è

$$\mathcal{D}_4^{100} = \frac{100!}{96!}.$$

La probabilità che i 4 amici ricevano i biglietti 31, 32, 33, 34 in questo preciso ordine è

$$p = \frac{1}{\mathcal{D}_4^{100}} = \frac{96!}{100!}.$$

3. Il numero dei modi in cui si possono ottenere biglietti con numeri consecutivi è

$$100 - 3 = 97.$$

Bisogna infatti considerare anche il caso $\{97, 98, 99, 100\}$. La probabilità di ricevere quattro biglietti consecutivi è

$$\frac{97}{\binom{100}{4}} = \frac{97!4!}{100!}.$$

4. La probabilità che A, B, C ricevano biglietti con numero maggiore di 50 è

$$p = \frac{50}{100} \frac{49}{99} \frac{48}{98}.$$

Per il primo ci sono 50 casi favorevoli (tutti i biglietti con numero da 51 a 100) sui 100 possibili. Per il secondo rimangono 99 biglietti da cui scegliere uno dei 49 numeri maggiori di 50 rimasti. Così via fino al terzo.

□

Esercizio 9.3. Il PIN (codice identificativo personale) di una carta di credito è dato da un numero di 5 cifre. Supponiamo che ogni successione possibile abbia la stessa probabilità.

1. Calcolare la probabilità che le cifre del PIN siano tutte diverse.
2. Calcolare la probabilità che il PIN contenga almeno due cifre uguali.
3. Calcolare la probabilità che le cifre del PIN siano tutte diverse se la prima cifra è diversa da 0.
4. Calcolare la probabilità che le cifre del PIN contengano esattamente due cifre uguali se la prima cifra è diversa da 0.

Soluzione 9.3. 1. In questo caso, l'ordine in cui sono disposte le cifre distingue un PIN da un altro. I casi possibili sono dunque

$$10^5.$$

I casi favorevoli, in cui tutte le 5 cifre sono diverse, sono invece

$$\mathcal{D}_5^{10} = \frac{10!}{5!}.$$

Le probabilità che le cifre del PIN siano tutte diverse è

$$p_1 = \frac{\mathcal{D}_5^{10}}{10^5}.$$

2. La probabilità p che il PIN contenga almeno 2 cifre uguali è

$$p = 1 - p_1 = 1 - \frac{10!}{5! 10^5},$$

dove p_1 è la probabilità che le cifre del PIN siano diverse.

3. In questo caso, il numero dei casi possibili è

$$9 \cdot 10 \cdot 10 \cdot 10 \cdot 10.$$

Per la prima cifra si hanno 9 possibilità (tutte le cifre meno lo 0), mentre si possono scegliere le cifre rimanenti senza ripetizione e tenendo conto dell'ordine fra tutte le cifre esclusa quella scelta nella prima posizione in modi \mathcal{D}_4^9 modi. Il numero dei casi favorevoli è

$$9 \cdot 9 \cdot 8 \cdot 7 \cdot 6 = 9 \cdot \mathcal{D}_4^9.$$

La probabilità che le cifre siano diverse se la prima è diversa da 0 è quindi

$$\frac{9 \cdot \mathcal{D}_4^9}{9 \cdot 10^4} = \frac{\mathcal{D}_4^9}{10^4}.$$

4. In questo caso, i casi possibili sono sempre

$$9 \cdot 10^4.$$

Per calcolare il numero di modi in cui un PIN con la prima cifra diversa da 0 può avere esattamente due cifre uguali si può procedere nel seguente modo:

- a) Si sceglie la cifra ripetuta senza perdere di generalità si può pensare che sia uguale alla prima. Vi sono 9 modi di sceglierla.
 b) Si sceglie il posto della cifra ripetuta nella stringa del PIN. Vi sono

$$\binom{4}{1}$$

posti in cui può apparire.

- c) Le altre cifre devono essere diverse dalla prima e si possono disporre nella stringa in

$$\mathcal{D}_3^9$$

modi. In totale

$$9 \cdot \binom{4}{1} \cdot \mathcal{D}_3^9.$$

Il procedimento illustrato nei punti a), b) e c) verrà chiamato d'ora in avanti *metodo della stringa*.

- d) Se la cifra ripetuta è diversa dalla prima, vi sono
- 9 modi per scegliere la prima;
 - 9 modi per scegliere la cifra ripetuta (ovvero 10 cifre meno quella che si è già scelta al primo posto);
 - $\binom{4}{2}$ modi di scegliere il posto;
 - \mathcal{D}_2^8 modi di disporre le due cifre rimanenti.

In totale

$$9 \cdot 9 \cdot \mathcal{D}_2^8 \cdot \binom{4}{2}.$$

Il numero dei casi favorevoli è

$$9 \cdot \binom{4}{1} \cdot \mathcal{D}_3^9 + 9 \cdot 9 \cdot \mathcal{D}_2^8 \cdot \binom{4}{2} = 9 \cdot \left[\binom{4}{1} + \binom{4}{2} \right] \cdot \mathcal{D}_3^9 = 9 \cdot \binom{5}{2} \cdot \mathcal{D}_3^9,$$

dove si è usata la formula

$$\binom{n}{r} = \binom{n-1}{r} + \binom{n-1}{r-1}.$$

Infatti, si poteva anche pensare di scegliere la cifra ripetuta (9 possibilità), i posti occupati dalla cifra ripetuta nella stringa ($\binom{5}{2}$ possibilità) ed infine il modo di disporre le 3 cifre rimanenti (\mathcal{D}_3^9 possibilità). □

Esercizio 9.4. Si lanciano contemporaneamente 4 dadi equilibrati con facce numerate da 1 a 6.

- Calcolare la probabilità di avere facce tutte diverse.
- Calcolare la probabilità di avere almeno due facce uguali.
- Calcolare la probabilità di avere esattamente due facce uguali.
- Calcolare la probabilità che la somma delle facce sia uguale a 5.
- Si lanciano infine solo 2 dadi. Calcolare la probabilità che la somma delle facce sia un numero dispari.

Soluzione 9.4. a) Per calcolare la probabilità, si usa in la formula

$$p = \frac{\# \text{casi favorevoli}}{\# \text{casi possibili}} \quad (9.1)$$

I casi possibili sono esattamente dati da tutti i modi di ottenere una qualsiasi configurazione dei 4 dadi. Quindi

$$\text{casi possibili} = 6 \cdot 6 \cdot 6 \cdot 6 = 6^4.$$

Il numero dei casi possibili coincide con quello delle disposizioni di 4 elementi su 6. I casi favorevoli sono invece dati dalle disposizioni semplici di 4 elementi su 6 in quanto le facce devono essere tutte diverse:

$$\text{casi favorevoli} = 6 \cdot 5 \cdot 4 \cdot 3 = D_4^6.$$

Tale probabilità è dunque

$$\mathbf{P}(\text{i dadi escono con facce tutte diverse}) = \frac{D_4^6}{6^4} = \frac{5}{18}.$$

- b) La probabilità di avere almeno due facce uguali si calcola subito usando il risultato precedente:

$$\begin{aligned} \mathbf{P}(\text{i dadi escono con almeno 2 facce uguali}) &= \\ 1 - \mathbf{P}(\text{i dadi escono con facce tutte diverse}) &= \\ 1 - \frac{D_4^6}{6^4} &= \frac{13}{18}. \end{aligned}$$

- c) Anche in questo caso si usa la formula (9.1) e la “regola della stringa” (si veda l’esercizio 9.3). Il numero di modi di ottenere esattamente due facce uguali è dunque:

$$\binom{4}{2} \cdot 6 \cdot D_2^6,$$

dove

$$\begin{aligned} \binom{4}{2} &= \# \text{ modi di scegliere i 2 dadi che hanno facce uguali,} \\ 6 &= \# \text{ modi di scegliere la faccia ripetuta,} \\ D_2^5 &= \# \text{ modi di scegliere le restanti facce.} \end{aligned}$$

Si noti che le restanti facce devono essere diverse fra loro e da quelle ripetute.

- d) Avendo a disposizione 4 dadi, affinché la somma sia pari a 5 l'unica possibilità è data da tre facce con il numero 1 e una faccia con il numero 2. Per i casi favorevoli, si ha che scelti i 3 posti dell'1, rimane un unico posto che può occupare il 2, ovvero

$$\binom{4}{3} \cdot 1 = 4.$$

Per i casi possibili vi sono 6^4 modi di ottenere una configurazione di 4 dadi. Pertanto la probabilità che la somma delle facce sia pari a 5 è data da:

$$p = \frac{4}{6^4}.$$

- e) La somma delle facce di due dadi è un numero dispari se una delle due è pari e l'altra dispari. Quindi

$$\# \text{casi favorevoli} = \binom{2}{1} \cdot 3 \cdot 3 = 18,$$

dove $\binom{2}{1}$ conta il numero di modi di scegliere il dado che esce con le facce pari. Quindi

$$\mathbf{P}(\text{la somma delle facce è dispari}) = \frac{2 \cdot 3^2}{6^2} = \frac{1}{2}.$$

Più semplicemente, si poteva anche ragionare così: la somma delle facce di due dadi o è pari o è dispari. Quindi

$$\# \text{casi possibili} = 2$$

$$\# \text{casi favorevoli} = 1,$$

ovvero

$$\mathbf{P}(\text{la somma delle facce è dispari}) = \frac{1}{2}.$$

□

Esercizio 9.5. Due fabbriche A e B producono capi di abbigliamento per una stessa marca Y . Per la fabbrica A , il 5% della produzione risulta difettoso, mentre per la B risulta difettoso il 7% della produzione. Inoltre, il 75% dei capi di abbigliamento commercializzati dalla marca Y proviene dalla fabbrica A , mentre il restante 25% dalla fabbrica B . Il capo acquistato si suppone scelto a caso con uguale probabilità fra tutti i capi commercializzati.

1. Calcolare la probabilità di comprare un capo di abbigliamento della marca Y che presenti difetti.
2. Calcolare la probabilità che il capo di abbigliamento provenga dalla fabbrica A subordinata all'evento che esso presenta dei difetti.

Soluzione 9.5. Nel seguito, si indica

- Con A l'evento

$$A = \{\text{il capo proviene dalla fabbrica } A\},$$

- Con B l'evento

$$B = \{\text{il capo proviene dalla fabbrica } B\},$$

- Con D l'evento

$$D = \{\text{il capo risulta difettoso}\}.$$

1. La probabilità di avere acquistato un capo difettoso della marca Y si calcola con la formula delle probabilità totali perché non si conosce a priori se il capo è stato prodotto nella fabbrica A o B .

$$\begin{aligned} \mathbf{P}(D) &= \mathbf{P}(D|A) \mathbf{P}(A) + \mathbf{P}(D|B) \mathbf{P}(B) \\ &= \frac{5}{100} \frac{75}{100} + \frac{7}{100} \frac{25}{100} = \frac{11}{200}. \end{aligned}$$

2. La probabilità che il capo provenga dalla fabbrica A se esso presenta dei difetti è:

$$\mathbf{P}(A|D) = \frac{\mathbf{P}(D|A) \mathbf{P}(A)}{\mathbf{P}(D)} = \frac{15}{22}.$$

Per calcolare tale probabilità subordinata si è usata la formula di Bayes.

□

Esercizio 9.6. Da un'indagine nelle scuole, risulta che la percentuale degli alunni che portano gli occhiali è il 10% nelle scuole elementari, il 25% nelle medie e il 40% nelle superiori.

1. Calcolare la probabilità che, scegliendo a caso 3 studenti, uno per fascia (cioè uno delle scuole elementari, uno delle medie e uno delle superiori), almeno uno dei tre porti gli occhiali.
2. Scegliendo uno studente da una qualunque delle fasce (e supponendo che la scelta di ogni fascia sia equiprobabile) calcolare la probabilità che questo abbia gli occhiali.
3. Sapendo che lo studente scelto porti gli occhiali, calcolare la probabilità che frequenti le scuole elementari.

Soluzione 9.6. 1. Il metodo più veloce per calcolare la probabilità che scegliendo a caso tre studenti, uno per fascia, almeno uno dei tre porti gli occhiali è calcolare la probabilità dell'evento negato, ovvero la probabilità che nessuno dei tre porti gli occhiali. Se B è l'evento che almeno uno dei tre studenti porti gli occhiali si ha che

$$\mathbf{P}(B) = 1 - \mathbf{P}(\tilde{B}) .$$

In questo caso $\mathbf{P}(\tilde{B}) = \frac{90}{100} \frac{75}{100} \frac{60}{100} = \frac{81}{200}$, da cui

$$\mathbf{P}(B) = 1 - \frac{81}{200} = \frac{119}{200} .$$

2. Sia O l'evento

$$O = \{\text{lo studente scelto porta gli occhiali}\} .$$

La probabilità di O si calcola utilizzando la formula delle probabilità totali in quanto, scelto a caso uno studente dalle tre fasce, non si conosce a priori se questi frequenti le elementari, le medie o le superiori. Indichiamo con

- $E = \{\text{lo studente frequenta le elementari}\}$,
- $M = \{\text{lo studente frequenta le medie}\}$,
- $S = \{\text{lo studente frequenta le superiori}\}$.

Allora si ottiene:

$$\begin{aligned} \mathbf{P}(O) &= \mathbf{P}(O|E) \mathbf{P}(E) + \mathbf{P}(O|M) \mathbf{P}(M) + \mathbf{P}(O|S) \mathbf{P}(S) \\ &= \frac{1}{10} \frac{1}{3} + \frac{1}{4} \frac{1}{3} + \frac{2}{5} \frac{1}{3} = \frac{1}{4} . \end{aligned}$$

Si noti che si assume la stessa probabilità di scegliere uno studente delle scuole elementari, delle medie e delle superiori.

3. La probabilità che lo scolaro frequenti le elementari se porta gli occhiali si calcola con la formula di Bayes:

$$\mathbf{P}(E|O) = \frac{\mathbf{P}(O|E)\mathbf{P}(E)}{\mathbf{P}(O)} = \frac{2}{15}.$$

□

Distribuzioni Discrete

Esercizio 10.1. Due amici A e B fanno un gioco con le carte. Il mazzo contiene 52 carte: 13 per ogni seme di cuori, quadri, fiori e picche. Estraggono 2 carte ciascuno a turno. Comincia il giocatore A . Vince chi estrae per la prima volta l'asso di cuori o due carte di fiori. Dopo aver estratto le due carte, le rimettono nel mazzo e lo rimescolano.

- Calcolare la probabilità che il giocatore A vinca alla terza prova (cioè dopo che i due giocatori hanno già fatto due estrazioni ciascuno).
- Calcolare le probabilità che vinca il giocatore A , che vinca il giocatore B o che non vinca nessuno dei due (ripetendo le prove fino a che uno dei due non vinca).
- Calcolare la previsione del tempo T che indica il numero della prova in cui il gioco viene deciso, cioè in cui uno dei due giocatori vince.
- La distribuzione di T è una distribuzione nota?

Soluzione 10.1. a) Si possono pensare le prove ripetute dei due giocatori come una successione di prove stocasticamente indipendenti ed equiprobabili. La probabilità che il giocatore A vinca alla terza prova è quindi pari alla probabilità che l'istante di primo successo coincida con la quinta prova (ciascuno dei due giocatori ha già fatto due prove ed il giocatore A vince alla sua terza prova, quindi in tutto sono già avvenute

$$2 + 2 + 1 = 5$$

prove). Sia T la variabile aleatoria che rappresenta l'istante di primo successo. Bisogna calcolare la probabilità che un giocatore vinca, ovvero che estragga l'asso di cuori oppure due carte di fiori. Tale probabilità è pari a

$$p = \frac{51}{\binom{52}{2}} + \frac{\binom{13}{2}}{\binom{52}{2}} \quad (10.1)$$

dove si è usata la formula $\frac{\# \text{casi favorevoli}}{\# \text{casi possibili}}$ per calcolare

$$1. \text{ La probabilità di estrarre l'asso di cuori} = \frac{1 \cdot \binom{51}{1}}{\binom{52}{2}}.$$

$$2. \text{ La probabilità di estrarre 2 carte di fiori} = \frac{\binom{13}{2}}{\binom{52}{2}}.$$

La probabilità che un giocatore vinca alla terza prova è dunque

$$\mathbf{P}(T = 5) = p(1-p)^4,$$

dove p è dato dalla (10.1)

- b) Se il giocatore A vince, il gioco si ferma ad una prova dispari. Allora la probabilità che A vinca è pari a

$$\begin{aligned} \mathbf{P}(A \text{ vince}) &= \sum_{k=0}^{\infty} \mathbf{P}(T = 2k + 1) = \\ &= \sum_{k=0}^{\infty} p(1-p)^{2k} = p \frac{1}{1 - (1-p)^2}. \end{aligned}$$

La probabilità che B vinca coincide quindi con la probabilità che il gioco si fermi ad una prova pari, ovvero

$$\begin{aligned} \mathbf{P}(B \text{ vince}) &= \sum_{k=1}^{\infty} \mathbf{P}(T = 2k) \\ &= p \sum_{k=1}^{\infty} (1-p)^{2k-1} \\ &= \frac{p}{1-p} \left(\frac{1}{1 - (1-p)^2} - 1 \right) \\ &= \frac{p}{1-p} \frac{(1-p)^2}{1 - (1-p)^2} \\ &= \frac{p(1-p)}{1 - (1-p)^2} = \frac{1-p}{2-p}. \end{aligned}$$

La probabilità che nessuno vinca è

$$\begin{aligned}
\mathbf{P}(\text{nessuno vince}) &= 1 - \mathbf{P}(A \text{ vince}) - \mathbf{P}(B \text{ vince}) \\
&= 1 - \sum_{k=0}^{\infty} \mathbf{P}(T = 2k + 1) - \sum_{k=1}^{\infty} \mathbf{P}(T = 2k) \\
&= 1 - \sum_{k=1}^{\infty} \mathbf{P}(T = k) \\
&= 0.
\end{aligned}$$

c)-d) La variabile aleatoria T che indica l'istante in cui il gioco viene deciso ha distribuzione geometrica di parametro p in quanto denota l'istante di primo successo in una serie di prove indipendenti ed equiprobabili. Quindi la previsione di T è pari a

$$\mathbf{P}(T) = \frac{1}{p} = \frac{\binom{52}{2}}{1 + \binom{13}{2}}.$$

□

Esercizio 10.2. Siano X ed Y due numeri aleatori stocasticamente indipendenti con distribuzione di Poisson di parametri rispettivamente μ e σ .

1. Sia $Z = X + Y$; si calcoli previsione e varianza di Z .
2. Qual è l'insieme dei valori possibili $I(Z)$ di Z ?
3. Calcolare $\mathbf{P}(Z = i)$, per $i \in I(Z)$.
4. Calcolare $\mathbf{cov}(Z, X)$.
5. Sia $u > 0$; calcolare la funzione generatrice $\phi_Z(u) = \mathbf{P}(u^Z)$ di Z .

Soluzione 10.2. 1. Per calcolare la previsione di Z , si usa la linearità della previsione

$$\mathbf{P}(Z) = \mathbf{P}(X + Y) = \mathbf{P}(X) + \mathbf{P}(Y) = \mu + \lambda.$$

Per calcolare la varianza, si usa la formula per la varianza della somma

$$\sigma^2(X + Y) = \sigma^2(X) + \sigma^2(Y) + 2\mathbf{cov}(X, Y).$$

Poiché X e Y sono stocasticamente indipendenti, ne segue che

$$\mathbf{cov}(X, Y) = 0,$$

quindi

$$\sigma^2(X + Y) = \sigma^2(X) + \sigma^2(Y) = \mu + \lambda.$$

Per la distribuzione di Poisson, vale infatti

$$\sigma^2(X) = \mathbf{P}(X) = \mu.$$

2. L'insieme dei valori possibili per Z è

$$I(Z) = \mathbb{N} = \{\inf(X) + \inf(Y), \dots\}.$$

3. Calcoliamo la distribuzione di probabilità di Z . L'evento $\{Z = i\}$ si può scrivere come

$$\begin{aligned} \{Z = i\} &= \{X = 0, Y = i\} + \{X = 1, Y = i - 1\} + \dots + \{X = i, Y = 0\} \\ &= \sum_{k=0}^i \{X = k, Y = i - k\}, \end{aligned}$$

in quanto gli eventi $\{X = k, Y = i - k\}$ sono disgiunti al variare di k . Usando la linearità della previsione, si ottiene

$$\mathbf{P}(Z = i) = \sum_{k=0}^i \mathbf{P}(X = k, Y = i - k).$$

Inoltre X e Y sono stocasticamente indipendenti, quindi

$$\mathbf{P}(X = k, Y = i - k) = \mathbf{P}(X = k) \mathbf{P}(Y = i - k),$$

da cui

$$\begin{aligned} \mathbf{P}(Z = i) &= \sum_{k=0}^i \mathbf{P}(X = k) \mathbf{P}(Y = i - k) \\ &= \sum_{k=0}^i \frac{\mu^k}{k!} e^{-\mu} \frac{\lambda^{(i-k)}}{(i-k)!} e^{-\lambda} \\ &= \frac{e^{-(\mu+\lambda)}}{i!} \sum_{k=0}^i \frac{i!}{k!(i-k)!} \mu^k \lambda^{(i-k)} \\ &= \frac{(\mu + \lambda)^i}{i!} e^{-(\mu+\lambda)}, \end{aligned}$$

dove si è usato lo sviluppo del binomio di Newton. Si ottiene che Z ha distribuzione di Poisson di parametro $\mu + \lambda$.

4. Per calcolare la covarianza tra Z e X si procede nel modo seguente:

$$\begin{aligned} \text{cov}(Z, X) &= \mathbf{P}(ZX) - \mathbf{P}(Z) \mathbf{P}(X) \\ &= \mathbf{P}((X + Y)X) - (\mathbf{P}(X) + \mathbf{P}(Y)) \mathbf{P}(X) \\ &= \mathbf{P}(X^2) + \mathbf{P}(XY) - \mathbf{P}(X)^2 - \mathbf{P}(Y) \mathbf{P}(X) \\ &= \sigma^2(X) \\ &= \mu. \end{aligned}$$

5. Dato $\mu > 0$, la funzione generatrice per Z è data da

$$\phi_Z(u) = \mathbf{P}(u^Z) = \mathbf{P}(u^{X+Y}).$$

Poiché X, Y sono stocasticamente indipendenti

$$\mathbf{P}(u^{X+Y}) = \mathbf{P}(u^X \cdot u^Y) = \mathbf{P}(u^X) \cdot \mathbf{P}(u^Y).$$

Basta calcolare $\mathbf{P}(u^X)$ usando la formula per la previsione di una funzione di X :

$$\begin{aligned} \mathbf{P}(u^X) &= \sum_{i=0}^{+\infty} u^i \mathbf{P}(X = i) \\ &= e^{-\mu} \sum_{i=0}^{+\infty} \frac{(u\mu)^i}{i!} \\ &= e^{(u-1)\mu}, \end{aligned}$$

dove l'ultimo passaggio si è svolto ricordando la serie notevole:

$$\sum_{i=0}^{+\infty} \frac{x^i}{i!} = e^x.$$

Ne segue che

$$\begin{aligned}\phi_Z(u) &= \mathbf{P}(u^X) \cdot \mathbf{P}(u^Y) \\ &= e^{(u-1)\mu} e^{(u-1)\lambda} \\ &= e^{(u-1)(\mu+\lambda)}.\end{aligned}$$

□

Esercizio 10.3. In un paese di 200 abitanti, 5 abitanti sono affetti da una particolare malattia genetica. Viene estratto dalla popolazione un campione di 3 individui (tutti i sottoinsiemi di tre individui hanno la stessa probabilità di essere scelti). Sia X il numero degli individui appartenenti al campione affetti dalla malattia.

1. Determinare l'insieme $I(X)$ dei valori possibili per X .
2. Determinare la distribuzione di X (cioè la probabilità che X assuma ognuno dei valori possibili).
3. Calcolare previsione e varianza di X .

Soluzione 10.3. 1. I valori possibili per X sono 0, 1, 2 e 3, ovvero, nel campione si possono trovare come minimo 0 persone affette dalla malattia e come massimo 3.

2. Consideriamo l'evento $\{X = i\}$, $i \in I(X)$. Determinare la distribuzione di X significa determinare la probabilità

$$\mathbf{P}(X = i), \quad i \in I(X).$$

Per fare ciò, si utilizza la formula

$$\frac{\# \text{ casi favorevoli}}{\# \text{ casi possibili}}.$$

Il numero dei casi possibili è dato dal numero di modi di scegliere 3 persone sui 200 abitanti, ovvero

$$\binom{200}{3}.$$

Il numero di casi favorevoli è dato dal numero di modi di scegliere i persone fra quelle affette dalla malattia e le rimanenti $(3-i)$ persone fra gli abitanti sani, ovvero

$$\binom{5}{i} \binom{195}{3-i}.$$

Si ottiene

$$\mathbf{P}(X = i) = \frac{\binom{5}{i} \binom{195}{3-i}}{\binom{200}{3}}.$$

La distribuzione di X è pertanto *ipergeometrica*.

3. La previsione di X si può calcolare direttamente visto che $I(X)$ consiste solo di 4 valori:

$$\begin{aligned}\mathbf{P}(X) &= \sum_{i=0}^3 i \mathbf{P}(X = i) \\ &= \frac{1}{\binom{200}{3}} \left(5 \binom{195}{2} + 20 \binom{195}{1} + 30 \right) \\ &= \frac{3}{40}.\end{aligned}$$

Per la varianza, il calcolo è analogo; basta applicare la formula

$$\sigma^2(X) = \mathbf{P}(X^2) - \mathbf{P}(X)^2$$

e calcolare $\mathbf{P}(X^2) = \sum_{i=0}^3 i^2 \mathbf{P}(X = i)$.

□

Esercizio 10.4. Ad una corsa di cavalli partecipano 10 concorrenti. Gli scommettitori vincono se indovinano i tre cavalli che si piazzano ai primi tre posti specificando l'ordine di arrivo. Si supponga che tutti gli ordinamenti abbiano la stessa probabilità di verificarsi e che gli scommettitori scelgano con uguale probabilità e indipendentemente i tre cavalli su cui scommettere.

1. Qual è la probabilità che uno scommettitore vinca?
2. Se gli scommettitori sono 100 in tutto, sia X il numero aleatorio che conta il numero di scommettitori che vincono. Calcolare $I(X)$ e $\mathbf{P}(X = i)$ per $i = 1, 2, 3$.
3. Calcolare previsione e varianza di X .
4. Siano gli scommettitori numerati da 1 a 100. Calcolare la probabilità che vi sia un vincitore e che il vincitore col numero minimo abbia un numero maggiore o uguale a 50.

Soluzione 10.4. 1. La probabilità che uno scommettitore vinca si calcola con la formula

$$\frac{\# \text{ casi favorevoli}}{\# \text{ casi possibili}}.$$

In questo caso, i casi possibili sono dati da tutte le disposizioni (semplici) di tre elementi su 10. Esse rappresentano infatti il numero di modi di classificarsi ai primi tre posti dei 10 cavalli. Una sola è la terna vincente, quindi la possibilità di vincere per uno scommettitore è

$$p = \frac{1}{\mathcal{D}_3^{10}} = \frac{7!}{10!} = \frac{1}{720}.$$

2. Se X è il numero aleatorio che conta il numero di scommettitori che vincono, si può scrivere

$$X = E_1 + E_2 + \cdots + E_{100},$$

dove l'evento E_i si verifica se l' i -esimo scommettitore vince. Gli eventi E_i , $i = 1 \dots 100$ sono stocasticamente indipendenti ed equiprobabili perché gli scommettitori scelgono indipendentemente e con la stessa probabilità i cavalli su cui scommettere. X ha quindi distribuzione binomiale $\mathcal{B}(n, p)$ di parametri $n = 100$ e $p = \frac{1}{720}$. L'insieme dei valori possibili $I(X)$ è

$$I(X) = \{0, 1, \dots, 100\}$$

e vale che

$$\mathbf{P}(X = i) = \binom{100}{i} \left(\frac{1}{720}\right)^i \left(1 - \frac{1}{720}\right)^{100-i}.$$

In particolare, si ottiene:

i=1

$$\mathbf{P}(X = 1) = 100 \cdot \frac{1}{720} \cdot \left(\frac{719}{720}\right)^{99}.$$

i=2

$$\mathbf{P}(X = 2) = \binom{100}{2} \left(\frac{1}{720}\right)^2 \cdot \left(\frac{719}{720}\right)^{98}.$$

i=3

$$\mathbf{P}(X = 3) = \binom{100}{3} \left(\frac{1}{720}\right)^3 \cdot \left(\frac{719}{720}\right)^{97}.$$

3. La previsione di X è

$$\begin{aligned} \mathbf{P}(X) &= \mathbf{P}(E_1 + \dots + E_{100}) = \\ &= \sum_{i=1}^{100} \mathbf{P}(E_i) = 100 \cdot \frac{1}{720} = \frac{5}{36}. \end{aligned}$$

Si è utilizzata la proprietà di linearità per il calcolo di $\mathbf{P}(X)$. Analogamente, usando la formula per il calcolo della varianza della somma di n numeri aleatori, si ottiene:

$$\begin{aligned} \sigma^2(X) &= \sigma^2(E_1 + \dots + E_{100}) \\ &= \sum_{i=1}^{100} \sigma^2(E_i) + \underbrace{\sum_{i,j=1}^{100} \text{cov}(E_i, E_j)}_0 \\ &= 100 \cdot \frac{1}{720} \cdot \left(1 - \frac{1}{720}\right). \end{aligned}$$

Si ricordi che gli eventi E_i sono stocasticamente indipendenti.

4. Affinché vi sia un vincitore ed il vincitore con il numero minimo abbia un numero maggiore od uguale a 50, deve essere contemporaneamente che tutti gli scommettitori con numero da 1 a 49 perdano e che almeno uno scommettitore con numero da 50 a 100 vinca. Sia E l'evento che tutti gli scommettitori con numero da 1 a 49 perdano ed F l'evento che vinca almeno uno scommettitore con numero da 50 a 100. La probabilità che vi sia un vincitore ed il vincitore con il numero minimo abbia un numero maggiore od uguale a 50 è allora

$$\mathbf{P}(EF) = \mathbf{P}(E)\mathbf{P}(F) = \left(\frac{719}{720}\right)^{49} \left[1 - \left(\frac{719}{720}\right)^{51}\right].$$

dove $\mathbf{P}(F) = 1 - \mathbf{P}(\tilde{F})$ ed \tilde{F} è l'evento che nessun scommettitore con numero da 50 a 100 vinca.

□

Esercizio 10.5. In un sondaggio d'opinione viene sottoposto a 100 intervistati un questionario con 5 domande. Ad ogni domanda si può rispondere solo sì o no. Per ogni intervistato la probabilità di ogni possibile scelta delle risposte è uguale e le scelte di ogni intervistato sono stocasticamente indipendenti da quelle degli altri. Sia N il numero degli intervistati che rispondono sì alle prime due domande o rispondono sì ad almeno quattro domande.

1. Qual è la distribuzione di N ?
2. Calcolare la previsione, la varianza e la funzione generatrice di N .

Soluzione 10.5. 1. Si chiama E_i l'evento che si verifica se l' i -esimo intervistato risponde sì alle prime 2 domande oppure sì ad almeno 4 domande. Si può allora scrivere N come

$$N = E_1 + E_2 + \dots + E_{100}.$$

Gli eventi sono stocasticamente indipendenti ed equiprobabili perché ogni intervistato sceglie indipendentemente dagli altri e con la stessa probabilità. Basta calcolare la probabilità degli E_i . Poniamo:

- $F_i = \{ \text{l}'i\text{-esimo intervistato risponde sì alle prime 2 domande} \}$
- $G_i = \{ \text{l}'i\text{-esimo intervistato risponde sì ad almeno 4 domande} \}$

Si ottiene che $E_i = F_i \vee G_i$ e

$$\mathbf{P}(E_i) = \mathbf{P}(F_i) + \mathbf{P}(G_i) - \mathbf{P}(F_i \wedge G_i).$$

Le probabilità di F_i , G_i e $F_i \wedge G_i$ sono date da:

a)

$$\mathbf{P}(F_i) = \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4}.$$

Infatti, per ogni domanda i casi possibili sono due (si può rispondere con un sì o con un no) mentre per le prime due domande si ha una sola scelta possibile (sì).

b)

$$\mathbf{P}(G_i) = \binom{5}{4} \left(\frac{1}{2}\right)^5 + \binom{5}{5} \left(\frac{1}{2}\right)^5.$$

Infatti, un intervistato risponde sì ad almeno quattro domande se risponde sì ad esattamente quattro domande oppure esattamente a cinque domande.

c)

$$\mathbf{P}(F_i \wedge G_i) = \binom{3}{2} \left(\frac{1}{2}\right)^5 + \binom{3}{3} \left(\frac{1}{2}\right)^5.$$

Nel caso in cui i due eventi accadano entrambi, bisogna scegliere solo le altre due o tre domande cui l'intervistato risponde sì.

Infine

$$\mathbf{P}(E_i) = \frac{1}{4} + \binom{5}{4} \frac{1}{2^5} + \frac{1}{2^5} - \frac{3}{2^5} - \frac{1}{2^5} = \frac{1}{2^2} + \frac{1}{2^4} = \frac{5}{16}.$$

Si ottiene che $I(N) = \{0, \dots, 100\}$ e vale

$$\mathbf{P}(N = i) = \binom{100}{i} \left(\frac{5}{16}\right)^i \left(1 - \frac{5}{16}\right)^{100-i}.$$

2. La previsione di N è

$$\mathbf{P}(N) = \sum_{i=1}^{100} \mathbf{P}(E_i) = 100 \cdot \frac{5}{16} = \frac{125}{4}.$$

La varianza di N è

$$\begin{aligned} \sigma^2(N) &= \sum_{i=1}^{100} \sigma^2(E_i) + \underbrace{\sum_{i,j=1}^{100} \mathbf{cov}(E_i, E_j)}_0 \\ &= 100 \cdot \frac{5}{16} \cdot \left(1 - \frac{5}{16}\right). \end{aligned}$$

La funzione generatrice di N è

$$\begin{aligned} \phi_N(t) &= \mathbf{P}(t^N) \\ &= \sum_{i=0}^{100} \binom{100}{i} \cdot \left(\frac{5t}{16}\right)^i \cdot \left(1 - \frac{5}{16}\right)^{100-i} \\ &= \left(\frac{5t}{16} + 1 - \frac{5}{16}\right)^{100}, \end{aligned}$$

dove di nuovo si è usata la formula per il binomio di Newton.

□

Esercizio 10.6. Un'urna contiene 8 palline di cui 4 bianche e 4 nere. Si scelgono 4 palline senza reimbussolamento. Sia E_i l'evento che l' i -esima pallina estratta sia bianca e siano $X = E_1 + E_2, Y = E_3 + E_4$.

- Calcolare la distribuzione congiunta di X e Y .
- Calcolare $\mathbf{P}(X), \mathbf{P}(Y), \sigma^2(X), \sigma^2(Y)$.
- Calcolare $\mathbf{cov}(X, Y)$, il coefficiente di correlazione $\rho(X, Y)$ e dire se X e Y sono stocasticamente indipendenti.

Soluzione 10.6. a) Si consideri il vettore aleatorio (X, Y) . L'insieme dei valori possibili di (X, Y) è

$$I(X, Y) = \{(i, j) | i = 0, 1, 2, j = 0, 1, 2\}.$$

Per calcolare la distribuzione congiunta di (X, Y) , bisogna quindi calcolare

$$\mathbf{P}(X = i, Y = j)$$

per ogni $(i, j) \in I(X, Y)$. Le ultime due estrazioni E_3 e E_4 dipendono dalle prime due, quindi si usa la formula delle probabilità subordinate:

$$\mathbf{P}(X = i, Y = j) = \mathbf{P}(Y = j | X = i) \mathbf{P}(X = i).$$

Si ha che la probabilità di estrarre i palline bianche nelle prime 2 estrazioni è

$$\mathbf{P}(X = i) = \frac{\binom{4}{i} \binom{4}{2-i}}{\binom{8}{2}}.$$

Qui i casi possibili sono $\binom{8}{2}$ perché si guardano solo le prime 2 estrazioni.

Inoltre

$$\begin{aligned} \mathbf{P}(Y = j | X = i) &= \frac{\binom{4-i}{j} \binom{4-(2-i)}{2-j}}{\binom{6}{2}} \\ &= \frac{\binom{4-i}{j} \binom{2+i}{2-j}}{\binom{6}{2}}. \end{aligned}$$

Dalle restanti 6 palline nell'urna se ne devono infatti estrarre ancora 2, di cui j fra le bianche rimaste $(4-i)$ e $(2-j)$ fra le nere rimaste $(4-(2-i) = 2+i)$. La distribuzione congiunta di (X, Y) è

$$\mathbf{P}(X = i, Y = j) = \frac{\binom{4-i}{j} \binom{2+i}{2-j}}{\binom{6}{2}} \cdot \frac{\binom{4}{i} \binom{4}{2-i}}{\binom{8}{2}}.$$

b) Per calcolare $\mathbf{P}(X)$ e $\mathbf{P}(Y)$ si usa il fatto che gli eventi E_i sono equiprobabili (ma non indipendenti!), quindi

$$\mathbf{P}(X) = \mathbf{P}(E_1) + \mathbf{P}(E_2) = 2 \cdot \frac{4}{8} = 1$$

e

$$\mathbf{P}(X) = \mathbf{P}(Y) = 1.$$

Gli eventi E_1 ed E_2 (e quindi anche E_3 e E_4) sono negativamente correlati con correlazione:

$$\begin{aligned} \mathbf{cov}(E_1, E_2) &= \mathbf{P}(E_1 E_2) - \mathbf{P}(E_1)\mathbf{P}(E_2) = \\ &= \mathbf{P}(E_2|E_1)\mathbf{P}(E_1) - \mathbf{P}(E_1)\mathbf{P}(E_2) = -\frac{1}{28}. \end{aligned}$$

La varianza di X è quindi

$$\begin{aligned} \sigma^2(X) &= \sigma^2(E_1 + E_2) = \sigma^2(E_1) + \sigma^2(E_2) + 2\mathbf{cov}(E_1, E_2) \\ &= \frac{1}{4} + \frac{1}{4} - \frac{1}{28} = \frac{13}{28}. \end{aligned}$$

Anche in questo caso $\sigma^2(Y) = \sigma^2(X) = \frac{13}{28}$.

c) I numeri aleatori X ed Y non possono essere stocasticamente indipendenti perché sono dati dalla somma di eventi fra loro correlati. Si ha:

$$\begin{aligned} \mathbf{cov}(X, Y) &= \mathbf{cov}(E_1 + E_2, E_3 + E_4) \\ &= \mathbf{cov}(E_1, E_3) + \mathbf{cov}(E_1, E_4) + \mathbf{cov}(E_2, E_3) + \mathbf{cov}(E_2, E_4) \\ &= 4 \cdot \left(-\frac{1}{28}\right) = -\frac{1}{7}. \end{aligned}$$

Qui si è usato il fatto che la covarianza è bilineare. Infine, il coefficiente di correlazione fra X e Y è dato da:

$$\rho(X, Y) = \frac{\mathbf{cov}(X, Y)}{\sigma(X)\sigma(Y)} = \frac{-\frac{1}{7}}{\sqrt{\frac{13}{28}} \cdot \sqrt{\frac{13}{28}}} = -\frac{4}{13}.$$

□

Esercizio 10.7. Siano E_1, E_2, F_1, F_2 eventi stocasticamente indipendenti con $\mathbf{P}(E_1) = \mathbf{P}(E_2) = \frac{1}{4}$, $\mathbf{P}(F_1) = \mathbf{P}(F_2) = \frac{1}{3}$. Siano $X = E_1 + E_2, Y = F_1 + F_2$.

- Calcolare i valori possibili e la distribuzione di probabilità di X e Y .
- Calcolare $\mathbf{P}(X + Y)$, $\sigma^2(X + Y)$.
- Calcolare $\mathbf{P}(X = Y)$, $\mathbf{P}(X = -Y)$.

Soluzione 10.7. a) Poiché E_1, E_2 sono eventi, ovvero numeri aleatori che possono assumere solo i valori 0 e 1, si ha che l'insieme dei valori possibili di X è

$$I(X) = \{0, 1, 2\}.$$

Analogamente per Y

$$I(Y) = \{0, 1, 2\}.$$

Calcolare la distribuzione di probabilità di X significa calcolare con quale probabilità X assume ciascuno dei valori possibili. Si ha per esempio che

$$\begin{aligned} \mathbf{P}(X = 0) &= \mathbf{P}(E_1 + E_2 = 0) \\ &= \mathbf{P}(E_1 = E_2 = 0) = \mathbf{P}(\tilde{E}_1)\mathbf{P}(\tilde{E}_2) = \frac{9}{16}. \end{aligned}$$

Poiché X è pari alla somma di due eventi equiprobabili e stocasticamente indipendenti, si può in realtà subito dire che la distribuzione di X è una binomiale $\mathcal{B}(n, p)$ di parametri $n = 2$ e $p = \frac{1}{4}$. Analogamente $Y \sim \mathcal{B}(2, \frac{1}{3})$ e si ha che

$$\begin{aligned} \mathbf{P}(X = i) &= \binom{2}{i} \left(\frac{1}{4}\right)^i \left(\frac{3}{4}\right)^{2-i}, \quad i = 0, 1, 2 \\ \mathbf{P}(Y = j) &= \binom{2}{j} \left(\frac{1}{3}\right)^j \left(\frac{2}{3}\right)^{2-j}, \quad j = 0, 1, 2. \end{aligned}$$

- Per calcolare la previsione della somma, basta usare la linearità

$$\begin{aligned} \mathbf{P}(X + Y) &= \mathbf{P}(E_1 + E_2 + F_1 + F_2) \\ &= \mathbf{P}(E_1) + \mathbf{P}(E_2) + \mathbf{P}(F_1) + \mathbf{P}(F_2) \\ &= 2 \cdot \frac{1}{4} + 2 \cdot \frac{1}{3} = \frac{7}{6}. \end{aligned}$$

Per le varianze, si applica invece la formula della varianza:

$$\sigma^2(X + Y) = \sigma^2(X) + \sigma^2(Y) + 2\text{cov}(X, Y).$$

Poiché X, Y hanno distribuzione binomiale

$$\begin{aligned} \sigma^2(X) &= 2 \cdot \frac{1}{4} \cdot \frac{3}{4} = \frac{3}{8} \\ \sigma^2(Y) &= 2 \cdot \frac{1}{3} \cdot \frac{2}{3} = \frac{4}{9}. \end{aligned}$$

Per calcolare la covarianza fra X e Y si usa il fatto che gli eventi E_1, E_2, F_1, F_2 sono stocasticamente indipendenti (oltre alla proprietà di bilinearità delle covarianze!)

$$\begin{aligned}\mathbf{cov}(X, Y) &= \mathbf{cov}(E_1 + E_2, F_1 + F_2) \\ &= \mathbf{cov}(E_1, F_1) + \mathbf{cov}(E_1, F_2) + \mathbf{cov}(E_2, F_1) + \mathbf{cov}(E_2, F_2) \\ &= 0.\end{aligned}$$

Quindi

$$\sigma^2(X + Y) = \frac{3}{8} + \frac{4}{9} = \frac{59}{72}.$$

c) Per calcolare $\mathbf{P}(X = Y)$ si osserva che l'evento

$$(X = Y)$$

è dato da

$$(X = Y) = (X = 0, X = 0) + (X = 1, Y = 1) + (X = 2, Y = 2).$$

Quindi

$$\begin{aligned}\mathbf{P}(X = Y) &= \sum_{i=0}^2 \mathbf{P}(X = i, Y = i) \\ &= \sum_{i=0}^2 \mathbf{P}(X = i)\mathbf{P}(Y = i) \\ &= \sum_{i=0}^2 \binom{2}{i} \left(\frac{1}{4}\right)^i \left(\frac{3}{4}\right)^{2-i} \binom{2}{i} \left(\frac{1}{3}\right)^i \left(\frac{2}{3}\right)^{2-i} \\ &= \sum_{i=0}^2 \binom{2}{i}^2 \left(\frac{1}{12}\right)^i \left(\frac{1}{2}\right)^{2-i} \\ &= \frac{1}{4} \sum_{i=0}^2 \binom{2}{i}^2 \left(\frac{1}{6}\right)^i = \frac{61}{144}.\end{aligned}$$

Invece l'evento

$$(X = -Y)$$

si verifica solo se

$$(X = -Y) = (X = 0, Y = 0).$$

Quindi

$$\mathbf{P}(X = -Y) = \mathbf{P}(X = 0, Y = 0) = \mathbf{P}(X = 0)\mathbf{P}(Y = 0) = \frac{1}{4}.$$

□

Distribuzioni Assolutamente Continue in una Dimensione

Esercizio 11.1. I numeri aleatori X, Y e Z sono stocasticamente indipendenti con distribuzione esponenziale di parametro $\lambda = 2$.

- a) Calcolare la densità di probabilità di $X + Y$ e di $X + Y + Z$.
- b) Siano E, F, G gli eventi $E = (X \leq 2), F = (X + Y > 2), G = (X + Y + Z \leq 3)$. Calcolare $\mathbf{P}(E), \mathbf{P}(F), \mathbf{P}(G)$ e $\mathbf{P}(EF)$.
- c) Determinare se E, F, G siano stocasticamente indipendenti.

Soluzione 11.1. a) La distribuzione esponenziale di parametro λ è un caso particolare della distribuzione gamma, corrispondente ai parametri $1, \lambda$. Se X, Y e Z sono numeri aleatori stocasticamente indipendenti con distribuzione esponenziale, ovvero $\Gamma(1, 2)$, per calcolare la densità della loro somma si può usare la proprietà della somma di numeri aleatori stocasticamente indipendenti con distribuzione gamma. Si ha che

$$\Gamma(\alpha, \lambda) + \Gamma(\beta, \lambda) \sim \Gamma(\alpha + \beta, \lambda),$$

quindi $W_1 = X + Y$ ha distribuzione $\Gamma(2, 2)$. Iterativamente

$$W_2 = X + Y + Z = W_1 + Z$$

ha distribuzione $\Gamma(3, 2)$.

- b) Si possono quindi calcolare:
 - i)

$$\begin{aligned}\mathbf{P}(E) &= \mathbf{P}(X \leq 2) = \int_0^2 2e^{-2x} dx = \\ &= 1 - e^{-4}.\end{aligned}$$

- ii)

$$\begin{aligned}\mathbf{P}(F) &= \mathbf{P}(X + Y > 2) = \int_2^{+\infty} 4xe^{-2x} dx = \\ &= [-2xe^{-2x}]_2^{+\infty} + 2 \int_2^{+\infty} e^{-2x} dx = \\ &= 4e^{-4} + [-e^{-2x}]_2^{+\infty} = 5e^{-4}.\end{aligned}$$

iii)

$$\begin{aligned}\mathbf{P}(G) &= \mathbf{P}(X + Y + Z \leq 3) = \int_0^3 4x^2 e^{-2x} dx = \\ &= 1 - \int_3^{+\infty} 4x^2 e^{-2x} dx = \\ &= 1 - \left\{ [-2x^2 e^{-2x}]_3^{+\infty} + 4 \int_3^{+\infty} x e^{-2x} dx \right\} = \\ &= 1 - 25e^{-6},\end{aligned}$$

dove per l'integrale $4 \int_3^{+\infty} x e^{-2x} dx$ si è usato il calcolo svolto al punto precedente.

iv)

$$\begin{aligned}\mathbf{P}(EF) &= \mathbf{P}(X \leq 2, X + Y > 2) \\ &= \mathbf{P}(X \leq 2, Y > 2 - X) = \mathbf{P}(X \leq 2, Y > 0) \\ &= \mathbf{P}(X \leq 2)\mathbf{P}(Y > 0) = \mathbf{P}(X \leq 2).\end{aligned}$$

Qui si è usato il fatto che il prodotto di eventi indica il verificarsi di due condizioni simultanee ed anche che X e Y sono per ipotesi stocasticamente indipendenti.

d) Per determinare se E, F, G sono stocasticamente indipendenti, bisogna verificare tutte le seguenti condizioni:

$$\begin{aligned}\mathbf{P}(EF) &= \mathbf{P}(E)\mathbf{P}(F) \\ \mathbf{P}(EG) &= \mathbf{P}(E)\mathbf{P}(G) \\ \mathbf{P}(FG) &= \mathbf{P}(F)\mathbf{P}(G) \\ \mathbf{P}(EFG) &= \mathbf{P}(E)\mathbf{P}(F)\mathbf{P}(G).\end{aligned}$$

Basta che una sola non sia verificata, per non avere l'indipendenza. Si vede subito che

$$\mathbf{P}(EF) \neq \mathbf{P}(E)\mathbf{P}(F)$$

usando il punto precedente. Quindi i tre eventi non sono stocasticamente indipendenti. □

Esercizio 11.2. Sia X un numero aleatorio con distribuzione normale standard. Poniamo $Y = 3X + 2$ e $Z = X^2$.

1. Calcolare la funzione di ripartizione e la densità di Y .
2. Stimare $\mathbf{P}(Y \geq y)$, dove $y > 0$.
3. Calcolare previsione e varianza di Z .
4. Calcolare la funzione di ripartizione e la densità di Z .

Soluzione 11.2. 1. Poniamo

$$n(t) = \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}}$$

e calcoliamo la funzione di ripartizione F_Y di $Y = 3X + 2$. Dato $y \in \mathbb{R}$

$$F_Y(y) = \mathbf{P}(Y \leq y) = \mathbf{P}(3X + 2 \leq y)$$

$$\mathbf{P}\left(X \leq \frac{y-2}{3}\right) = \int_{-\infty}^{\frac{y-2}{3}} n(t) dt = \int_{-\infty}^y \frac{1}{3\sqrt{2\pi}} e^{-\frac{(z-2)^2}{18}} dz,$$

dove si è usato il cambio di variabile $z = \frac{t-2}{3}$. La densità f_Y di Y si ottiene come derivata di F_Y . Si utilizza la formula di derivazione di una funzione composta, ottenendo

$$f_Y(y) = \frac{d}{dy} F_Y(y) = \frac{1}{3} \frac{1}{\sqrt{2\pi}} e^{-\frac{(y-2)^2}{2 \cdot 9}}.$$

Ne segue che Y ha densità normale di parametri $N(2, 9)$.

2. Per stimare la probabilità $\mathbf{P}(Y \geq y)$, $y > 0$, si utilizza il fatto che

$$\frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \left(\frac{1}{x} - \frac{1}{x^3}\right) \leq \mathbf{P}(X \geq x) \leq \frac{1}{x} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}},$$

se X ha distribuzione gaussiana standard. Poiché $\mathbf{P}(Y \geq y) = \mathbf{P}(X \geq \frac{y-2}{3})$, $y > 0$, si ottiene

$$\frac{1}{\sqrt{2\pi}} e^{-\frac{(y-2)^2}{2 \cdot 9}} \left(\frac{3}{y-2} - \frac{27}{(y-2)^3}\right) \leq \mathbf{P}(Y > y) \leq \frac{3}{y-2} \frac{1}{\sqrt{2\pi}} e^{-\frac{(y-2)^2}{2 \cdot 9}}.$$

3. La previsione di Z è data da

$$\mathbf{P}(Z) = \mathbf{P}(X^2) = \int_{-\infty}^{+\infty} x^2 \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx = \sigma^2(X) = 1,$$

dove si è utilizzato la formula $\mathbf{P}(\psi(x)) = \int \psi(x) f_X(x) dx$. Per calcolare la varianza di Z , si usa la formula

$$\sigma^2(Z) = \mathbf{P}(Z^2) - \mathbf{P}(Z)^2.$$

Ci rimane da calcolare

$$\begin{aligned}
 \mathbf{P}(Z^2) &= \mathbf{P}((X^2)^2) = \mathbf{P}(X^4) \\
 &= \int_{-\infty}^{+\infty} x^4 \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx \\
 &= \left[-x^3 \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \right]_{-\infty}^{+\infty} + 3 \int_{-\infty}^{+\infty} x^2 \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx \\
 &= 3.
 \end{aligned}$$

4. Per calcolare la funzione di ripartizione F_Z di Z , si procede come nel primo punto, ovvero

$$F_Z(z) = \mathbf{P}(Z \leq z) = \mathbf{P}(X^2 \leq z).$$

Poiché $Z = X^2$ è un numero aleatorio sempre positivo, si possono distinguere due casi:

- a) Per $z < 0$ si ha che $F_Z(z) = 0$.
- b) Se invece $z \geq 0$

$$\begin{aligned}
 F_Z(z) &= \mathbf{P}(X^2 \leq z) \\
 &= \mathbf{P}(-\sqrt{z} \leq X \leq \sqrt{z}) \\
 &= \mathbf{P}(X \leq \sqrt{z}) - \mathbf{P}(X \leq -\sqrt{z}) \\
 &= \int_{-\infty}^{\sqrt{z}} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt - \int_{-\infty}^{-\sqrt{z}} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt \\
 &= \int_{-\sqrt{z}}^{\sqrt{z}} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt.
 \end{aligned}$$

Riassumendo

$$F_Z(z) = \begin{cases} 0 & z < 0 \\ \int_{-\sqrt{z}}^{\sqrt{z}} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt & z \geq 0. \end{cases}$$

Per calcolare la densità f_Z , si deriva la funzione di ripartizione. Per $z \geq 0$

$$\begin{aligned}
 f_Z(z) &= \frac{d}{dz} \left(\int_{-\infty}^{\sqrt{z}} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt - \int_{-\infty}^{-\sqrt{z}} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt \right) \\
 &= \frac{1}{2} z^{-\frac{1}{2}} n(\sqrt{z}) - \left(-\frac{1}{2} z^{-\frac{1}{2}} \right) n(-\sqrt{z}) \\
 &= z^{-\frac{1}{2}} \cdot n(\sqrt{z}) \\
 &= z^{-\frac{1}{2}} \frac{1}{\sqrt{2\pi}} e^{-\frac{z}{2}}.
 \end{aligned}$$

Si ottiene

$$f_Z(z) = \begin{cases} 0 & z < 0 \\ \frac{1}{\sqrt{2\pi}} z^{-\frac{1}{2}} e^{-\frac{z}{2}} & z \geq 0. \end{cases}$$

Quindi Z ha distribuzione gamma di parametri $\Gamma(\frac{1}{2}, \frac{1}{2})$, ovvero distribuzione chi-quadro di parametro 1.

□

Esercizio 11.3. Sia X un numero aleatorio con distribuzione esponenziale di parametro $\lambda = 2$.

1. Si calcolino i momenti di ordine n , ovvero $\mathbf{P}(X^n)$, $n \in \mathbb{N}$.
2. Consideriamo la famiglia di numeri aleatori $Z_u = e^{uX}$, $u < \lambda$. Fissato $u < \lambda$, calcolare la previsione $\Psi_X(u) = \mathbf{P}(e^{uX})$ di Z_u . La funzione $\Psi_X(u)$ si chiama *funzione di Laplace* di X .

Soluzione 11.3. 1. Il momento di ordine n si calcola sfruttando la formula

$$\mathbf{P}(\Psi(x)) = \int \Psi(x) f_X(x) dx,$$

dove $\Psi : \mathbb{R} \rightarrow \mathbb{R}$ è una funzione tale che l'integrale di cui sopra esiste finito. In questo caso, $\Psi(x) = x^n$. Si ottiene quindi

$$\mathbf{P}(X^n) = \int_0^{+\infty} x^n \lambda e^{-\lambda x} dx.$$

Per il calcolo di tale integrale ci si riconduce a quello della costante di normalizzazione della densità $\Gamma(n+1, \lambda)$, ovvero

$$\mathbf{P}(X^n) = \lambda \int_0^{+\infty} x^n e^{-\lambda x} dx = \lambda \frac{\Gamma(n+1)}{\lambda^{n+1}} = \frac{n!}{\lambda^n}.$$

In particolare per $n = 1$ si ha che $\mathbf{P}(X) = \frac{1}{\lambda}$.

2. Calcoliamo la previsione di $Z_u = e^{uX}$, $u \in \mathbb{R}$.

$$\begin{aligned} \mathbf{P}(Z_u) &= \mathbf{P}(e^{uX}) \\ &= \int_0^{+\infty} \lambda e^{ux} e^{-\lambda x} dx \\ &= \int_0^{+\infty} \lambda e^{(u-\lambda)x} dx. \end{aligned}$$

Si noti che u è da considerarsi come un parametro al fine di calcolare l'integrale e che questi è ben definito in quanto $u < \lambda$. Si ottiene che

$$\mathbf{P}(Z_u) = \frac{\lambda}{u-\lambda} \left[e^{(u-\lambda)x} \right]_0^{+\infty} = \frac{\lambda}{u-\lambda}.$$

□

Esercizio 11.4. Il numero aleatorio X ha distribuzione uniforme nell'intervallo $(-1, 1)$.

a) Scrivere la densità di X .

Si consideri $Z = \log |X|$.

b) Calcolare $I(Z)$ e $\mathbf{P}(Z)$.

c) Calcolare la funzione di ripartizione e la densità di Z .

d) Calcolare $\mathbf{P}(Z < -\frac{1}{2} | X > -\frac{1}{2})$.

Soluzione 11.4. a) La densità di X è

$$f(x) = \begin{cases} \frac{1}{2} & \text{per } x \in (-1, 1) \\ 0 & \text{altrimenti} \end{cases}.$$

b) Il numero aleatorio X ha come insieme dei valori possibili

$$I(X) = [-1, 1],$$

quindi $Z = \log |X|$ ha come insieme dei valori possibili

$$I(Z) = (-\infty, 0].$$

Il numero aleatorio Z non è definito se X assume il valore $0 \in I(X)$. Per calcolare la previsione si deve procedere nel modo seguente:

$$\begin{aligned} \mathbf{P}(Z) &= \mathbf{P}(Z | X > 0) \mathbf{P}(X > 0) + \mathbf{P}(Z | X < 0) \mathbf{P}(X < 0) \\ &= \mathbf{P}(\log X | X > 0) \cdot \frac{1}{2} - \mathbf{P}(\log(-X) | X < 0) \cdot \frac{1}{2}, \end{aligned}$$

dove si è usato il fatto che

$$\mathbf{P}(X > 0) = \mathbf{P}(X < 0) = \frac{1}{2}.$$

Verificare facendo il conto!

Basta dunque calcolare

$$\begin{aligned} \mathbf{P}(\log X | X > 0) &= \int_0^1 \log x dx \\ &= [x \log x - x]_0^1 = -1, \end{aligned} \tag{11.1}$$

$$\begin{aligned} \mathbf{P}(\log(-X) | X < 0) &= \int_{-1}^0 \log(-x) dx \\ &= \int_0^1 \log y dy = -1, \end{aligned} \tag{11.2}$$

da cui

$$\mathbf{P}(Z) = \mathbf{P}(\log X) = -1.$$

- c) Per calcolare la funzione di ripartizione di Z , bisogna di nuovo escludere il valore 0. Si ottiene:

$$\begin{aligned} F_Z(z) &= \mathbf{P}(Z \leq z) \\ &= \mathbf{P}(Z \leq z, X > 0) + \mathbf{P}(Z \leq z, X < 0). \end{aligned}$$

Se $z \geq 0$, allora $F_Z(z) = 1$. Sia $z < 0$. Si ottiene:

$$F_Z(z) = \mathbf{P}(\log X \leq z, X > 0) + \mathbf{P}(\log(-X) \leq z, X < 0).$$

Bisogna calcolare

$$\begin{aligned} \mathbf{P}(\log X \leq z, X > 0) &= \mathbf{P}(X \leq e^z, X > 0) & (11.3) \\ &= \mathbf{P}(0 < X \leq e^z) \\ &= \int_0^{e^z} \frac{1}{2} dx = \frac{1}{2} e^z. \end{aligned}$$

$$\begin{aligned} \mathbf{P}(\log(-X) \leq z, X < 0) &= \mathbf{P}(X \geq -e^z, X < 0) & (11.4) \\ &= \mathbf{P}(-e^z \leq X < 0) \\ &= \int_{-e^z}^0 \frac{1}{2} dx = \frac{1}{2} e^z. \end{aligned}$$

Quindi

$$F_Z(z) = e^z \quad \text{se } z < 0$$

La densità di Z è allora

$$f_Z(z) = \begin{cases} e^z & \text{se } z < 0 \\ 0 & \text{altrimenti.} \end{cases}$$

- d) Infine, $\mathbf{P}(Z < -\frac{1}{2} | X > -\frac{1}{2})$ si calcola con la formula della probabilità subordinata:

$$P\left(Z < -\frac{1}{2} \mid X > -\frac{1}{2}\right) = \frac{P\left(Z < -\frac{1}{2}, X > -\frac{1}{2}\right)}{\mathbf{P}\left(X > -\frac{1}{2}\right)},$$

dove

$$\begin{aligned} \mathbf{P}\left(Z < -\frac{1}{2}, X > -\frac{1}{2}\right) &= \mathbf{P}\left(\log|X| < -\frac{1}{2}, X > -\frac{1}{2}\right) = \\ &= \mathbf{P}\left(\log X < -\frac{1}{2}, X > 0\right) + \mathbf{P}\left(\log(-X) < -\frac{1}{2}, -\frac{1}{2} < X < 0\right), \end{aligned}$$

dove si è usato il fatto che l'evento

$$\left(X > -\frac{1}{2}\right) = (X > 0) + \left(-\frac{1}{2} < X < 0\right).$$

Ne segue che

$$\mathbf{P}\left(\log X < -\frac{1}{2}, X > 0\right) = \mathbf{P}\left(0 < X < e^{-\frac{1}{2}}\right) = \frac{e^{-\frac{1}{2}}}{2}$$

ed inoltre

$$\begin{aligned} \mathbf{P}\left(\log(-X) < -\frac{1}{2}, -\frac{1}{2} < X < 0\right) &= \mathbf{P}\left(X > -e^{-\frac{1}{2}}, -\frac{1}{2} < X < 0\right) \\ &= \mathbf{P}\left(-\frac{1}{2} < X < 0\right) = \int_{-\frac{1}{2}}^0 \frac{1}{2} dx = \frac{1}{4}. \end{aligned}$$

Quindi

$$\mathbf{P}\left(Z < -\frac{1}{2} | X > -\frac{1}{2}\right) = \frac{1}{2} \left(\frac{1}{\sqrt{e}} + \frac{1}{2}\right).$$

□

Distribuzioni assolutamente continue in più variabili

Esercizio 12.1. Sia X un numero aleatorio con densità

$$f(x) = \begin{cases} Kx^2 & \text{per } -1 \leq x \leq 1 \\ 0 & \text{altrimenti.} \end{cases}$$

- Calcolare K .
- Calcolare la funzione di ripartizione, la previsione e la varianza di X .
- Sia Y un numero aleatorio stocasticamente indipendente da X con densità esponenziale di parametro $\lambda = 2$. Scrivere la densità congiunta e la funzione di ripartizione di (X, Y) .

Soluzione 12.1. a) La costante K di normalizzazione è tale che

$$\int_{-1}^1 Kx^2 dx = 1.$$

Quindi

$$K = \frac{1}{\int_{-1}^1 x^2 dx} = \frac{3}{2}.$$

b) La funzione di ripartizione di X è data da

$$F(x) = \mathbf{P}(X \leq x) = \int_{-\infty}^x f(t) dt.$$

Quindi

$$F(x) = \begin{cases} 0 & \text{per } x \leq -1 \\ \int_{-1}^x \frac{3}{2} t^2 dt = \frac{1}{2}(x^3 + 1) & \text{per } x \in [-1, 1] \\ 1 & \text{per } x \geq 1. \end{cases}$$

Inoltre, la previsione di X è pari a

$$\mathbf{P}(X) = \int_{\mathbb{R}} tf(t) dt = \int_{-1}^1 \frac{3}{2} x^3 dx = 0.$$

La varianza è data da

$$\begin{aligned}\sigma^2(X) &= \mathbf{P}(X^2) - \mathbf{P}(X)^2 \\ &= \mathbf{P}(X^2) = \int_{-1}^1 x^2 \cdot \frac{3}{2}x^2 dx \\ &= \frac{3}{2} \int_{-1}^1 x^4 dx = \frac{3}{5}.\end{aligned}$$

c) La densità di Y è data da

$$g(y) = \begin{cases} 2e^{-2y} & \text{se } y \geq 0 \\ 0 & \text{altrimenti.} \end{cases}$$

Se X, Y sono stocasticamente indipendenti, allora la densità congiunta è data dal prodotto delle densità marginali:

$$f(x, y) = f_X(x)g_Y(y) = \begin{cases} 2e^{-2y} \frac{3}{2}x^2 = 3e^{-2y}x^2 & \text{se } x \in [-1, 1], y \geq 0 \\ 0 & \text{altrimenti.} \end{cases}$$

Analogamente, la funzione di ripartizione congiunta coincide con il prodotto delle marginali:

$$F(x, y) = F_X(x)F_Y(y) = \begin{cases} (1 - e^{-2y}) \frac{x^3 + 1}{2} & \text{se } x \in [-1, 1] \text{ e } y \geq 0 \\ 1 - e^{-2y} & \text{se } x > 1 \text{ e } y \geq 0 \\ 0 & \text{altrimenti.} \end{cases}$$

□

Esercizio 12.2. Sia (X, Y) un vettore aleatorio con distribuzione uniforme sul disco unitario di centro l'origine degli assi.

1. Si calcoli la densità congiunta $f(x, y)$ di (X, Y) .
2. Qual è la densità marginale f_X di X ?
3. Sia $Z = X^2 + Y^2$, calcolare $\mathbb{P}(\frac{1}{4} \leq Z \leq 1)$.
4. Calcolare la funzione di ripartizione e la densità di Z .

Soluzione 12.2. 1. Poiché (X, Y) hanno distribuzione congiunta uniforme sul disco

$$D_1 = \{(x, y) : x^2 + y^2 \leq 1\}.$$

la densità congiunta $f(x, y)$ è costante su D_1 e 0 fuori. Si ottiene che

$$f(x, y) = \begin{cases} \frac{1}{\text{area } D_1} = \frac{1}{\pi} & (x, y) \in D_1 \\ 0 & \text{altrimenti.} \end{cases}$$

Il dominio della densità è rappresentato in figura 12.1.

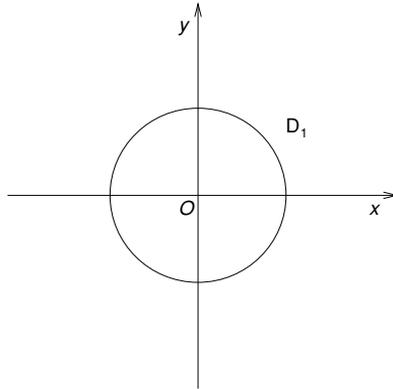


Figura 12.1. Rappresentazione dell'area D_1 sul piano.

Il valore di f su D_1 si determina imponendo la condizione

$$1 = \int \int_{\mathbb{R}^2} f(x, y) dx dy = \int \int_{D_1} c dx dy,$$

da cui

$$c = \frac{1}{\int \int_{D_1} dx dy} = \frac{1}{\text{area } D_1} = \frac{1}{\pi}.$$

2. Per calcolare la densità marginale di X , si distinguono quattro casi:

- Caso $x > 1$: $f_X(x) = 0$.
- Caso $1 \geq x \geq 0$: fissata la coordinata x , la y varia lungo la retta ortogonale all'asse delle ascisse e passante per la x . Gli estremi sono i punti dove tale retta interseca il grafico di D_1 , come mostrato in figura 12.2.

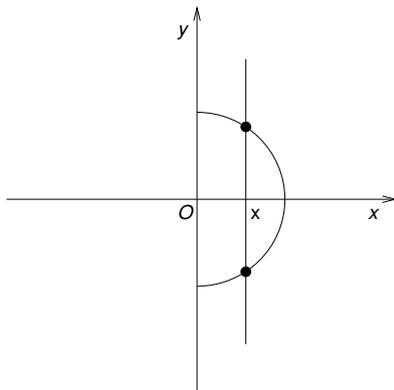


Figura 12.2. Caso $1 \geq x \geq 0$.

Si ottiene

$$f_X(x) = \int_{-\sqrt{1-x^2}}^{\sqrt{1-x^2}} f(x, t) dt = \int_{-\sqrt{1-x^2}}^{\sqrt{1-x^2}} \frac{1}{\pi} dt = \frac{2\sqrt{1-x^2}}{\pi}.$$

- Caso $-1 \leq x < 0$: per simmetria si ottiene, come mostrato in figura 12.3,

$$f_X(x) = \frac{2\sqrt{1-x^2}}{\pi}.$$

- Caso $x < -1$: anche qui si ha $f_X(x) = 0$.

Riassumendo

$$f_X(x) = \begin{cases} \frac{2\sqrt{1-x^2}}{\pi} & x \in [-1, 1] \\ 0 & \text{altrimenti.} \end{cases}$$

3. Consideriamo $Z = X^2 + Y^2$; calcolare $\mathbf{P}\left(\frac{1}{4} \leq Z \leq 1\right)$ significa calcolare la probabilità che il vettore aleatorio (X, Y) appartenga alla regione A del piano compresa fra il disco di centro O e raggio $\frac{1}{2}$ ed il disco di centro O e di raggio 1.

$$\mathbf{P}\left(\frac{1}{4} \leq Z \leq 1\right) = \mathbf{P}\left(\frac{1}{4} \leq X^2 + Y^2 \leq 1\right).$$

Si ottiene

$$\mathbf{P}\left(\frac{1}{4} \leq X^2 + Y^2 \leq 1\right) = \int \int_A f(x, y) dx dy.$$

Il calcolo di tale integrale è più facile se si passa alle coordinate polari

$$x = \rho \cos \theta, \quad y = \rho \sin \theta.$$

Per cambiare le coordinate all'interno dell'integrale bisogna utilizzare il modulo del determinante Jacobiano. Nel caso delle coordinate polari questo è

$$|J| = \rho.$$

Ne segue

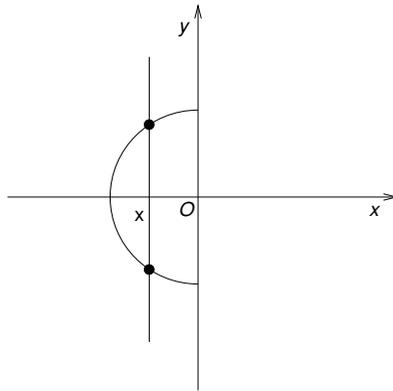


Figura 12.3. Caso $-1 \leq x < 0$.

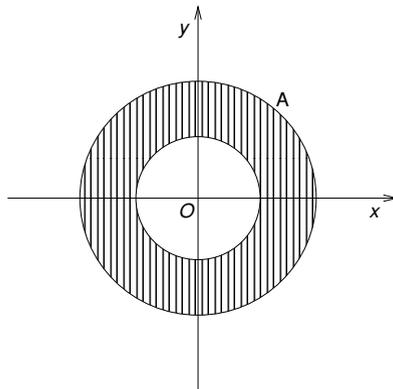


Figura 12.4. Area della regione $\{(x, y) \mid \frac{1}{4} \leq x^2 + y^2 \leq 1\}$.

$$\begin{aligned} \int \int_{A(x,y)} f(x,y) \, dx \, dy &= \int \int_{A(\rho,\theta)} f(\rho,\theta) \, d\rho \, d\theta = \\ &= \int_0^{2\pi} d\theta \int_{\frac{1}{2}}^1 \frac{1}{\pi} \, d\rho = \int_{\frac{1}{2}}^1 2\rho \, d\rho = [\rho^2]_{\frac{1}{2}}^1 = \frac{3}{4}. \end{aligned}$$

4. Per calcolare la funzione di ripartizione $F_Z(z)$ di Z si utilizza di nuovo la simmetria sferica.

- ($z < 0$) In questo caso $F_Z(z) = 0$.
- ($1 \geq z \geq 0$)

$$\begin{aligned} F_Z(z) &= \mathbf{P}(Z \leq z) \\ &= \mathbf{P}(X^2 + Y^2 \leq z) \\ &= \int \int_{D_z} f(x,y) \, dx \, dy, \end{aligned}$$

dove $D_z = \{(x,y) : x^2 + y^2 \leq z\}$. Ne segue che

$$F_Z(z) = \int_0^{2\pi} \int_0^{\sqrt{z}} \frac{1}{\pi} \rho \, d\rho \, d\theta = \int_0^{\sqrt{z}} 2\rho \, d\rho = [\rho^2]_0^{\sqrt{z}} = z.$$

- ($z > 1$) In questo caso $F_Z(z) = P(X^2 + Y^2 \leq z) = 1$.

Riassumendo

$$F_Z(z) = \begin{cases} 0 & z < 0 \\ z & 0 \leq z < 1 \\ 1 & z > 1. \end{cases}$$

La funzione di densità di Z è

$$f_Z(z) = \begin{cases} 1 & 0 \leq z \leq 1 \\ 0 & \text{altrimenti.} \end{cases}$$

Il numero aleatorio Z ha quindi densità uniforme nell'intervallo $[0, 1]$. \square

Esercizio 12.3. Sia (X, Y) un vettore aleatorio con densità congiunta

$$f(x, y) = \begin{cases} kxy & (x, y) \in T \\ 0 & \text{altrimenti.} \end{cases}$$

dove $T = \{(x, y) \in \mathbb{R}^2 \mid 0 \leq y \leq (-x + 2), 0 < x < 2\}$.

1. Calcolare la costante di normalizzazione k .
2. Si calcolino la probabilità congiunta $\mathbf{P}(X > 1, Y < \frac{1}{2})$ e la probabilità subordinata $\mathbf{P}(X > 1 \mid Y < \frac{1}{2})$.
3. Sia $Z = X + Y$. Si calcoli la probabilità $\mathbf{P}(0 < Z < 1)$.
4. Calcolare la funzione di ripartizione e la densità di Z .

Soluzione 12.3. 1. Per calcolare la costante di normalizzazione k si impone che

$$\int \int_{\mathbb{R}^2} f(x, y) \, dx \, dy = 1.$$

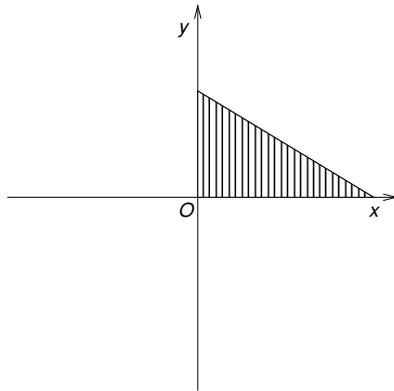


Figura 12. . Rappresentazione dell'area T nel piano.

Si calcola l'integrale di f utilizzando il teorema di Fubini-Tonelli:

$$\begin{aligned} \int \int_{\mathbb{R}^2} f(x, y) \, dx \, dy &= k \int_0^2 x \int_0^{-x+2} y \, dy \, dx = \\ &= k \int_0^2 x \left[\frac{y^2}{2} \right]_0^{-x+2} dx = k \int_0^2 x \frac{1}{2} (2-x)^2 dx = \\ &= \frac{k}{2} \int_0^2 (4x - 4x^2 + x^3) dx = \frac{k}{2} \left[2x^2 - \frac{4}{3}x^3 + \frac{1}{4}x^4 \right]_0^2 = \frac{2}{3} k. \end{aligned}$$

Ne segue che

$$k = \frac{3}{2}.$$

2. La probabilità congiunta $\mathbf{P}(X > 1, Y < \frac{1}{2})$ è data dall'integrale della densità congiunta sulla regione D di piano individuata dall'intersezione

$$D = \{(x, y) \in \mathbb{R}^2 \mid x > 1, y < \frac{1}{2}\} \cap T.$$

come si può osservare nella figure 12.6.

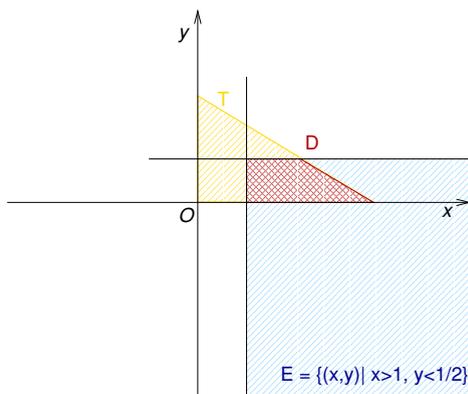


Figura 12.6. Rappresentazione dell'area D nel piano.

Si utilizza il metodo delle rette normali per trovare gli estremi dell'integrale. In questo caso è più semplice fissare le y e vedere come varia la x . Gli estremi di integrazione sono dati dall'intersezione fra i bordi del dominio D e la retta passante per y , come si può osservare nella figura 12.7.

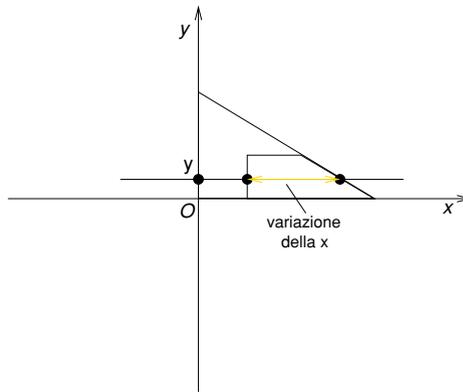


Figura 12.7. Applicazione del metodo delle rette normali sull'area D .

$$\begin{aligned}
 \mathbf{P}(X > 1, Y < \frac{1}{2}) &= \int \int_D f(x, y) \, dx \, dy \\
 &= \int_0^{\frac{1}{2}} \left(y \int_1^{-y+2} \frac{3}{2} x \, dx \right) \, dy \\
 &= \int_0^{\frac{1}{2}} y \left[\frac{3}{4} x^2 \right]_1^{-y+2} \, dy \\
 &= \frac{3}{4} \int_0^{\frac{1}{2}} y (3 - 4y + y^2) \, dy \\
 &= \frac{3}{4} \left[\frac{3}{2} y^2 - \frac{4}{3} y^3 + \frac{1}{4} y^4 \right]_0^{\frac{1}{2}} \\
 &= \frac{43}{256}.
 \end{aligned}$$

La probabilità subordinata $\mathbf{P}(X > 1|Y < \frac{1}{2})$ si ottiene come

$$\mathbf{P}(X > 1|Y < \frac{1}{2}) = \frac{\mathbf{P}(X > 1, Y < \frac{1}{2})}{\mathbf{P}(Y < \frac{1}{2})}.$$

Basta quindi calcolare la probabilità $\mathbf{P}(Y < \frac{1}{2})$. A tal fine, non è necessario conoscere la densità marginale delle Y . Basta riflettere che tale probabilità è data dall'integrale della densità *congiunta* $f(x, y)$ sul dominio D_1 dato dall'intersezione di $E_1 = \{(x, y) \in \mathbb{R}^2 | y < \frac{1}{2}\}$ e di T

$$D_1 = E_1 \cap T.$$

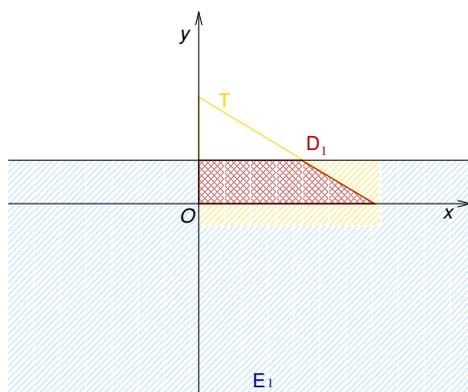


Figura 12.8. Rappresentazione nel piano dell'area D_1 .

Calcolare la probabilità che il numero aleatorio Y sia minore di 1 è equivalente a calcolare la probabilità congiunta che Y sia minore di 1 senza imporre alcuna condizione sulla X . Si ottiene

$$\begin{aligned} \mathbf{P}(Y < \frac{1}{2}) &= \int \int_{D_1} f(x, y) \, dx \, dy = \\ \int_0^{\frac{1}{2}} \frac{3}{2} y \int_0^{-y+2} x \, dx \, dy &= \frac{3}{4} \int_0^{\frac{1}{2}} y (4 - 4y + y^2) \, dy = \\ \frac{3}{4} \left[2y^2 - \frac{4}{3}y^3 + \frac{1}{4}y^4 \right]_0^{\frac{1}{2}} &= \frac{67}{256}. \end{aligned}$$

La probabilità subordinata è data da

$$\mathbf{P}(X > 1 | Y < \frac{1}{2}) = \frac{\mathbf{P}(X > 1, Y < \frac{1}{2})}{\mathbf{P}(Y < \frac{1}{2})} = \frac{43}{67}.$$

3. Consideriamo il numero aleatorio $Z = X + Y$. Per calcolare la probabilità $\mathbf{P}(0 < Z < 1)$ si può utilizzare la densità congiunta di (X, Y) . Si ottiene

$$\begin{aligned} \mathbf{P}(0 < Z < 1) &= \mathbf{P}(0 < X + Y < 1) \\ &= \mathbf{P}(-Y < X < 1 - Y) \\ &= \mathbf{P}(0 < X < 1 - Y). \end{aligned}$$

Si noti che, in questo caso, X e Y sono entrambe positive, quindi la condizione $X > -Y$ si riduce a $X > 0$. Dalla figura 12.9 si ricava la regione su cui è necessario integrare la densità congiunta di X, Y .

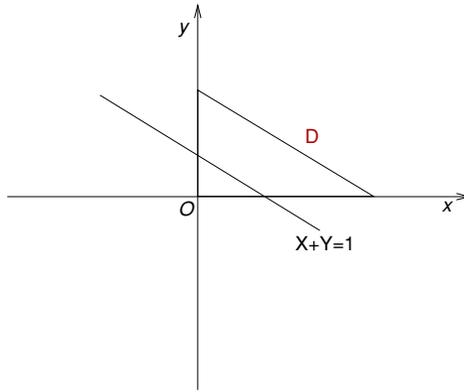


Figura 12.9. Regione dove $0 < Z < 1$.

$$\begin{aligned}
 \mathbf{P}(0 < X < 1 - Y) &= \int_0^1 \frac{3}{2} y \int_0^{1-y} x \, dx \, dy \\
 &= \frac{3}{4} \int_0^1 y (1-y)^2 \, dy \\
 &= \frac{3}{4} \int_0^1 (y - 2y^2 + y^3) \, dy \\
 &= \frac{3}{4} \left[\frac{1}{2} y^2 - \frac{2}{3} y^3 + \frac{1}{4} y^4 \right]_0^1 \\
 &= \frac{1}{16}.
 \end{aligned}$$

4. La funzione di ripartizione di Z è data da

$$F_Z(z) = \mathbf{P}(Z \leq z) = \mathbf{P}(X + Y \leq z) = \mathbf{P}(X \leq z - Y).$$

Se si considera la retta $x + y - z = 0$, la funzione di ripartizione di Z è data dall'integrale della densità congiunta di (X, Y) calcolato sulla regione di piano intercettata da tale retta sul triangolo T come da figura 12.10.

Si ottiene:

- Per $z < 0$: $\mathbf{P}(Z < z) = 0$.
- Per $z > 2$: $\mathbf{P}(Z < z) = 1$.
- Per $0 \leq z \leq 2$:

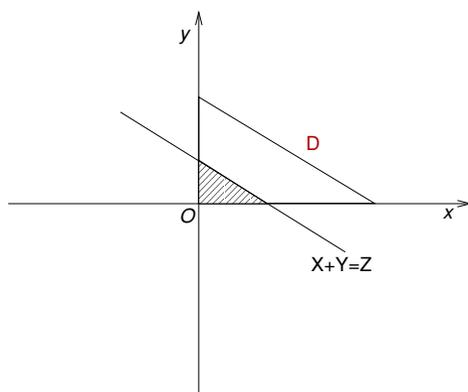


Figura 12.10. Regione di piano intercettata sul triangolo T .

$$\begin{aligned}
 \mathbf{P}(Z < z) &= \int_0^z \frac{3}{2} y \int_0^{z-y} x \, dx \, dy \\
 &= \frac{3}{4} \int_0^z y (z-y)^2 \, dy \\
 &= \frac{3}{4} \int_0^z (z^2 y - 2zy^2 + y^3) \, dy \\
 &= \frac{3}{4} \left[\frac{1}{2} z^2 y^2 - \frac{2}{3} zy^3 + \frac{1}{4} y^4 \right]_0^z \\
 &= \frac{3}{4} \left(\frac{1}{2} z^4 - \frac{2}{3} z^4 + \frac{1}{4} z^4 \right) \\
 &= \frac{z^4}{16}.
 \end{aligned}$$

Riassumendo

$$F_Z(z) = \begin{cases} 0 & z < 0 \\ \frac{z^4}{16} & 0 \leq z \leq 2 \\ 1 & z > 2. \end{cases}$$

La densità si ottiene come derivata della funzione di ripartizione

$$f_Z(z) = \begin{cases} 0 & z < 0 \\ \frac{z^3}{4} & 0 \leq z \leq 2 \\ 1 & z > 2 \end{cases}$$

o mediante la formula

$$f_Z(z) = \int_{\mathbb{R}} f(x, z-x) dx.$$

□

Esercizio 12.4. Siano X, Y due numeri aleatori con distribuzione congiunta

$$f(x, y) = \begin{cases} Kx & \text{per } y \leq x \leq y + 1, 0 \leq y \leq 2 \\ 0 & \text{altrimenti.} \end{cases}$$

- Calcolare K .
- Calcolare la densità marginale e la previsione di X .
- Calcolare $\text{cov}(X, Y)$.
- Calcolare $\mathbf{P}(0 < X - Y < 1)$.

Soluzione 12.4. a) Come negli esercizi precedenti, per prima cosa si disegna il grafico della regione R di definizione della densità congiunta come mostrato in figura 12.11.

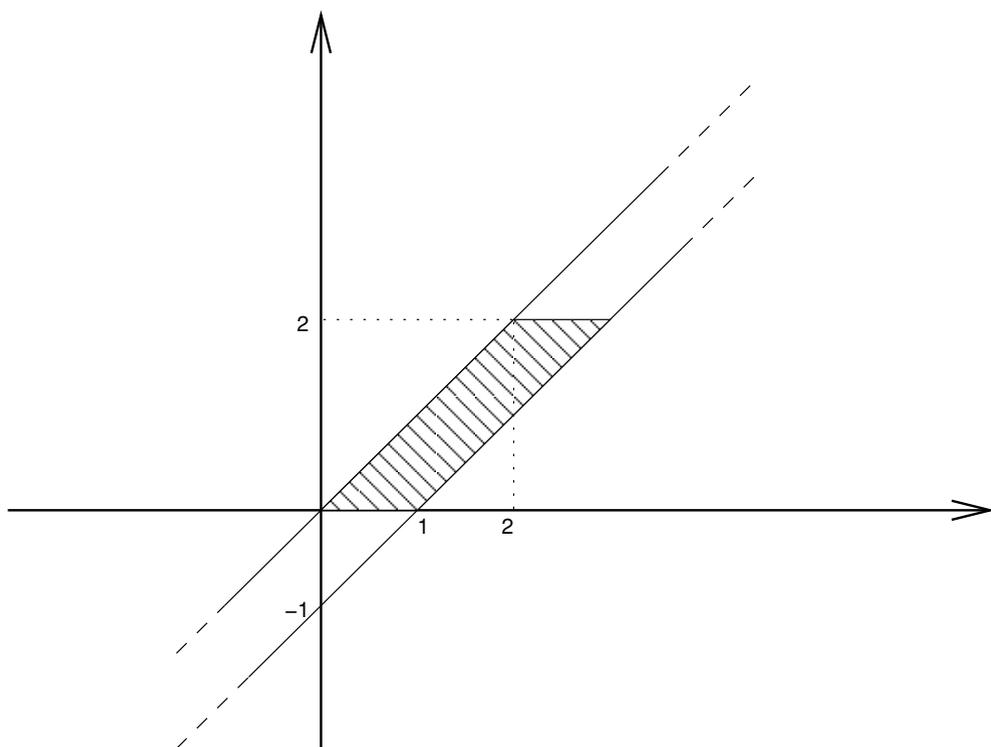


Figura 12.11. Regione di definizione della densità.

La costante K (o di normalizzazione) rende pari a 1 l'integrale della funzione che dà la densità, quindi

$$\int \int_{\mathbb{R}^2} f(x, y) dx dy = 1 \implies K = \frac{1}{\int \int_{\mathbb{R}^2} x dx dy},$$

dove

$$\begin{aligned} \int \int_{\mathbb{R}^2} x dx dy &= \int_0^2 dy \int_y^{y+1} x dx \\ &= \int_0^2 \left[\frac{x^2}{2} \right]_y^{y+1} dy \\ &= \int_0^2 \left(\frac{(y+1)^2}{2} - \frac{y^2}{2} \right) dy \\ &= \frac{1}{6} [(y+1)^3 - y^3]_0^2 \\ &= \frac{1}{6} (27 - 8 - 1) = 3. \end{aligned}$$

Si conclude che $K = \frac{1}{3}$.

b) Per calcolare la densità marginale di X si applica la formula

$$f_X(x) = \int_{\mathbb{R}} f(x, y) dy.$$

Per trovare gli estremi di integrazione si usa il metodo delle rette normali come mostrato in figura 12.12.

Bisogna però fare attenzione al fatto che gli estremi di integrazione cambiano a seconda che $0 < x < 1$, $1 < x < 2$, $2 < x < 3$ (si veda la figura 12.13).

Si ha che se $0 < x < 1$, allora la y varia fra le rette

$$y = 0 \quad \text{e} \quad y = x$$

Se invece $1 < x < 2$, allora la y varia fra le rette

$$y = x - 1 \quad \text{e} \quad y = x$$

Infine se $2 < x < 3$, la y varia fra

$$y = x - 1 \quad \text{e} \quad y = 2.$$

– Per $0 < x < 1$:

$$f_X(x) = \int_0^x \frac{1}{3} x dy = \frac{1}{3} x^2.$$

– Per $1 < x < 2$:

$$f_X(x) = \int_{x-1}^x \frac{1}{3} x dy = \frac{1}{3} x.$$

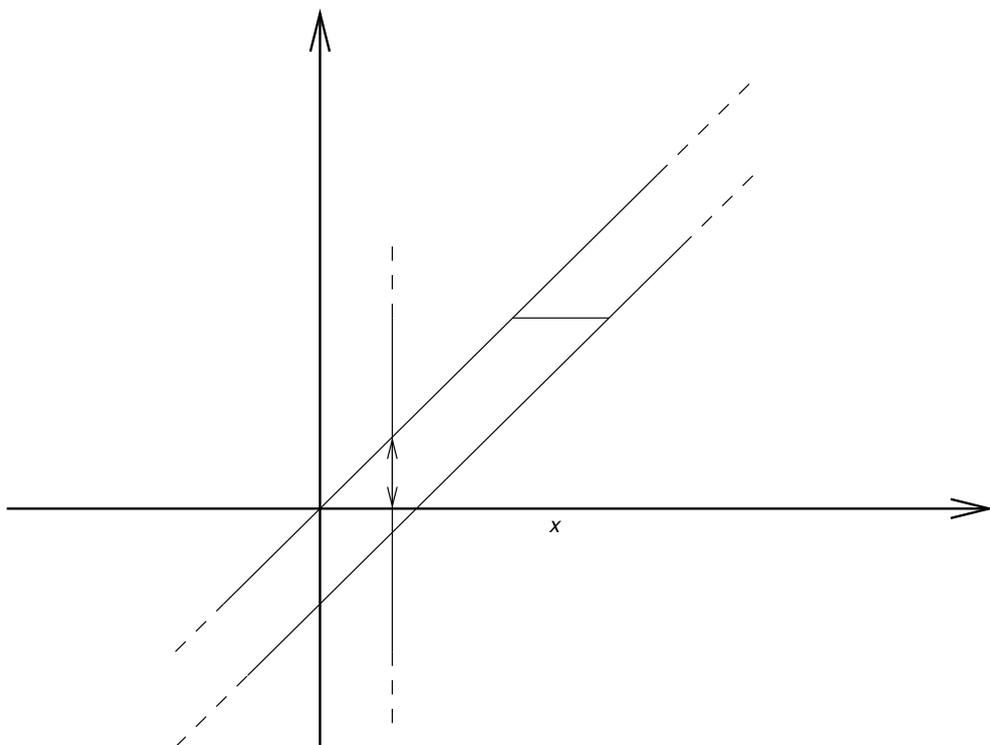


Figura 12.12. Metodo delle rette normali.

– Per $2 < x < 3$:

$$f_X(x) = \int_{x-1}^2 \frac{1}{3}x dy = \frac{1}{3}x(3-x).$$

Riassumendo

$$f_X(x) = \begin{cases} \frac{1}{3}x^2 & \text{per } 0 < x < 1 \\ \frac{1}{3}x & \text{per } 1 < x < 2 \\ \frac{1}{3}x(3-x) & \text{per } 2 < x < 3 \\ 0 & \text{altrimenti.} \end{cases}$$

Si verifica ora che $f_X(x)$ è una densità di probabilità. Deve essere che

$$\int_{\mathbb{R}} f_X(x) dx = 1,$$

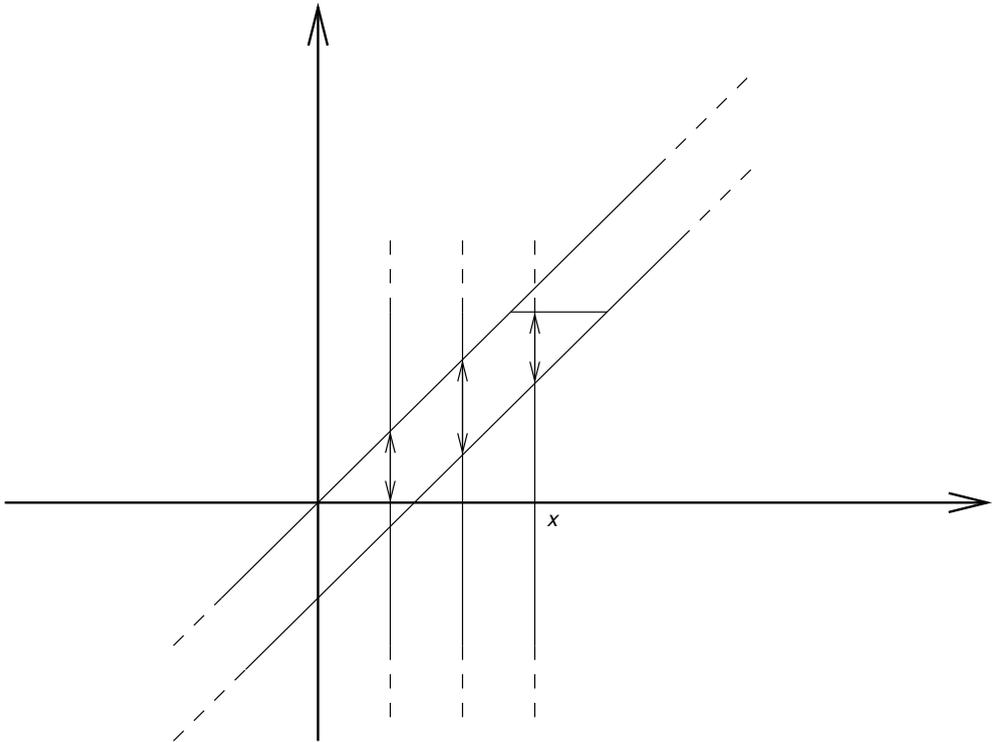


Figura 12.13. Attenzione agli intervalli!

Quindi

$$\begin{aligned}
 \int_0^1 \frac{1}{3}x^2 dx + \int_1^2 \frac{1}{3}x dx + \int_2^3 \frac{1}{3}x(3-x) dx &= \\
 &= \left[\frac{1}{9}x^3 \right]_0^1 + \left[\frac{1}{6}x^2 \right]_1^2 + \left[\frac{x^2}{2} - \frac{x^3}{9} \right]_2^3 \\
 &= \frac{1}{9} + \frac{1}{2} + \frac{9}{2} - 3 - 2 + \frac{8}{9} \\
 &= 1
 \end{aligned}$$

La previsione di X è data da:

$$\begin{aligned}
\mathbf{P}(X) &= \int_{\mathbb{R}} xf(x)dx \\
&= \int_0^1 x \frac{1}{3}x^2 dx + \int_1^2 x \frac{1}{3}x dx + \int_2^3 x \frac{1}{3}x(3-x)dx \\
&= \left[\frac{1}{12}x^4 \right]_0^1 + \left[\frac{1}{9}x^3 \right]_1^2 + \left[\frac{x^3}{3} - \frac{x^4}{12} \right]_2^3 \\
&= \frac{1}{12} + \frac{7}{9} + 9 - \frac{27}{4} - \frac{8}{3} + \frac{4}{3} \\
&= \frac{16}{9}.
\end{aligned}$$

c) La covarianza $\mathbf{cov}(X, Y)$ è data da:

$$\mathbf{cov}(X, Y) = \mathbf{P}(XY) - \mathbf{P}(X)\mathbf{P}(Y),$$

dove

$$\begin{aligned}
\mathbf{P}(XY) &= \int \int_{\mathbb{R}^2} xyf(x, y)dxdy = \\
&= \int_0^2 dy \int_y^{y+1} xy \frac{1}{3}x dx = \int_0^2 \frac{1}{9}y [(y+1)^3 - y^3] dy \\
&= \left[\frac{1}{12}y^4 + \frac{1}{9}y^3 + \frac{1}{18}y^2 \right]_0^2 = \frac{22}{9}.
\end{aligned}$$

Per calcolare la previsione di Y non importa calcolare prima la densità marginale di Y . Vale infatti che

$$\begin{aligned}
\mathbf{P}(Y) &= \int_{\mathbb{R}} yf_Y(y)dy \\
&= \int_{\mathbb{R}} \left[y \int_{\mathbb{R}} f(x, y)dx \right] dy \\
&= \int \int_{\mathbb{R}} yf(x, y)dxdy.
\end{aligned}$$

Quindi si può calcolare la previsione di Y direttamente:

$$\begin{aligned}
\mathbf{P}(Y) &= \int \int_{\mathbb{R}} yf(x, y)dxdy \\
&= \int_0^2 dy \int_y^{y+1} \frac{1}{3}xy dx \\
&= \int_0^2 y \frac{1}{6} [(y+1)^2 - y^2] dy \\
&= \left[\frac{y^3}{6} + \frac{y^2}{12} \right]_0^2 = \frac{4}{3} + \frac{1}{3} = \frac{5}{3}.
\end{aligned}$$

Si ottiene infine

$$\mathbf{cov}(X, Y) = \frac{22}{9} - \frac{11}{12} \cdot \frac{5}{3} = \frac{11}{12}.$$

X e Y sono quindi positivamente correlati.

d) Per calcolare $\mathbf{P}(0 < X - Y < 1)$, si osserva che

$$\mathbf{P}(0 < X - Y < 1) = \mathbf{P}(Y < X < Y + 1) = 1,$$

in quanto la regione del piano

$$I_1 = \{(x, y) \mid y < x < y + 1\}$$

contiene interamente il dominio di definizione della densità (figura 12.14).

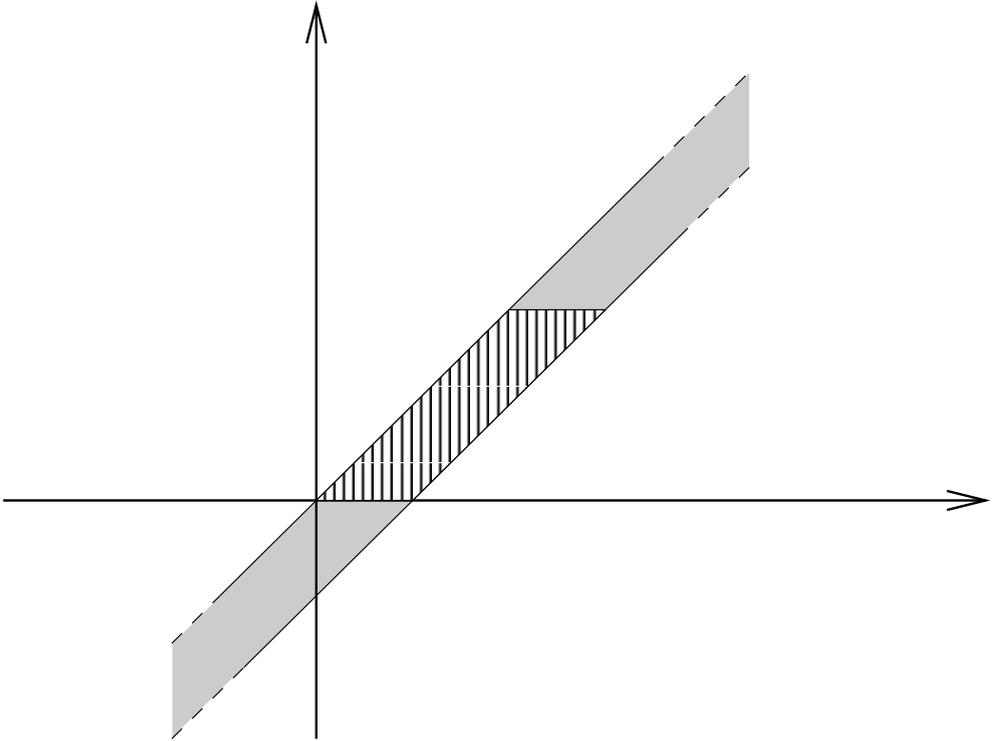


Figura 12.14. La regione I_1 .

□

Esercizio 12.5. Siano X, Y due numeri aleatori stocasticamente indipendenti con la seguente funzione di densità:

$$f(x) = \begin{cases} K(x^3 - 1) & \text{per } 1 \leq x \leq 2 \\ 0 & \text{altrimenti} \end{cases}$$

- Calcolare K .
- Calcolare la funzione di ripartizione congiunta, previsione, varianza e covarianza di X e Y .
- Sia $Z = X^2$. Calcolare la funzione di ripartizione, previsione e varianza di Z .
- Calcolare i coefficienti di correlazione $\rho(X, Z)$, $\rho(X + Y, Z)$.

Soluzione 12.5. a) Si calcola K usando il fatto che è la costante di normalizzazione. Si ottiene

$$K = \frac{1}{\int_0^2 (x^3 - 1) dx} = \frac{1}{\left[\frac{x^4}{4} - x\right]_1^2} = \frac{4}{11}.$$

- I numeri aleatori X ed Y sono stocasticamente indipendenti quindi la loro funzione di ripartizione congiunta è data dal prodotto delle funzioni di ripartizione marginali:

$$F(x, y) = \mathbf{P}(X \leq x, Y \leq y) = F_X(x)F_Y(y).$$

Basta quindi calcolare

$$F(x) = \mathbf{P}(X \leq x) = \int_{-\infty}^x f(t) dt.$$

Si ottiene

$$F_X(x) = \begin{cases} 0 & x < 1 \\ \int_1^x \frac{4}{11}(t^3 - 1) dt = \frac{4}{11} \left(\frac{x^4}{4} - x + \frac{3}{4} \right) & x \in [1, 2] \\ 0 & x \geq 2 \end{cases}$$

Quindi

$$F(x, y) = \begin{cases} 0 & x < 1 \text{ o } y < 1 \\ \left(\frac{4}{11}\right)^2 (x^4 - x + \frac{3}{4})(y^4 - y + \frac{3}{4}) & (x, y) \in [1, 2] \times [1, 2] \\ \frac{4}{11}(x^4 - x + \frac{3}{4}) & \text{se } x \in [1, 2], y > 2 \\ \frac{4}{11}(y^4 - y + \frac{3}{4}) & \text{se } x > 2, y \in [1, 2] \\ 1 & x > 2, y > 2. \end{cases}$$

Poiché X ed Y sono stocasticamente indipendenti, si può dire subito che

$$\mathbf{cov}(X, Y) = 0.$$

Infine, si calcolano previsione e varianza usando le note formule.

1. Previsione

$$\begin{aligned} \mathbf{P}(X) &= \mathbf{P}(Y) = \int_{\mathbb{R}} tf(t)dt = \\ \frac{4}{11} \int_1^2 t(t^3 - 1)dt &= \frac{4}{11} \left[\frac{t^5}{5} - \frac{t^2}{2} \right]_1^2 = \frac{4}{11} \cdot \frac{47}{10} = \frac{94}{55}. \end{aligned}$$

2. Per la varianza, bisogna calcolare prima la previsione quadratica

$$\begin{aligned} \mathbf{P}(X^2) &= \mathbf{P}(Y^2) = \frac{4}{11} \int_1^2 t^2(t^3 - 1)dt = \\ \frac{4}{11} \left[\frac{t^6}{6} - \frac{t^3}{3} \right]_1^2 &= \frac{4}{11} \cdot \frac{49}{6} = \frac{98}{33}. \end{aligned}$$

Quindi

$$\sigma^2(X) = \mathbf{P}(X^2) - \mathbf{P}(X)^2 = \frac{98}{33} - \left(\frac{94}{55} \right)^2.$$

c) Calcoliamo la funzione di ripartizione di $Z = X^2$.

$$F(z) = \mathbf{P}(Z \leq z) = \mathbf{P}(X^2 \leq z).$$

La previsione di Z coincide con la previsione quadratica di X . Infatti

$$\mathbf{P}(Z) = \mathbf{P}(X^2) = \frac{98}{33}.$$

Per calcolare la varianza, si nota che

$$\begin{aligned} \mathbf{P}(Z^2) &= \mathbf{P}(X^4) = \int_1^2 \frac{4}{11} t^4(t^3 - 1)dt \\ &= \frac{4}{11} \left[\frac{t^8}{8} - \frac{t^5}{5} \right]_1^2 \\ &= \frac{4}{11} \cdot \frac{1027}{40} = \frac{1027}{110}. \end{aligned}$$

La varianza è quindi data da

$$\sigma^2(Z) = \mathbf{P}(Z^2) - \mathbf{P}(Z)^2 = \frac{1027}{110} - \left(\frac{98}{33} \right)^2.$$

□

Esercizio 12.6. I numeri aleatori X ed Y sono stocasticamente indipendenti. La densità di probabilità $f_X(x)$ di X è data da:

$$f_X(x) = \begin{cases} 2x & \text{per } 0 \leq x \leq 1 \\ 0 & \text{altrimenti,} \end{cases}$$

mentre la densità di probabilità di Y è

$$f_Y(y) = \begin{cases} e^{-y} & \text{per } y \geq 0 \\ 0 & \text{altrimenti.} \end{cases}$$

- Calcolare $\mathbf{P}(X)$, $\mathbf{P}(Y)$, $\sigma^2(X)$, $\sigma^2(Y)$.
- Scrivere la densità congiunta e la funzione di ripartizione congiunta di (X, Y) .
- Sia $Z = X + Y$. Calcolare $\mathbf{P}(Z)$, $\sigma^2(Z)$ e la funzione di ripartizione e la densità di Z .

Soluzione 12.6. a) Calcoliamo i momenti di X e di Y usando le formule note.

$$\mathbf{P}(X) = \int_{\mathbb{R}} x f_X(x) dx = \int_0^1 2x^2 dx = \frac{2}{3}.$$

$$\begin{aligned} \sigma^2(X) &= \mathbf{P}(X^2) - \mathbf{P}(X)^2 \\ &= \int_0^1 2x^3 dx - \frac{4}{9} = \frac{1}{18}. \end{aligned}$$

Il numero aleatorio Y ha densità esponenziale di parametro $\lambda = 1$, quindi si può scrivere subito

$$\begin{aligned} \mathbf{P}(Y) &= \frac{1}{\lambda} = 1, \\ \sigma^2(Y) &= \frac{1}{\lambda^2} = 1. \end{aligned}$$

- I numeri aleatori X e Y sono stocasticamente indipendenti, quindi la loro densità congiunta è

$$f(x, y) = f_X(x) f_Y(y),$$

ovvero

$$f(x, y) = \begin{cases} 2xe^{-y} & \text{per } 0 \leq x \leq 1 \text{ e } y \geq 0 \\ 0 & \text{altrimenti.} \end{cases}$$

Calcoliamo la funzione di ripartizione congiunta

$$F(x, y) = \int_{-\infty}^x \int_{-\infty}^y f(s, t) ds dt$$

dopo aver disegnato il dominio di definizione della densità come da figura 12.15.

Si ottiene che

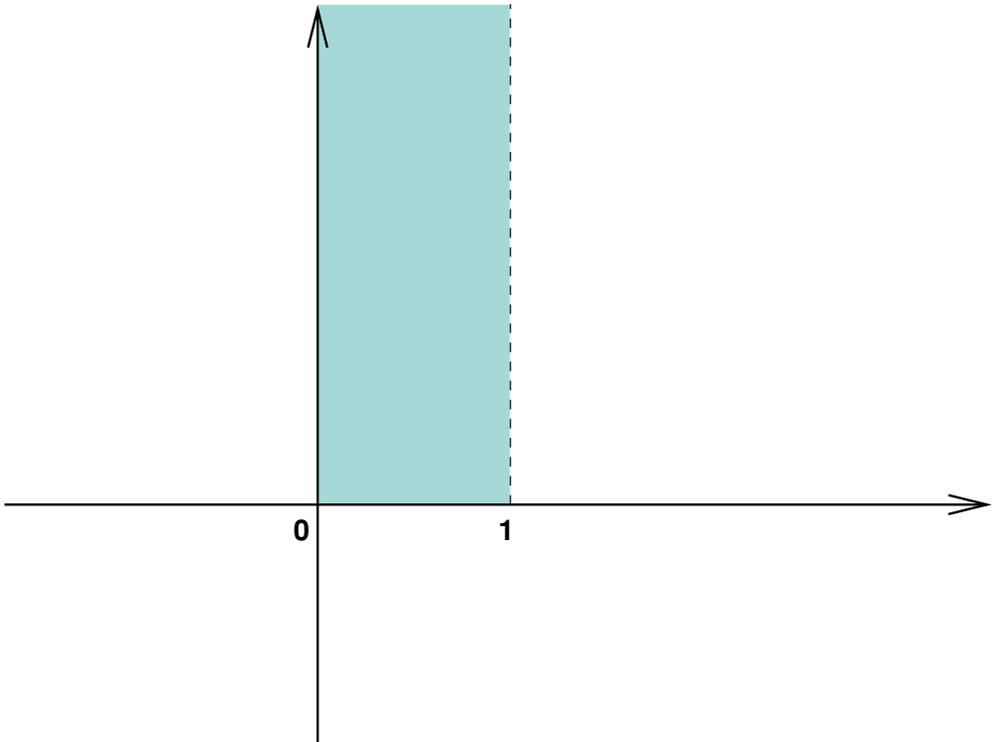


Figura 12.15. Il dominio della densità.

$$F(x, y) = \begin{cases} \int_0^x \int_0^y 2se^{-t} ds dt = x^2(1 - e^{-y}) & \text{per } 0 \leq x \leq 1 \text{ e } y \geq 0 \\ \int_0^1 \int_0^y 2se^{-t} ds dt = 1 - e^{-y} & \text{per } x > 1 \text{ e } y \geq 0 \\ 0 & \text{altrimenti.} \end{cases}$$

- c) Consideriamo ora $Z = X + Y$. Per calcolare $\mathbf{P}(Z)$ e $\sigma^2(Z)$ basta usare:
- (i) *additività* della previsione

$$\mathbf{P}(Z) = \mathbf{P}(X) + \mathbf{P}(Y) = \frac{2}{3} + 1 = \frac{5}{3}.$$

- (ii) la formula per la varianza della somma di due numeri aleatori

$$\begin{aligned} \sigma^2(Z) &= \sigma^2(X + Y) = \sigma^2(X) + \sigma^2(Y) + 2\mathbf{cov}(X, Y) \\ &= \sigma^2(X) + \sigma^2(Y) = \frac{19}{18}. \end{aligned}$$

Per calcolare la funzione di ripartizione di $Z = X + Y$, si utilizza il fatto che

$$\begin{aligned}
 F_Z(z) &= \mathbf{P}(Z \leq z) = \mathbf{P}(X + Y \leq z) \\
 &= \mathbf{P}(Y \leq z - X) \\
 &= \int \int_{\mathcal{D}_z} f(s, t) ds dt
 \end{aligned}$$

dove, per ogni z fissato, \mathcal{D}_z è la parte di piano ottenuta dall'intersezione del dominio di definizione della densità e del semipiano

$$z = \{(x, y) | y \leq z - x\}.$$

Le figure 12.16 e 12.17 mostrano la regione di piano intersecata da z sul dominio della densità al variare di z .

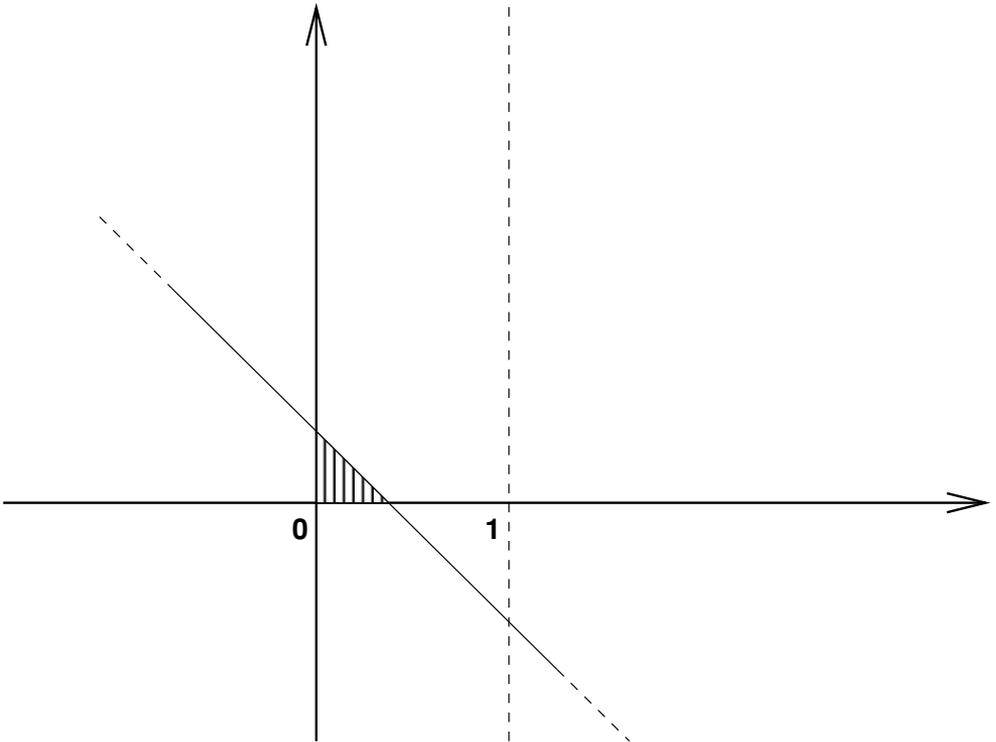


Figura 12.16. Caso $0 < z < 1$.

- Si ottiene che per
- (i) $z < 0$, $F_z(z) = 0$.
 - (ii) $0 < z < 1$,

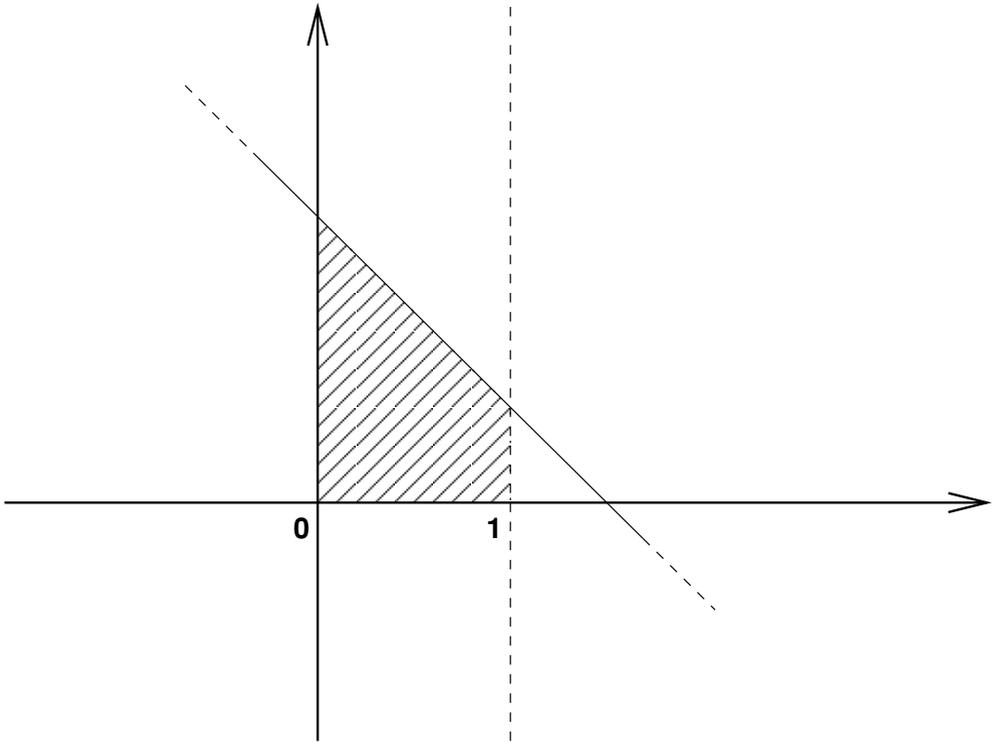


Figura 12.17. Caso $z > 1$.

$$F_Z(z) = \int_0^z 2x \int_0^{z-x} e^{-y} dy dx = \int_0^z 2x (1 - e^{-(z-x)}) dx = z^2 + 2(1 - z) - 2e^{-z}.$$

(iii) $z > 1$

$$F_Z(z) = \int_0^1 \int_0^{z-x} 2xe^{-y} dy dx = \int_0^1 2x (1 - e^{-(z-x)}) dx = 1 - 2e^{-z}.$$

Per calcolare la densità di Z , basta allora derivare la funzione di ripartizione. Si ottiene:

$$f_Z(z) = \begin{cases} 2z - 2 + 2e^{-z} & \text{per } 0 \leq z < 1 \\ 2e^{-z} & \text{per } z > 1 \\ 0 & \text{altrimenti.} \end{cases}$$

□

Esercizio 12.7. I numeri aleatori X ed Y hanno densità gaussiana bi-dimensionale

$$p(x, y) = \frac{1}{2\pi} e^{-\frac{1}{2}(x^2+y^2)}.$$

Siano $U = 2X + 3Y$ e $V = X - Y$. Calcolare

1. La matrice di covarianza di U e V .
2. La densità congiunta di U e V .

Soluzione 12.7. 1. Calcoliamo la matrice di covarianza di U e V :

$$C = \begin{pmatrix} \sigma^2(U) & \mathbf{cov}(U, V) \\ \mathbf{cov}(U, V) & \sigma^2(V) \end{pmatrix}.$$

Per calcolare C si utilizza la formula della varianza di una somma di numeri aleatori:

- $\sigma^2(U)$

$$\begin{aligned} \sigma^2(U) &= \sigma^2(2X + 3Y) \\ &= 4\sigma^2(X) + 9\sigma^2(Y) + 2 \cdot 6 \mathbf{cov}(X, Y) \\ &= 4 \cdot 1 + 9 \cdot 1 + 12 \cdot 0 \\ &= 13. \end{aligned}$$

- $\sigma^2(V)$

$$\begin{aligned} \sigma^2(V) &= \sigma^2(X - Y) \\ &= \sigma^2(X) + \sigma^2(Y) - 2 \mathbf{cov}(X, Y) \\ &= 2. \end{aligned}$$

- $\mathbf{cov}(U, V)$

$$\begin{aligned} \mathbf{cov}(U, V) &= \mathbf{cov}(2X + 3Y, X - Y) \\ &= 2\sigma^2(X) - 2 \mathbf{cov}(X, Y) + 3 \mathbf{cov}(X, Y) - 3\sigma^2(Y) \\ &= 2 - 3 \\ &= -1. \end{aligned}$$

La matrice di covarianza è

$$C = \begin{pmatrix} 13 & -1 \\ -1 & 2 \end{pmatrix}.$$

2. Per calcolare la densità congiunta di (U, V) si calcola la funzione di ripartizione di (U, V) .

$$F(u, v) = \mathbf{P}(U \leq u, V \leq v) = \mathbf{P}(2X + 3Y \leq u, X - Y \leq v).$$

Tale probabilità è data dall'integrale della densità congiunta di (X, Y) calcolato sul dominio $D_{u,v}$ di \mathbb{R}^2 dove

$$D_{u,v} = \{(x, y) \in \mathbb{R}^2 \mid 2x + 3y \leq u, x - y \leq v\}.$$

Si ottiene

$$F(u, v) = \int \int_{D_{u,v}} f(x, y) \, dx \, dy$$

Per risolvere tale integrale, operiamo il cambio di variabili

$$z = 2x + 3y, \quad t = x - y,$$

in modo da trasformare il dominio $D_{u,v}$ in una regione

$$\hat{D}_{u,v} = \{(x, y) \in \mathbb{R}^2 \mid z \leq u, t \leq v\}.$$

di lati paralleli agli assi. Se calcoliamo x, y in funzione di z e t , si ottiene

$$x = \frac{1}{5}(z + 3t), \quad y = \frac{1}{5}(z - 2t).$$

Ne segue che la matrice Jacobiana è pari a

$$J_{\Psi} = \begin{pmatrix} \frac{\partial \Psi_1}{\partial z} & \frac{\partial \Psi_1}{\partial t} \\ \frac{\partial \Psi_2}{\partial z} & \frac{\partial \Psi_2}{\partial t} \end{pmatrix} = \begin{pmatrix} \frac{1}{5} & \frac{3}{5} \\ \frac{1}{5} & -\frac{2}{5} \end{pmatrix},$$

dove $(x, y) = \Psi(z, t) = (\Psi_1(z, t), \Psi_2(z, t)) = \left(\frac{z + 3t}{5}, \frac{z - 2t}{5}\right)$. Si ricava quindi il determinante Jacobiano

$$|\det J_{\Psi}| = \frac{1}{5},$$

Si ottiene:

$$\begin{aligned} F(u, v) &= \int \int_{D_{u,v}} f(x, y) \, dx \, dy \\ &= \int \int_{\hat{D}_{u,v}} f(\Psi(z, t)) |\det J_{\Psi}| \, dz \, dt \\ &= \int_{-\infty}^u \int_{-\infty}^v \frac{1}{2\pi} e^{-\frac{1}{2} \left(\left(\frac{z+3t}{5} \right)^2 + \left(\frac{z-2t}{5} \right)^2 \right)} \frac{1}{5} \, dz \, dt \\ &= \int_{-\infty}^u \int_{-\infty}^v \frac{1}{10\pi} e^{-\frac{1}{2} \cdot \frac{1}{25} (2z^2 + 13t^2 + 2zt)} \, dz \, dt. \end{aligned}$$

La densità congiunta di (U, V) è allora

$$\frac{1}{10\pi} e^{-\frac{1}{50}(2z^2+13t^2+2zt)}.$$

Si osserva che (U, V) hanno ancora distribuzione congiunta gaussiana bidimensionale con matrice di varianza e covarianza pari a C .

Si calcoli per esercizio e verifichi la matrice inversa di A , dove

$$A = \begin{pmatrix} \frac{2}{25} & \frac{1}{25} \\ \frac{1}{25} & \frac{13}{25} \end{pmatrix}.$$

□

Esercizio 12.8. Un vettore aleatorio (X, Y, Z) ha densità congiunta data da

$$f(x, y, z) = k e^{-\frac{1}{2}(2x^2 - 2xy + y^2 + z^2 + 2x - 6y)}.$$

1. Calcolare k .
2. Calcolare le previsioni $\mathbf{P}(X)$, $\mathbf{P}(Y)$ e $\mathbf{P}(Z)$.
3. Calcolare la densità marginale congiunta del vettore aleatorio (X, Z) .
4. Calcolare il coefficiente di correlazione fra X e Z e fra X e Y .
5. Sia $W = X + Z$; calcolare la densità di probabilità di W .

Soluzione 12.8. 1. Se si scrive la densità nella forma standard

$$f(x, y, z) = k e^{-\frac{1}{2} A v \cdot v + b \cdot v}$$

dove A è la matrice simmetrica,

$$A = \begin{pmatrix} 2 & -1 & 0 \\ -1 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

b è il vettore di \mathbb{R}^3

$$b = \begin{pmatrix} -1 \\ 3 \\ 0 \end{pmatrix}$$

e v il vettore delle variabili

$$v = \begin{pmatrix} x \\ y \\ z \end{pmatrix}$$

si può ricavare la costante di normalizzazione k nel modo seguente:

$$k = \sqrt{\frac{\det A}{(2\pi)^3}} e^{-\frac{1}{2} A^{-1} b \cdot b}.$$

Basta dunque calcolare il determinante e la matrice inversa di A . Si ottiene

$$\det A = 1,$$

$$A^{-1} = \begin{pmatrix} 1 & 1 & 0 \\ 1 & 2 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

da cui

$$A^{-1}b = \begin{pmatrix} 2 \\ 5 \\ 0 \end{pmatrix}$$

e

$$k = e^{-\frac{1}{2}A^{-1}b \cdot b} \sqrt{\frac{\det A}{(2\pi)^3}} = e^{-\frac{13}{2}} \sqrt{\frac{1}{(2\pi)^3}}.$$

2. Le previsioni di X, Y, Z sono rispettivamente date da

$$\mathbf{P}(X) = [A^{-1}b]_1 = 2$$

$$\mathbf{P}(Y) = [A^{-1}b]_2 = 5$$

$$\mathbf{P}(Z) = [A^{-1}b]_3 = 0.$$

3. Il vettore aleatorio (X, Z) ha densità gaussiana bidimensionale di matrice D di covarianza data da

$$D = \begin{pmatrix} [A^{-1}]_{11} & [A^{-1}]_{13} \\ [A^{-1}]_{31} & [A^{-1}]_{33} \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

e vettore d delle previsioni

$$d = \begin{pmatrix} 2 \\ 0 \end{pmatrix}$$

Per dimostrare ciò, basta calcolare la densità congiunta $f_{X,Z}(x, z)$ a partire da $f(x, y, z)$

$$\begin{aligned} f_{X,Z}(x, z) &= \int_{\mathbb{R}} f(x, y, z) \, dy \\ &= \int_{\mathbb{R}} k e^{-\frac{1}{2}(2x^2 - 2xy + y^2 + z^2 + 2x - 6y)} \, dy \\ &= k e^{-\frac{1}{2}(2x^2 + z^2 + 2x)} \int_{\mathbb{R}} k e^{-\frac{1}{2}(y^2 - 2xy) + 3y} \, dy \\ &= k e^{-\frac{1}{2}(2x^2 + z^2) - x} \int_{\mathbb{R}} k e^{-\frac{1}{2}y^2 + (3+x)y} \, dy. \end{aligned}$$

Si considera

$$I = \int_{\mathbb{R}} k e^{-\frac{1}{2}y^2 + (3+x)y} \, dy$$

come l'integrale di una distribuzione gaussiana uni-dimensionale i cui coefficienti dipendono da x . Utilizzando la stessa notazione del punto del punto 1 di questo esercizio, si ottiene

$$A = 1$$

$$b = 3 + x$$

da cui

$$\begin{aligned} \int_{\mathbb{R}} k e^{-\frac{1}{2}y^2 + (3+x)y} dy &= \sqrt{\frac{2\pi}{\det A}} e^{\frac{1}{2}A^{-1}b \cdot b} \\ &= \sqrt{2\pi} e^{\frac{1}{2}(3+x)^2}. \end{aligned}$$

Tale risultato si poteva ottenere calcolando l'integrale I completando il quadrato

$$-\frac{1}{2}y^2 + (3+x)y.$$

Ne segue che

$$\begin{aligned} f_{X,Z}(x, z) &= k e^{-\frac{1}{2}(2x^2+z^2)-x} \cdot \sqrt{2\pi} e^{\frac{1}{2}(3+x)^2} \\ &= \frac{e^{-\frac{13}{2}+\frac{9}{2}}}{2\pi} e^{-\frac{1}{2}(x^2+z^2)+2x} \\ &= \frac{e^{-2}}{2\pi} e^{-\frac{1}{2}(x^2+z^2)+2x}. \end{aligned}$$

4. Il coefficiente di correlazione fra X e Z si ottiene dalla formula

$$\rho(X, Z) = \frac{\mathbf{cov}(X, Z)}{\sigma(X)\sigma(Z)}.$$

Dalla matrice di covarianza:

$$\mathbf{cov}(X, Z) = 0,$$

quindi

$$\rho(X, Y) = \frac{1}{\sqrt{2}\sqrt{1}} = \frac{\sqrt{2}}{2}.$$

5. La densità di probabilità di W si calcola utilizzando la formula

$$f_W(w) = \int_{\mathbb{R}} f_{X,Z}(x, w-x) dx.$$

Quindi, con lo stesso metodo usato al punto precedente:

$$\begin{aligned}
f_W(w) &= \int_{\mathbb{R}} \frac{e^{-2}}{2\pi} e^{-\frac{1}{2}(x^2+(w-x)^2)+2x} dx \\
&= \frac{e^{-2}}{2\pi} e^{-\frac{1}{2}w^2} \int_{\mathbb{R}} e^{-\frac{1}{2}(2x^2)+(2+w)x} dx \\
&= \frac{e^{-2}}{2\pi} e^{-\frac{1}{2}w^2} \sqrt{\pi} e^{\frac{1}{2} \cdot \frac{1}{2} (2+w)^2} \\
&= \frac{e^{-2+1}}{2\sqrt{\pi}} e^{-\frac{1}{4}w^2+w} \\
&= \frac{1}{2\sqrt{\pi}} e^{-\frac{1}{4}(w-2)^2}.
\end{aligned}$$

Il numero aleatorio W ha densità normale di previsione

$$\mathbf{P}(W) = \mathbf{P}(X) + \mathbf{P}(Z) = 2$$

e varianza

$$\sigma^2(W) = \sigma^2(X) + \sigma^2(Z) + 2\mathbf{cov}(X, Z) = 2.$$

□

Catene di Markov

Esercizio 13.1. Una catena di Markov $(X_n)_{n \in \mathbb{N}}$ con insieme degli stati $S = \{1, 2, 3, 4\}$ ha la seguente matrice di transizione

$$\begin{pmatrix} \frac{1}{4} & \frac{3}{4} & 0 & 0 \\ 0 & 0 & \frac{2}{3} & \frac{1}{3} \\ \frac{1}{4} & 0 & \frac{3}{4} & 0 \\ 0 & \frac{1}{3} & 0 & \frac{2}{3} \end{pmatrix}$$

e distribuzione iniziale

$$\mu(1) = \mu(2) = \mu(3) = \mu(4) = \frac{1}{4}.$$

- Dire quali sono le classi di equivalenza fra stati ed i loro periodi.
- Calcolare $p_{2,1}^{(2)}$, $p_{1,4}^{(2)}$, $p_{1,1}^{(2)}$.
- Dire se esistono ed in caso positivo calcolare

$$\lim_{n \rightarrow \infty} p_{1,3}^{(n)} \quad \text{e} \quad \lim_{n \rightarrow \infty} \mathbf{P}(X_n = 2).$$

Soluzione 13.1. a) Per individuare le classi di equivalenza fra gli stati ed il loro periodo si può costruire un grafico della probabilità di transizione usando la matrice P . In questo caso si rappresentano gli stati (figura 13.1) e si congiungono con una freccia quelli che hanno probabilità positiva di



Figura 13.1. Gli stati.

passare l'uno dall'altro in un passo. Per esempio, poiché

$$[P]_{1,2} = \frac{3}{4}$$

questo significa che la catena ha possibilità positiva di transitare dallo stato 1 allo stato 2 in un passo. Questo si può rappresentare nel grafico unendo 1 e 2 con una freccia (entrante in 2) (figura 13.2).

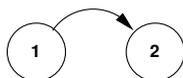


Figura 13.2. La catena ha probabilità positiva di transitare da 1 a 2.

Analogamente, $[P]_{1,1} = \frac{1}{4}$ significa che la catena ha probabilità positiva di rimanere ferma nello stato 1. Questo si rappresenta nel seguente modo (figura 13.3).



Figura 13.3. La catena può con probabilità positiva di rimanere ferma nello stato 1.

Usando questo criterio, si può dunque costruire il grafico 13.4.

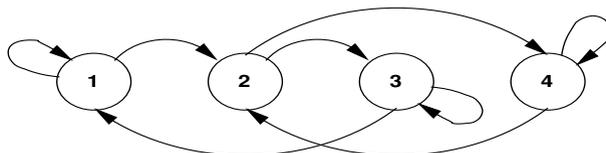


Figura 13.4. Il grafico delle relazioni fra gli stati.

Dal grafico si deduce che tutti gli stati comunicano fra loro, ovvero esistono dei percorsi che, con probabilità positiva, congiungono ogni stato agli altri. Esiste quindi un'unica classe di equivalenza [1].

Inoltre, dal grafico si deduce subito che il periodo della catena è 1 in quanto esiste un percorso lungo 1 contenuto in

$$A_{1,1}^+ = \{n \mid p_{1,1}^{(n)}\}.$$

- b) Per calcolare $p_{2,1}^{(2)}$, ovvero la probabilità di transitare in 2 passi dallo stato 2 allo stato 1, si utilizza il fatto che

$$p_{2,1}^{(2)} = \sum_{i \in S} p_{2,i}^{(1)} p_{i,1}^{(1)}$$

ovvero si può “spezzare” tale probabilità da 2 a 1 nella somma di tutti i possibili percorsi da 2 a 1.

Dal grafico relativo alla matrice P si ottiene

$$p_{2,1}^{(2)} = p_{2,3}^{(1)} p_{3,1}^{(1)} = \frac{2}{3} \cdot \frac{1}{4} = \frac{1}{6}$$

Si noti che si può ricavare $p_{2,1}^{(2)}$ dal prodotto riga per colonna

$$p_{2,1}^{(2)} = P_2 \cdot P^1$$

dove P_2 denota la seconda riga e P^1 la prima colonna della matrice P . Analogamente si calcolano $p_{1,4}^{(2)}$ e $p_{1,1}^{(2)}$.

- c) Poiché la catena è irriducibile e aperiodica, si può applicare il teorema ergodico che garantisce l'esistenza del limite

$$\lim_{n \rightarrow \infty} p_{1,3}^{(n)} = \pi_3,$$

dove π_3 si ottiene dalla soluzione del sistema lineare

$$\begin{aligned} \pi &= {}^t \pi P \\ \left\{ \begin{array}{l} \pi_1 = \sum \pi_i p_{i,1} \\ \pi_2 = \sum \pi_i p_{i,2} \\ \pi_3 = \sum \pi_i p_{i,3} \\ \pi_1 + \pi_2 + \pi_3 + \pi_4 = 1. \end{array} \right. \end{aligned}$$

In questo caso

$$\left\{ \begin{array}{l} \pi_1 = \frac{1}{4} \pi_1 + \frac{1}{4} \pi_3 \\ \pi_2 = \frac{3}{4} \pi_1 + \frac{1}{3} \pi_4 \\ \pi_3 = \frac{2}{3} \pi_2 + \frac{3}{4} \pi_3 \\ \sum_{i=1}^4 \pi_i = 1. \end{array} \right.$$

Risolvendo tale sistema con i metodi standard per la risoluzione di equazioni lineari, si ottiene

$$\pi_3 = \frac{12}{25}.$$

Quindi

$$\lim_{n \rightarrow \infty} p_{1,3}^{(n)} = \frac{12}{25}.$$

Inoltre, per calcolare $\lim_{n \rightarrow \infty} \mathbf{P}(X_n = 2)$, si osservi che

$$\mathbf{P}(X_n = 2) = \sum_{i=1}^4 \mathbf{P}(X_n = 2 | X_0 = i) \mathbf{P}(X_0 = i) = \frac{1}{4} \sum_{i=1}^4 p_{i,2}^{(n)}$$

Poiché per ogni i

$$\lim_{n \rightarrow \infty} p_{i,2}^{(n)} = \pi_2,$$

si ottiene

$$\begin{aligned} \lim_{n \rightarrow \infty} \mathbf{P}(X_n = 2) &= \lim_{n \rightarrow \infty} \frac{1}{4} \sum_{i=1}^4 p_{i,2}^{(n)} \\ &= \frac{1}{4} \cdot 4\pi_2 \\ &= \pi_2 \end{aligned}$$

dove $\pi_2 = \frac{9}{50}$.

□

Esercizio 13.2. Una catena di Markov X_n , $n = 0, 1, 2, \dots$ con insieme degli stati

$$S = \{1, 2, 3, 4, 5, 6\}$$

ha la seguente matrice di transizione

$$\begin{pmatrix} 0 & \frac{1}{3} & 0 & \frac{1}{3} & 0 & \frac{1}{3} \\ \frac{1}{2} & 0 & \frac{1}{2} & 0 & 0 & 0 \\ 0 & \frac{1}{3} & 0 & \frac{2}{3} & 0 & 0 \\ 0 & 0 & \frac{2}{3} & 0 & \frac{1}{3} & 0 \\ 0 & 0 & 0 & \frac{1}{2} & 0 & \frac{1}{2} \\ \frac{1}{3} & 0 & \frac{1}{3} & 0 & \frac{1}{3} & 0 \end{pmatrix}$$

e distribuzione iniziale

$$\mu(1) = \frac{1}{3}, \mu(2) = \frac{2}{3}, \mu(3) = \mu(4) = \mu(5) = \mu(6) = 0.$$

1. Dire quali sono le classi di equivalenza fra stati e i loro periodi.
2. Dire se esistono e, in caso, calcolare

$$\lim_{n \rightarrow \infty} p_{1,5}^{(2n)}, \lim_{n \rightarrow \infty} p_{3,5}(n), \lim_{n \rightarrow \infty} p_{2,5}^{(2n)} \text{ e } \lim_{n \rightarrow \infty} \mathbf{P}(X_n = 5).$$

3. Calcolare $\mathbf{P}(X_2 < 3)$.

Soluzione 13.2. 1. Come nell'esercizio precedente, per calcolare le classi di equivalenza si disegna il grafico degli stati come in figura 13.5, congiungendo gli stati tra loro con una freccia a seconda che esista una probabilità positiva di passare dallo stato da cui la freccia parte a quello in cui la freccia entra.



Figura 13.5. Grafico degli stati.

Usando la matrice P di transizione si ottiene il grafico mostrato in figura 13.6, da cui si ottiene che esiste un'unica classe di equivalenza. Inoltre, il numero di passi che si deve compiere da uno stato per tornare nello stesso stato è sempre pari. Ne segue che il periodo d della classe di equivalenza è 2. Infatti vale

$$d = \text{MCD } A_{s,s}^+$$

dove $A_{s,s}^+ = \{n \mid p_{s,s}^{(n)} > 0\}$.

2. Per studiare i limiti si considerano le classi di equivalenza della matrice P^2 ; esse sono due, ciascuna di periodo 1. Per ottenere la loro composizione non è necessario calcolare esplicitamente tutta la matrice P^2 ; per esempio, la classe di 1 sarà formata da tutti gli stati che comunicano con 1 in un numero pari di passi. Si ottiene

$$\begin{aligned} [1] &= \{1, 3, 5\} \\ [2] &= \{2, 4, 6\}. \end{aligned}$$

Poiché 2 e 5 non comunicano in un numero pari di passi si ottiene subito

$$p_{2,5}^{(2n)} = 0, \forall n$$

quindi

$$\lim_{n \rightarrow \infty} p_{2,5}^{(2n)} = 0.$$

Lo stato 5 appartiene alla classe [1] calcolata rispetto a P^2 , quindi si può applicare il teorema ergodico a tale classe in quanto essa ha periodo 1 rispetto a P^2 . La sottomatrice di P^2 relativa alla sottoclasse [1] è:

$$\begin{pmatrix} \frac{5}{18} & \frac{9}{18} & \frac{2}{9} \\ \frac{1}{6} & \frac{11}{18} & \frac{2}{9} \\ \frac{1}{6} & \frac{1}{2} & \frac{1}{3} \end{pmatrix}.$$

Applicando il teorema ergodico si ha che

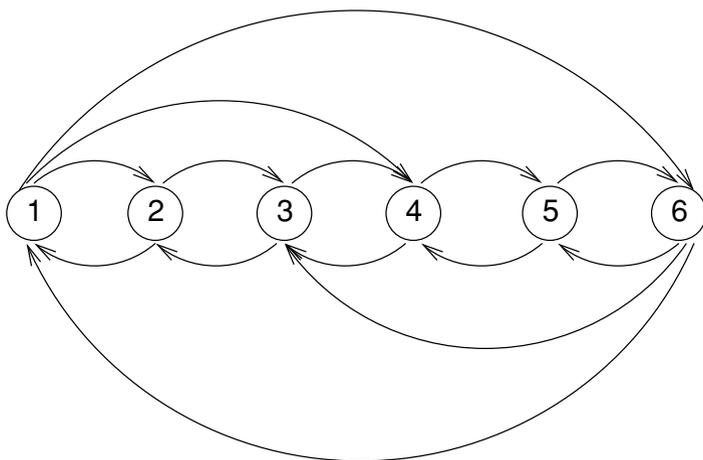


Figura 13.6. Grafico delle probabilità di transizione.

$$\lim_{n \rightarrow \infty} p_{1,5}^{(2n)} = \pi_5$$

dove π_5 è la soluzione del sistema

$$\begin{cases} \pi_1 = \frac{5}{18}\pi_1 + \frac{1}{6}\pi_3 + \frac{1}{6}\pi_5 \\ \pi_3 = \frac{9}{18}\pi_1 + \frac{11}{18}\pi_3 + \frac{1}{2}\pi_5 \\ \pi_1 + \pi_3 + \pi_5 = 1. \end{cases}$$

La soluzione del sistema è

$$\pi_1 = \frac{3}{16} \quad \pi_3 = \frac{9}{16} \quad \pi_5 = \frac{1}{4}.$$

Allora

$$\lim_{n \rightarrow \infty} P_{1,5}^{(2n)} = \frac{1}{4}.$$

Per quanto riguarda il comportamento di $p_{3,5}^{(n)}$ per n che tende all'infinito, si ha che:

a) Sui passi pari, ovvero per $n = 2k$

$$p_{3,5}^{(2k)} \xrightarrow[k \rightarrow \infty]{} \pi_5.$$

b) Sui passi dispari, ovvero per $n = 2k + 1$

$$p_{3,5}^{(2k+1)} = 0$$

perché la probabilità di andare dallo stato 3 allo stato 5 in un numero dispari di passi è zero. Poiché il limite su due sottosuccessioni è diverso, si conclude che

$$\lim_{n \rightarrow \infty} p_{3,5}^{(n)}$$

non esiste. Per calcolare

$$\lim_{n \rightarrow \infty} \mathbf{P}(X_n = 5)$$

si usa la formula delle probabilità totali:

$$\begin{aligned} \mathbf{P}(X_n = 5) &= \sum_{i=1}^6 \mathbf{P}(X_n = 5 | X_0 = i) \mathbf{P}(X_0 = i) \\ &= \sum_{i=1}^6 p_{i,5}^{(n)} \mu_i \\ &= \frac{1}{3} p_{1,5}^{(n)} + \frac{2}{3} p_{2,5}^{(n)}. \end{aligned}$$

Si devono distinguere due casi:

a) cosa succede sui passi pari, ovvero per $n = 2k$

$$\frac{1}{3} p_{1,5}^{(2k)} + \frac{2}{3} p_{2,5}^{(2k)} = \frac{1}{3} p_{1,5}^{(2k)} \xrightarrow[k \rightarrow \infty]{} \frac{\pi_5}{3};$$

b) cosa succede sui passi dispari, ovvero per $n = 2k + 1$

$$\frac{1}{3} p_{1,5}^{(2k+1)} + \frac{2}{3} p_{2,5}^{(2k+1)} = \frac{2}{3} p_{2,5}^{(2k+1)} = \frac{2}{3} \sum_{i=1}^6 p_{2,i}^{(1)} p_{i,5}^{(2k)}$$

che tende a $\frac{2}{3} \pi_5 \sum_{i=1}^6 p_{2,i}^{(1)} = \frac{2}{3} \pi_5$ per $k \rightarrow \infty$.

Poiché si ottengono due limiti diversi, si conclude che il limite

$$\lim_{n \rightarrow \infty} \mathbf{P}(X_n = 5)$$

non esiste.

3. Per calcolare $\mathbf{P}(X_2 < 3)$ basta osservare che

$$\mathbf{P}(X_2 < 3) = \sum_{i=1}^2 \mathbf{P}(X_2 = i),$$

in quanto l'evento $(X_2 < 3) = (X_2 = 1) + (X_2 = 2)$. Basta dunque calcolare

$$\begin{aligned} \mathbf{P}(X_2 = 1) &= \sum_{i=1}^6 \mathbf{P}(X_2 = 1 | X_0 = i) \mathbf{P}(X_0 = i) \\ &= \sum_{i=1}^6 p_{i,1}^{(2)} \mu_i \\ &= \frac{1}{3} p_{1,1}^{(2)} + \frac{2}{3} p_{2,1}^{(2)} = \frac{5}{54} \end{aligned}$$

e

$$\begin{aligned} \mathbf{P}(X_2 = 2) &= \sum_{i=1}^6 \mathbf{P}(X_2 = 2 | X_0 = i) \mathbf{P}(X_0 = i) \\ &= \sum_{i=1}^6 p_{i,2}^{(2)} \mu_i \\ &= \frac{1}{3} p_{1,2}^{(2)} + \frac{2}{3} p_{2,2}^{(2)} = \frac{2}{9}. \end{aligned}$$

Concludendo, $\mathbf{P}(X_2 < 3) = \frac{17}{54}$.

□

Esercizio 13.3. Una catena di Markov X_n , $n = 0, 1, 2, \dots$ con insieme degli stati

$$S = \{1, 2, 3, 4\}$$

ha la seguente matrice di transizione

$$P = \begin{pmatrix} 0 & \frac{1}{2} & \frac{1}{2} & 0 \\ \frac{2}{3} & 0 & 0 & \frac{1}{3} \\ \frac{1}{6} & 0 & 0 & \frac{5}{6} \\ 0 & \frac{3}{4} & \frac{1}{4} & 0 \end{pmatrix}$$

e distribuzione iniziale

$$\mu(1) = \frac{1}{3}, \mu(2) = \frac{1}{3}, \mu(3) = \frac{1}{3}, \mu(4) = 0.$$

1. Dire quali sono le classi di equivalenza fra stati e i loro periodi.
2. Calcolare $\mathbf{P}(X_5 = 2 | X_2 = 3)$, $p_{1,4}(2)$ e $\mathbf{P}(X_2)$.
3. Dire se esistono e in caso positivo calcolare

$$\lim_{n \rightarrow \infty} p_{1,3}^{(2n)}, \lim_{n \rightarrow \infty} p_{1,4}^{(2n)}, \lim_{n \rightarrow \infty} p_{2,3}^{(n)} \text{ e } \lim_{n \rightarrow \infty} \mathbf{P}(X_n = 2).$$

Soluzione 13.3. 1. Per trovare le classi di equivalenza si disegna il grafico degli stati come in figura 13.7.

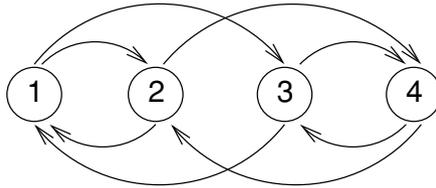


Figura 13.7. Grafico degli stati.

Poiché a partire dallo stato 1 si raggiungono tutti gli stati e viceversa, esiste un'unica classe di equivalenza. Inoltre, partendo dallo stato 1, si ritorna sempre in esso con un numero pari di passi. Si conclude che il periodo della classe è 2.

2. Per calcolare la probabilità subordinata $\mathbf{P}(X_5 = 2 | X_2 = 3)$ si applica il fatto che la catena è omogenea. Vale infatti che

$$\mathbf{P}(X_5 = 2 | X_2 = 3) = p_{3,2}^{(3)} = [P^3]_{3,2} = 0.$$

Per calcolare tale probabilità basta calcolare l'elemento alla riga 3 e colonna 2 della matrice P^3 , che si ottiene come prodotto della terza riga di P^2 con la seconda colonna di P .

Senza fare dei conti, si può comunque subito notare che la probabilità di andare dallo stato 3 allo stato 2 in un numero dispari di passi è 0!

Analogamente

$$p_{1,4}^{(2)} = [P^2]_{1,4} = \frac{7}{12}.$$

Calcoliamo la previsione della catena al tempo $t = 2$ usando la formula della previsione per un numero aleatorio con distribuzione discreta e successivamente quella delle probabilità totali.

$$\begin{aligned} \mathbf{P}(X_2) &= \sum_{i=1}^4 i \mathbf{P}(X_2 = i) \\ &= \sum_{i=1}^4 i \sum_{j=1}^4 \mathbf{P}(X_2 = i | X_0 = j) \mathbf{P}(X_0 = j) \\ &= \sum_{i=1}^4 i \frac{1}{3} (\mathbf{P}(X_2 = i | X_0 = 1) + \mathbf{P}(X_2 = i | X_0 = 2) + \mathbf{P}(X_2 = i | X_0 = 3)) \\ &= \sum_{i=1}^4 i \frac{1}{3} ([P^2]_{1,i} + [P^2]_{2,i} + [P^2]_{3,i}). \end{aligned}$$

La matrice di P^2 è data da:

$$P^2 = \begin{pmatrix} \frac{5}{12} & 0 & 0 & \frac{7}{12} \\ 0 & \frac{7}{12} & \frac{5}{12} & 0 \\ 0 & \frac{17}{24} & \frac{7}{24} & 0 \\ \frac{13}{24} & 0 & 0 & \frac{11}{24} \end{pmatrix}.$$

La previsione di X_2 è quindi:

$$\mathbf{P}(X_2) = \frac{1}{3} \left(\frac{5}{12} + 2 \left(\frac{7}{12} + \frac{17}{24} \right) + 3 \left(\frac{5}{12} + \frac{7}{24} \right) \right) = \frac{41}{24}.$$

3. Calcoliamo i limiti. La catena di Markov osservata solo sui passi pari può essere considerata come una catena di Markov relativa alla matrice di transizione P^2 . Si osserva subito che lo stato 3 non può essere raggiunto a partire dallo stato 1 con un numero pari di passi. Infatti, le classi di equivalenza relative a P^2 sono

$$[1] = \{1, 4\},$$

$$[2] = \{2, 3\}.$$

Ne segue

$$\lim_{n \rightarrow \infty} p_{1,3}^{(2n)} = 0.$$

Lo stato 4 appartiene alla classe di equivalenza 1 relativamente a P^2 ed essa ha inoltre periodo 1. Si può applicare il teorema ergodico a tale sottocatenata irriducibile aperiodica per calcolare

$$\lim_{n \rightarrow \infty} p_{1,4}^{(2n)}.$$

Posto $\pi_4 = \lim_{n \rightarrow \infty} p_{1,4}^{(2n)}$ e $\pi_1 = \lim_{n \rightarrow \infty} p_{1,1}^{(2n)}$, dal teorema ergodico si ottiene

$$\begin{cases} \pi_1 + \pi_4 = 1 \\ \frac{5}{12} \pi_1 + \frac{13}{24} \pi_4 = \pi_1. \end{cases}$$

La soluzione del sistema è

$$\pi_1 = \frac{13}{27} \quad \pi_4 = \frac{14}{27}.$$

Ne segue che

$$\lim_{n \rightarrow \infty} p_{1,4}^{(2n)} = \frac{14}{27}.$$

Per calcolare $\lim_{n \rightarrow \infty} p_{2,3}^{(n)}$ osserviamo il comportamento della catena sui passi pari e su quelli dispari.

a) Per prima cosa si osserva che $2 \in [3]$ relativamente a P^2 , quindi dal teorema ergodico si ha che sui passi pari (ovvero se $n = 2k$)

$$p_{2,3}^{(2k)} \xrightarrow[k \rightarrow \infty]{} \pi_3$$

dove π_3 è la soluzione del sistema

$$\begin{cases} \pi_2 + \pi_3 = 1 \\ \frac{5}{12} \pi_2 + \frac{7}{24} \pi_3 = \pi_3. \end{cases}$$

b) Non esiste un percorso con numero di passi dispari dallo stato 2 allo stato 3, quindi

$$p_{2,3}(2k+1) = 0 \quad \forall k.$$

Si ottiene lo stesso risultato calcolando

$$\begin{aligned} p_{2,3}^{(2k+1)} &= \sum_j p_{2,j}^{(2k)} p_{j,3}(1) = \\ p_{2,1}^{(2k)} p_{1,3}(1) + p_{2,4}^{(2k)} p_{4,3}(1) &= 0. \end{aligned}$$

Riassumendo

$$\begin{aligned} p_{2,3}^{(2k)} &\xrightarrow[k \rightarrow \infty]{} \pi_3 > 0 \\ p_{2,3}^{(2k+1)} &\xrightarrow[k \rightarrow \infty]{} 0. \end{aligned}$$

Quindi, $\lim_{n \rightarrow \infty} p_{2,3}^{(n)}$ non esiste.

Infine, per calcolare $\lim_{n \rightarrow \infty} \mathbf{P}(X_n = 2)$ si procede come nel caso precedente.

Per prima cosa, si usa la formula delle probabilità totali per calcolare $\mathbf{P}(X_n = 2)$:

$$\begin{aligned} \mathbf{P}(X_n = 2) &= \sum_{i=1}^4 \mathbf{P}(X_n = 2 | X_0 = i) \mathbf{P}(X_0 = i) = \\ &= \frac{1}{3} (\mathbf{P}(X_n = 2 | X_0 = 1) + \mathbf{P}(X_n = 2 | X_0 = 2) + \mathbf{P}(X_n = 2 | X_0 = 3)) = \\ &= \frac{1}{3} (p_{1,2}^{(n)} + p_{2,2}^{(n)} + p_{3,2}^{(n)}). \end{aligned}$$

Utilizzando i risultati discussi nel caso precedente, si ottiene

a) Se $n = 2k$

$$\frac{1}{3} (p_{1,2}^{(2k)} + p_{2,2}^{(2k)} + p_{3,2}^{(2k)}) = \frac{1}{3} (p_{2,2}^{(2k)} + p_{3,2}^{(2k)})$$

che tende a $\frac{2}{3}\pi_2$ per $k \rightarrow \infty$.

b) Se $n = 2k + 1$

$$\begin{aligned} \frac{1}{3} (p_{1,2}^{(2k+1)} + p_{2,2}^{(2k+1)} + p_{3,2}^{(2k+1)}) &= \\ &= \frac{1}{3} p_{1,2}^{(2k+1)} = \\ &= \frac{1}{3} \sum_{i=1}^4 p_{1,i}^{(1)} p_{i,2}^{(2k)} = \\ &= \frac{1}{3} (p_{1,2}^{(1)} p_{2,2}^{(2k)} + p_{1,3}^{(1)} p_{3,2}^{(2k)}) \end{aligned}$$

che tende a $\frac{1}{3}\pi_2 (p_{1,2}^{(1)} + p_{1,3}^{(1)}) = \frac{1}{3}\pi_2$ per $k \rightarrow \infty$.

Riassumendo, posto $p_n = \mathbf{P}(X_n = 2)$, si ha che

$$\begin{aligned} p_{2n} &\xrightarrow[n \rightarrow \infty]{} \frac{2}{3}\pi_2, \\ p_{2n+1} &\xrightarrow[n \rightarrow \infty]{} \frac{1}{3}\pi_2, \end{aligned}$$

ovvero $\lim_{n \rightarrow \infty} \mathbf{P}(X_n = 2)$ non esiste. □

Esercizio 13.4. Una catena di Markov $(X_n)_{n \in \mathbb{N}}$ con insieme degli stati $S = \{1, 2, 3, 4, 5\}$ ha la seguente matrice di transizione

$$P = \begin{pmatrix} 0 & \frac{1}{2} & \frac{1}{2} & 0 & 0 \\ \frac{1}{2} & 0 & \frac{1}{2} & 0 & 0 \\ 0 & \frac{2}{3} & 0 & \frac{1}{3} & 0 \\ 0 & 0 & \frac{2}{3} & 0 & \frac{1}{3} \\ \frac{2}{3} & 0 & 0 & \frac{1}{3} & 0 \end{pmatrix}$$

e distribuzione iniziale

$$\mu(1) = 0, \quad \mu(2) = \frac{2}{3}, \quad \mu(3) = \frac{1}{3}, \quad \mu(4) = \mu(5) = 0.$$

- a) Dire quali sono le classi di equivalenza fra stati ed i loro periodi;
 b) dire se esistono e in caso positivo calcolare

$$\lim_{n \rightarrow \infty} p_{1,5}^{(n)}, \quad \lim_{n \rightarrow \infty} p_{3,5}^{(n)}, \quad \lim_{n \rightarrow \infty} (p_{2,3}^{(n)} + p_{3,5}^{(n)}), \quad \lim_{n \rightarrow \infty} \mathbf{P}(X_n = 5)$$

- c) Calcolare $\mathbf{P}(X_1 \leq 2)$ e $\mathbf{P}(X_2 = 5)$.

Soluzione 13.4. a) Per individuare le classi di equivalenza fra gli stati ed il loro periodo, si costruisce il grafico degli stati (figura 13.8).

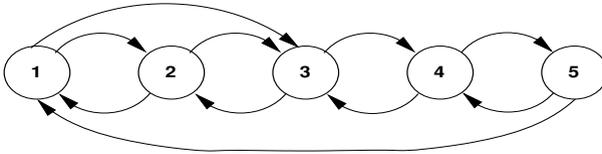


Figura 13.8. Grafico degli stati.

Si osserva per prima cosa che tutti gli stati comunicano fra loro. Si consideri l'insieme

$$A_1^+ = \{n \mid p_{11}^{(n)} > 0\}$$

ovvero l'insieme delle lunghezze dei cammini che iniziano e finiscono in 1. Si osserva che esiste un cammino di lunghezza 2 (per esempio, da 1 a 2 e da 2 a 1) e di lunghezza 3 (da 1 a 3, da 3 a 2, da 2 a 1). Si ha

$$2, 3 \in A_1^+.$$

Il periodo d della classe di equivalenza [1] è dato dal

$$d = MCD(A_1^+),$$

quindi d non può che essere pari ad 1 perché deve dividere sia 2 che 3. Concludendo, esiste un'unica classe di equivalenza di periodo 1.

- b) Dal teorema ergodico segue che i limiti esistono e sono finiti perché la catena ha un'unica classe di periodo 1. Come prima cosa si noti che

$$\lim_{n \rightarrow \infty} p_{1,5}^{(n)} = \lim_{n \rightarrow \infty} p_{3,5}^{(n)} = \pi_5$$

in quanto lo stato di partenza (1 o 3) non conta. Inoltre

$$\begin{aligned} \lim_{n \rightarrow \infty} (p_{2,3}^{(n)} + p_{3,5}^{(n)}) &= \lim_{n \rightarrow \infty} p_{2,3}^{(n)} + \lim_{n \rightarrow \infty} p_{3,5}^{(n)} \\ &= \pi_3 + \pi_5 \end{aligned}$$

ed infine

$$\begin{aligned} \lim_{n \rightarrow \infty} \mathbf{P}(X_n = 5) &= \lim_{n \rightarrow \infty} \sum_{i=1}^5 \mathbf{P}(X_n = 5 | X_0 = i) \mathbf{P}(X_0 = i) \\ &= \lim_{n \rightarrow \infty} \sum_{i=1}^5 \mu(i) p_{i,5}^{(n)} \\ &= \pi_5 \sum_{i=1}^5 \mu(i) = \pi_5 \cdot 1 = \pi_5 \end{aligned}$$

in quanto $\lim_{n \rightarrow \infty} p_{i,5}^{(n)} = \pi_5$, $\forall i = 1, \dots, 5$ e $\sum_{i=1}^5 \mu(i) = 1$. Per ottenere i π_i , basta risolvere il sistema

$$\begin{cases} \pi = \pi^t P \\ \sum_{i=1}^5 \pi_i = 1 \end{cases}$$

ovvero

$$\begin{cases} \pi_1 = \frac{1}{2}\pi_2 + \frac{2}{3}\pi_5 \\ \pi_2 = \frac{1}{2}\pi_1 + \frac{2}{3}\pi_3 \\ \pi_4 = \frac{1}{3}\pi_3 + \frac{1}{3}\pi_5 \\ \pi_5 = \frac{1}{3}\pi_4 \\ \sum_{i=1}^5 \pi_i = 1. \end{cases}$$

Nel sistema è stata già tolta l'equazione sovrabbondante. Risolvendo si ottiene

$$\pi_3 = \frac{7}{540} \quad \text{e} \quad \pi_5 = \frac{14}{135}.$$

Quindi, riassumendo

$$\lim_{n \rightarrow \infty} p_{1,5}^{(n)} = \lim_{n \rightarrow \infty} p_{3,5}^{(n)} = \lim_{n \rightarrow \infty} \mathbf{P}(X_n = 5) = \pi_5 = \frac{14}{135},$$

$$\lim_{n \rightarrow \infty} \left(p_{2,3}^{(n)} + p_{3,5}^{(n)} \right) = \pi_3 + \pi_5 = \frac{7}{540} + \frac{14}{135} = \frac{7}{60}.$$

c) Per calcolare tali probabilità si osserva che

$$\begin{aligned} \mathbf{P}(X_1 \leq 2) &= \mathbf{P}(X_1 = 1) + \mathbf{P}(X_1 = 2) \\ &= \sum_{i=1}^5 \mathbf{P}(X_1 = 1 | X_0 = i) \mu(i) + \sum_{i=1}^5 \mathbf{P}(X_1 = 2 | X_0 = i) \mu(i) \\ &= \frac{2}{3} p_{2,1} + \frac{1}{3} p_{3,1} + \frac{2}{3} p_{2,2} + \frac{1}{3} p_{3,2} \\ &= \frac{1}{3} + \frac{2}{9} = \frac{5}{9}. \end{aligned}$$

La seconda probabilità si calcola di nuovo usando la formula delle probabilità totali:

$$\begin{aligned} \mathbf{P}(X_2 = 5) &= \sum_{i=1}^5 \mathbf{P}(X_2 = 5 | X_0 = i) \mu(i) \\ &= \frac{2}{3} p_{2,5}^{(2)} + \frac{1}{3} p_{3,5}^{(2)} \\ &= \frac{2}{3} [P^2]_{2,5} + \frac{1}{3} [P^2]_{3,5} \\ &= \frac{1}{3} \cdot \frac{1}{9} = \frac{1}{27}. \end{aligned}$$

□

Statistica

Esercizio 14.1. Gli eventi E_1, E_2, \dots sono stocasticamente indipendenti subordinatamente alla conoscenza del parametro aleatorio Θ con $\mathbf{P}(E_i|\Theta = \theta) = \theta$. La densità a priori di Θ è data da

$$\pi_0 = \begin{cases} 3\theta^2 & 0 \leq \theta \leq 1 \\ 0 & \text{altrimenti.} \end{cases}$$

Si osservano i valori dei primi quattro eventi:

$$E_1 = 0, E_2 = 1, E_3 = 1, E_4 = 1.$$

1. Calcolare la densità a posteriori di Θ .
2. Calcolare la probabilità a priori che Θ appartenga all'intervallo $\left[\frac{1}{2}, 1\right]$.
3. Calcolare la probabilità a posteriori che Θ appartenga all'intervallo $\left[\frac{1}{2}, 1\right]$.
4. Calcolare il valore massimo della densità a posteriori di Θ .
5. Calcolare la previsione a posteriori di $E = E_5 \wedge E_6$.

Soluzione 14.1. 1. La densità a posteriori si calcola usando la formula

$$\pi_4(\Theta|E_1 = 0, E_2 = 1, E_3 = 1, E_4 = 1) = k\mathbf{P}(E_1 = 0, E_2 = 1, E_3 = 1, E_4 = 1|\Theta = \theta)\pi_0(\theta).$$

Poiché gli E_i sono stocasticamente indipendenti subordinatamente alla conoscenza di Θ , la probabilità

$$\mathbf{P}(E_1 = 0, E_2 = 1, E_3 = 1, E_4 = 1|\Theta = \theta)$$

si fattorizza in

$$\begin{aligned} \mathbf{P}(E_1 = 0, E_2 = 1, E_3 = 1, E_4 = 1 | \Theta = \theta) = \\ \mathbf{P}(E_1 = 0 | \Theta = \theta) \cdot \mathbf{P}(E_2 = 1 | \Theta = \theta) \mathbf{P}(E_3 = 1 | \Theta = \theta) \cdot \mathbf{P}(E_4 = 1 | \Theta = \theta) = \\ (1 - \theta) \cdot \theta \cdot \theta \cdot \theta. \end{aligned}$$

Quindi, la densità a posteriori è data da

$$\pi_4(\Theta | E_1 = 0, E_2 = 1, E_3 = 1, E_4 = 1) = k\theta^5(1 - \theta),$$

dove k è un'opportuna costante di proporzionalità. La densità a posteriori π_4 è una distribuzione $\beta(6, 2)$, quindi

$$k = \frac{\Gamma(6 + 2)}{\Gamma(6)\Gamma(2)} = \frac{7!}{5!} = 42.$$

2. La probabilità a priori che Θ appartenga all'intervallo $\left[\frac{1}{2}, 1\right]$ è:

$$\begin{aligned} \mathbf{P}\left(\frac{1}{2} \leq \Theta \leq 1\right) &= \int_{\frac{1}{2}}^1 \pi_0(\theta) d\theta = \\ &= \int_{\frac{1}{2}}^1 3\theta^2 d\theta = \left[\theta^3\right]_{\frac{1}{2}}^1 = \frac{7}{8}. \end{aligned}$$

3. La probabilità a posteriori che Θ appartenga all'intervallo $\left[\frac{1}{2}, 1\right]$ è:

$$\begin{aligned} \mathbf{P}\left(\frac{1}{2} \leq \Theta \leq 1 | E_1 = 0, E_2 = E_3 = E_4 = 1\right) &= \\ &= \int_{\frac{1}{2}}^1 \pi_4(\theta | E_1 = 0, E_2 = E_3 = E_4 = 1) d\theta = \\ &= 42 \int_{\frac{1}{2}}^1 (\theta^5 - \theta^6) d\theta = 42 \left[\frac{\theta^6}{6} - \frac{\theta^7}{7} \right]_{\frac{1}{2}}^1 = \frac{15}{16}. \end{aligned}$$

4. Il valore massimo della densità a posteriori di Θ si trova calcolando la derivata di $\pi_4(\theta | E_1 = 0, E_2 = 1, E_3 = 1, E_4 = 1)$:

$$\frac{d}{d\theta} \pi_4 = 42(5\theta^4 - 6\theta^5) = 42\theta^4(5 - 6\theta).$$

La derivata si annulla nel punto $\bar{\theta} = \frac{5}{6}$. Poiché

$$\frac{d^2}{d^2\theta} \pi_4 \Big|_{\theta=\frac{5}{6}} = 42\theta^3(20 - 30\theta) \Big|_{\theta=\frac{5}{6}} < 0$$

si conclude che $\bar{\theta} = \frac{5}{6}$ è un punto di massimo.

5. La previsione a posteriori dell'evento $E = E_5 \wedge E_6 = \min(E_5, E_6) = E_5 E_6$ coincide con la sua probabilità a posteriori.

$$\begin{aligned}
 & \mathbf{P}(E_5 E_6 | E_1 = 0, E_2 = E_3 = E_4 = 1) = \\
 & \int_0^1 \mathbf{P}(E_5 E_6 | \theta) \mathbf{P}(\theta | E_1 = 0, E_2 = E_3 = E_4 = 1) d\theta = \\
 & \int_0^1 \mathbf{P}(E_5 E_6 | \theta) \pi_4(\theta | E_1 = 0, E_2 = E_3 = E_4 = 1) d\theta = \\
 & \int_0^1 \mathbf{P}(E_5 | \theta) \mathbf{P}(E_6 | \theta) \cdot 42 \theta^5 (1 - \theta) d\theta = \\
 & \int_0^1 \theta^2 42 \theta^5 (1 - \theta) d\theta = \\
 & 42 \int_0^1 \theta^7 (1 - \theta) d\theta = \\
 & \frac{\Gamma(8)}{\Gamma(6) \Gamma(2)} \cdot \frac{\Gamma(8) \Gamma(2)}{\Gamma(10)} = \\
 & \frac{\Gamma(8)}{\Gamma(6)} \cdot \frac{7 \cdot 6 \cdot \Gamma(6)}{9 \cdot 8 \cdot \Gamma(8)} = \\
 & \frac{7}{12}.
 \end{aligned}$$

Qui si è usato il fatto che vale la seguente formula ¹

$$\int_0^1 \theta^{\alpha-1} (1 - \theta)^{\beta-1} d\theta = \frac{\Gamma(\alpha) \Gamma(\beta)}{\Gamma(\alpha + \beta)}.$$

□

¹ Si veda la dimostrazione per la somma $\Gamma(\alpha, \lambda) + \Gamma(\beta, \lambda)$, dove $\Gamma(\alpha, \lambda), \Gamma(\beta, \lambda)$ sono stocasticamente indipendenti.

Esercizio 14.2. Gli eventi E_1, E_2, \dots sono stocasticamente indipendenti subordinatamente alla conoscenza del parametro aleatorio Θ con $\mathbf{P}(E_i|\Theta = \theta) = \theta$. La densità a priori di Θ è data da

$$\pi_0 = \begin{cases} K \theta^2 (1 - \theta) & 0 \leq \theta \leq 1 \\ 0 & \text{altrimenti.} \end{cases}$$

Si osservano i valori dei primi cinque eventi:

$$E_1 = 0, E_2 = 1, E_3 = 1, E_4 = 0, E_5 = 1.$$

1. Calcolare la costante K .
2. Calcolare la densità a posteriori di Θ e la probabilità a posteriori dell'evento ($\Theta < \frac{1}{2}$).
3. Calcolare la previsione a posteriori di $X = E_6 + E_7$ e le probabilità a posteriori di $E = E_6 E_7$ e di $F = E_6 \vee E_7$.

Soluzione 14.2. 1. Per calcolare la costante k si impone che l'integrale della densità sia pari a 1, ovvero

$$\int_0^1 \pi_0(\theta) d\theta = 1.$$

Ne segue che

$$k = \frac{1}{\int_0^1 \theta^2 (1 - \theta) d\theta}.$$

Il valore di tale integrale è noto e pari a:

$$\int_0^1 \theta^2 (1 - \theta) d\theta = \frac{\Gamma(3) \Gamma(2)}{\Gamma(3 + 2)},$$

da cui

$$k = \frac{\Gamma(3 + 2)}{\Gamma(3) \Gamma(2)} = \frac{4!}{2! \cdot 1!} = 12.$$

2. La densità a posteriori è data da

$$\begin{aligned} \pi_5(\Theta|E_1 = 0, E_2 = E_3 = 1, E_4 = 0, E_5 = 1) &= \\ \pi_0(\theta) \cdot \mathbf{P}(E_1 = 0, E_2 = E_3 = 1, E_4 = 0, E_5 = 1|\theta) &= \\ \pi_0(\theta) \cdot \mathbf{P}(E_1 = 0|\theta) \cdot \mathbf{P}(E_2 = 1|\theta) \mathbf{P}(E_3 = 1|\theta) \cdot \mathbf{P}(E_4 = 0|\theta) \cdot \mathbf{P}(E_5 = 1|\theta) &= \\ c \cdot \theta^2 (1 - \theta) \cdot (1 - \theta)^2 \theta^3 &= \\ c \cdot \theta^5 (1 - \theta)^3. & \end{aligned}$$

Si è indicato con c la costante di normalizzazione della densità a posteriori, che risulta essere una $\beta(6, 4)$. Ne segue che

$$c = \frac{\Gamma(6+4)}{\Gamma(6)\Gamma(4)} = \frac{9!}{5!3!} = 7 \cdot 8 \cdot 9 = 504.$$

Se si pone $W = \tilde{E}_1 E_2 E_3 \tilde{E}_4 E_5$, la densità a posteriori è data da

$$\pi_5(\theta) = \begin{cases} 504 \theta^5 (1-\theta)^3 & 0 \leq \theta \leq 1 \\ 0 & \text{altrimenti.} \end{cases}$$

Per trovare la probabilità a posteriori dell'evento $(\Theta < \frac{1}{2})$ basta dunque integrare la densità a posteriori fra 0 ed $\frac{1}{2}$:

$$\mathbf{P}(\Theta < \frac{1}{2}) = 504 \int_0^{\frac{1}{2}} \theta^5 (1-\theta)^3 d\theta = 504 \int_0^{\frac{1}{2}} (\theta^5 + 3\theta^7 - 3\theta^6 - \theta^8) d\theta = \frac{55}{256}.$$

3. La previsione a posteriori di $X = E_6 + E_7$ è data da

$$\begin{aligned} \mathbf{P}(X|W) &= \sum_{i=0}^2 i \mathbf{P}(X = i|W) \\ &= \mathbf{P}(X = 1|W) + 2 \mathbf{P}(X = 2|W). \end{aligned}$$

Si ottiene:

$$\begin{aligned} \mathbf{P}(X = 1|W) &= \mathbf{P}(E_6 = 1, E_7 = 0|W) + \mathbf{P}(E_6 = 0, E_7 = 1|W) \\ &= 2 \int_0^1 \mathbf{P}(E_6 = 0, E_7 = 1|\Theta = \theta) \pi_5(\Theta = \theta|W) d\theta \\ &= 2 \int_0^1 \frac{\Gamma(10)}{\Gamma(6)\Gamma(4)} \theta \cdot (1-\theta) \cdot \theta^5 \cdot (1-\theta)^3 d\theta \\ &= 2 \frac{\Gamma(10)}{\Gamma(6)\Gamma(4)} \int_0^1 \theta^6 (1-\theta)^4 d\theta \\ &= 2 \frac{\Gamma(10)}{\Gamma(6)\Gamma(4)} \frac{\Gamma(7)\Gamma(5)}{\Gamma(12)} \\ &= \frac{24}{55} \end{aligned}$$

ed inoltre

$$\begin{aligned}
\mathbf{P}(X = 2|W) &= \mathbf{P}(E_6 = 1, E_7 = 1|W) \\
&= \int_0^1 \mathbf{P}(E_6 = 1, E_7 = 1|\Theta = \theta)\pi_5(\Theta = \theta|W) d\theta \\
&= \frac{\Gamma(10)}{\Gamma(6)\Gamma(4)} \int_0^1 \theta^2 \cdot \theta^5 \cdot (1 - \theta)^3 d\theta \\
&= \frac{\Gamma(10)}{\Gamma(6)\Gamma(4)} \int_0^1 \theta^7 (1 - \theta)^3 d\theta \\
&= \frac{\Gamma(10)}{\Gamma(6)\Gamma(4)} \frac{\Gamma(8)\Gamma(4)}{\Gamma(12)} \\
&= \frac{21}{55}.
\end{aligned}$$

La previsione a posteriori di X è

$$\mathbf{P}(X|W) = 2 \cdot \frac{21}{55} + \frac{24}{55} = \frac{66}{55} = \frac{6}{5}.$$

Si noti che $X = E_6 + E_7$ è un numero aleatorio ma non un evento perché può assumere tre valori: 0, 1 oppure 2.

Le probabilità a posteriori degli eventi $E = E_6 E_7$ e $F = E_6 \vee E_7$ si calcolano con lo stesso metodo.

$$\mathbf{P}(E|W) = \mathbf{P}(E_6 E_7 = 1|W) = \mathbf{P}(E_6 = E_7 = 1|W) = \frac{21}{55}$$

ed inoltre

$$\begin{aligned}
\mathbf{P}(F|W) &= \mathbf{P}(E_6 \vee E_7 = 1|W) = \\
\mathbf{P}(E_6 = 1, E_7 = 0|W) &+ \mathbf{P}(E_6 = 0, E_7 = 1|W) + \mathbf{P}(E_6 = 1 = E_7 = 1|W) = \frac{9}{11}.
\end{aligned}$$

□

Esercizio 14.3. Gli eventi E_1, E_2, \dots sono stocasticamente indipendenti subordinatamente alla conoscenza del parametro aleatorio Θ con $\mathbf{P}(E_i|\Theta = \theta) = \theta$. La densità a priori di Θ è data da

$$\pi_0(\theta) = \begin{cases} K\theta^2(1-\theta)^2 & \text{per } 0 \leq \theta \leq 1 \\ 0 & \text{altrimenti} \end{cases}$$

Si osservano i valori dei primi 4 eventi: $E_1 = 0, E_2 = 1, E_3 = E_4 = 1$.

- Calcolare la costante K .
- Calcolare la densità e la previsione a posteriori di Θ .
- Calcolare la probabilità a posteriori dell'evento $F = E_5^2$ e la varianza a posteriori dell'evento \tilde{E}_6 .

Soluzione 14.3. a) La costante K è detta “di normalizzazione” perché rende l'integrale della parte funzionale della densità pari a 1. Per calcolare K , si impone quindi che

$$\int_0^1 \pi_0(\theta) d\theta = 1$$

ovvero

$$K = \frac{1}{\int_0^1 \theta^2(1-\theta)^2 d\theta}.$$

L'integrale al denominatore è noto ed è pari a

$$\int_0^1 \theta^2(1-\theta)^2 d\theta = \frac{\Gamma(3)^2}{\Gamma(6)}$$

ovvero

$$K = \frac{\Gamma(6)}{\Gamma(3)^2} = \frac{5!}{(2!)^2} = 30.$$

- La densità a posteriori di Θ dati gli eventi $E_1 = 0, E_2 = E_3 = E_4 = 1$ è data dalla formula

$$\begin{aligned} \pi_4(\theta|E_1 = 0, E_2 = 1, E_3 = 1, E_4 = 1) &= \\ K\pi_0\mathbf{P}(E_1 = 0, E_2 = 1, E_3 = 1, E_4 = 1|\theta) &= \\ \mathbf{P}(E_1 = 0|\theta) \cdot \mathbf{P}(E_2 = 1|\theta) \cdot \mathbf{P}(E_3 = 1|\theta) \cdot \mathbf{P}(E_4 = 1|\theta) &= \\ K\theta^5(1-\theta)^3. & \end{aligned}$$

dove $K = \frac{\Gamma(10)}{\Gamma(6)\Gamma(4)} = 504$ e $\theta \in [0, 1]$. Per $\theta \notin [0, 1]$, la densità a posteriori è pari a zero. Per calcolare la previsione a posteriori di Θ , si applica la formula della previsione per i numeri aleatori con distribuzione assolutamente continua, ovvero

$$\begin{aligned}
\mathbf{P}(\Theta|E_1 = 0, E_2 = E_3 = E_4 = 1) &= \int_0^1 \theta \pi_4(\theta|E_1 = 0, E_2 = 1, E_3 = 1, E_4 = 1) d\theta \\
&= \frac{\Gamma(10)}{\Gamma(6)\Gamma(4)} \int_0^1 \theta^6 (1-\theta)^3 d\theta \\
&= \frac{\Gamma(10)}{\Gamma(6)\Gamma(4)} \cdot \frac{\Gamma(7)\Gamma(4)}{\Gamma(11)} \\
&= \frac{3}{5}.
\end{aligned}$$

- c) L'evento $F = E_5^2$ coincide con E_5 in quanto assume solo il valore 0 o 1. La probabilità a posteriori di F è data da

$$\begin{aligned}
\mathbf{P}(F|E_1 = 0, E_2 = 1, E_3 = 1, E_4 = 1) &= \\
\mathbf{P}(E_5^2|E_1 = 0, E_2 = 1, E_3 = 1, E_4 = 1) &= \\
\mathbf{P}(E_5|E_1 = 0, E_2 = 1, E_3 = 1, E_4 = 1) &= \\
\int_0^1 \theta \pi_4(\theta|E_1 = 0, E_2 = 1, E_3 = 1, E_4 = 1) d\theta &= \frac{3}{5}.
\end{aligned}$$

Per calcolare la varianza a posteriori dell'evento \tilde{E}_6 , si considera per prima cosa la solita formula per la varianza. Per semplificare la notazione si ponga $A = \tilde{E}_1 E_2 E_3 E_4$. Si ottiene

$$\begin{aligned}
\sigma^2(\tilde{E}_6|\tilde{E}_1 E_2 E_3 E_4) &= \\
\mathbf{P}(\tilde{E}_6^2|\tilde{E}_1 E_2 E_3 E_4) - \mathbf{P}(\tilde{E}_6|\tilde{E}_1 E_2 E_3 E_4)^2 &= \\
\mathbf{P}(\tilde{E}_6|\tilde{E}_1 E_2 E_3 E_4) - \mathbf{P}(\tilde{E}_6|\tilde{E}_1 E_2 E_3 E_4)^2 &= \\
\mathbf{P}(\tilde{E}_6|\tilde{E}_1 E_2 E_3 E_4)(1 - \mathbf{P}(\tilde{E}_6|\tilde{E}_1 E_2 E_3 E_4)), &
\end{aligned}$$

dove si è usato di nuovo il fatto che

$$\tilde{E}_6^2 = \tilde{E}_6.$$

Basta quindi calcolare la previsione a posteriori di \tilde{E}_6 . Per fare ciò, si osserva che non conosciamo l'evento \tilde{E}_5 , quindi bisogna applicare la formula delle probabilità totali (a posteriori). Quindi

$$\begin{aligned}
\mathbf{P}(\tilde{E}_6^2 | \tilde{E}_1 E_2 E_3 E_4) &= \mathbf{P}(\tilde{E}_6 | \tilde{E}_1 E_2 E_3 E_4) = \\
\mathbf{P}(\tilde{E}_6 E_5 | \tilde{E}_1 E_2 E_3 E_4) &+ \mathbf{P}(\tilde{E}_6 \tilde{E}_5 | \tilde{E}_1 E_2 E_3 E_4) = \\
\int_0^1 \mathbf{P}(\tilde{E}_6 E_5 | \theta) \pi_4(\theta | \tilde{E}_1 E_2 E_3 E_4) d\theta &+ \int_0^1 \mathbf{P}(\tilde{E}_6 \tilde{E}_5 | \theta) \pi_4(\theta | \tilde{E}_1 E_2 E_3 E_4) d\theta = \\
\int_0^1 \theta(1-\theta) \pi_4(\theta | \tilde{E}_1 E_2 E_3 E_4) d\theta &+ \int_0^1 (1-\theta)(1-\theta) \pi_4(\theta | \tilde{E}_1 E_2 E_3 E_4) d\theta = \\
\int_0^1 (1-\theta)[\theta + 1-\theta] \pi_4(\theta | \tilde{E}_1 E_2 E_3 E_4) d\theta &= \\
\int_0^1 (1-\theta) \pi_4(\theta | \tilde{E}_1 E_2 E_3 E_4) d\theta &= \\
1 - \int_0^1 \theta \pi_4(\theta | \tilde{E}_1 E_2 E_3 E_4) d\theta &= \frac{2}{5}.
\end{aligned}$$

Come si vede, le probabilità a posteriori di E_6 è la stessa di E_5 .

□

Esercizio 14.4. Gli eventi E_1, E_2, \dots sono stocasticamente indipendenti subordinatamente alla conoscenza del parametro aleatorio Θ con $\mathbf{P}(E_i|\Theta = \theta) = \theta$. La densità a priori di Θ è data da

$$\pi_0(\theta) = \begin{cases} K\theta^2\sqrt{1-\theta} & \text{per } 0 \leq \theta \leq 1 \\ 0 & \text{altrimenti.} \end{cases}$$

Si osservano i valori dei primi 4 eventi: $E_1 = 1, E_2 = 0, E_3 = 0, E_4 = 1$.

- Calcolare la costante K .
- Calcolare la densità a posteriori di Θ ed il suo punto di massimo.
- Calcolare la covarianza a posteriori degli eventi E_6 ed E_7 .

Soluzione 14.4. a) La costante K rende 1 l'integrale della densità, quindi

$$K = \frac{1}{\int_0^1 \theta^2(1-\theta)^{\frac{1}{2}} d\theta}.$$

Si sa che

$$\int_0^1 \theta^2(1-\theta)^{\frac{1}{2}} d\theta = \frac{\Gamma(3)\Gamma(\frac{3}{2})}{\Gamma(3 + \frac{3}{2})},$$

quindi

$$K = \frac{\Gamma(\frac{9}{2})}{\Gamma(3)\Gamma(\frac{3}{2})} = \frac{\frac{7}{2} \cdot \frac{5}{2} \cdot \Gamma(\frac{3}{2})}{2!\Gamma(\frac{3}{2})} = \frac{35}{8}.$$

Si ricordi la proprietà $\Gamma(\alpha + 1) = \alpha\Gamma(\alpha)$.

- La densità a posteriori si calcola utilizzando il fatto che gli eventi sono stocasticamente indipendenti subordinatamente alla conoscenza di Θ . Si ha che

$$\begin{aligned} \pi_4(\theta|E_1 = 1, E_2 = E_3 = 0, E_4 = 1) &= \\ K\mathbf{P}(E_1 = 1, E_2 = E_3 = 0, E_4 = 1|\theta)\pi_0(\theta) &= \\ K\mathbf{P}(E_1 = 1|\theta) \cdot \mathbf{P}(E_2 = 0|\theta) \cdot \mathbf{P}(E_3 = 0|\theta) \cdot \mathbf{P}(E_4 = 1|\theta)\pi_0(\theta) &= \\ K\theta^4(1-\theta)^{\frac{5}{2}}, & \end{aligned}$$

dove

$$K = \frac{\Gamma(5 + \frac{7}{2})}{\Gamma(5)\Gamma(\frac{7}{2})} = \frac{\frac{15}{2} \cdot \frac{13}{2} \cdot \frac{11}{2} \cdot \frac{9}{2} \Gamma(\frac{7}{2})}{\Gamma(5)\Gamma(\frac{7}{2})} = \frac{6435}{128}.$$

Quindi

$$\pi_4(\theta|E_1 = 1, E_2 = E_3 = 0, E_4 = 1) = \begin{cases} \frac{6435}{128}\theta^4(1-\theta)^{\frac{5}{2}} & \theta \in [0, 1] \\ 0 & \text{altrimenti.} \end{cases}$$

Il punto di massimo della densità si trova cercando gli zeri della derivata prima.

$$\begin{aligned}
 \pi'_4(\theta|E_1 = 1, E_2 = E_3 = 0, E_4 = 1) &= K \left[4\theta^3(1-\theta)^{\frac{5}{2}} - \frac{5}{2}\theta^4(1-\theta)^{\frac{3}{2}} \right] \\
 &= K\theta^3(1-\theta)^{\frac{3}{2}} \left[4(1-\theta) - \frac{5}{2}\theta \right] \\
 &= K\frac{\theta^3}{2}(1-\theta)^{\frac{3}{2}}[8-13\theta].
 \end{aligned}$$

Oltre che negli estremi dell'intervallo, la derivata si annulla quindi in

$$\bar{\theta} = \frac{8}{13}.$$

Poiché $\pi'_4 > 0$ per $\theta \in \left[0, \frac{8}{13}\right)$ e $\pi'_4 < 0$ per $\theta \in \left(\frac{8}{13}, 1\right]$, si ha che $\bar{\theta} = \frac{8}{13}$ è il punto di massimo della densità a posteriori.

c) La covarianza a posteriori degli eventi E_6 ed E_7 è data da:

$$\begin{aligned}
 \mathbf{cov}(E_6, E_7|E_1 = 1, E_2 = E_3 = 0, E_4 = 1) &= \\
 &= \mathbf{P}(E_6 E_7|E_1 = 1, E_2 = E_3 = 0, E_4 = 1) - \\
 &= \mathbf{P}(E_6|E_1 = 1, E_2 = E_3 = 0, E_4 = 1)\mathbf{P}(E_7|E_1 = 1, E_2 = E_3 = 0, E_4 = 1).
 \end{aligned}$$

Nell'esercizio 14.3 si è dimostrato che

$$\begin{aligned}
 \mathbf{P}(E_6|E_1 = 1, E_2 = E_3 = 0, E_4 = 1) &= \\
 \mathbf{P}(E_7|E_1 = 1, E_2 = E_3 = 0, E_4 = 1) &= \\
 \int_0^1 \theta \pi_4(\theta|E_1 = 1, E_2 = E_3 = 0, E_4 = 1) d\theta &= \frac{10}{17}.
 \end{aligned}$$

Analogamente

$$\begin{aligned}
 \mathbf{P}(E_6 E_7|E_1 = 1, E_2 = E_3 = 0, E_4 = 1) &= \\
 \int_0^1 \theta^2 \pi_4(\theta|E_1 = 1, E_2 = E_3 = 0, E_4 = 1) d\theta &= \frac{120}{323}.
 \end{aligned}$$

Si conclude quindi che

$$\mathbf{cov}(E_6, E_7|E_1 = 1, E_2 = E_3 = 0, E_4 = 1) = \frac{120}{323} - \left(\frac{10}{17}\right)^2.$$

□

Esercizio 14.5. I numeri aleatori X_1, X_2, \dots sono stocasticamente indipendenti subordinatamente alla conoscenza del parametro aleatorio Θ con densità subordinata marginale

$$f(x_i|\theta) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{(x_i - \theta)^2}{2}\right), \quad x_i \in \mathbb{R}.$$

La distribuzione a priori di Θ è una normale standard. Si osservano i valori dei primi quattro esperimenti

$$x_1 = 0.1, \quad x_2 = 2, \quad x_3 = -1, \quad x_4 = 0.5.$$

- Scrivere la densità a priori di Θ .
- Calcolare la densità a posteriori di Θ ed il suo punto di massimo.
- Calcolare previsione e varianza a posteriori di Θ .

Soluzione 14.5. a) Poiché Θ ha distribuzione a priori di normale standard, si può scrivere subito la sua densità a priori

$$\pi_0(\theta) = \frac{1}{\sqrt{2\pi}} e^{-\frac{\theta^2}{2}}, \quad \theta \in \mathbb{R}.$$

- Per calcolare la densità a posteriori di Θ si usa il fatto che i numeri aleatori sono stocasticamente indipendenti subordinatamente a Θ :

$$\begin{aligned} \pi_4(\theta | x_1 = 0.1, x_2 = 2, x_3 = -1, x_4 = 0.5) &= \\ k f(x_1, x_2, x_3, x_4 | \theta) \pi_0(\theta) &= \\ k \prod_{i=1}^4 f(x_i | \theta) \pi_0(\theta) &= k \exp\left(-\frac{\sum_{i=1}^4 (x_i - \theta)^2 + \theta^2}{2}\right) = \\ k \exp\left(-\frac{5}{2}\left(\theta - \frac{8}{25}\right)^2\right), \end{aligned}$$

dove si è messo nella costante k tutto ciò che non dipende da θ . Si ottiene che la distribuzione a posteriori è una gaussiana $N\left(\frac{8}{25}, \frac{1}{5}\right)$, quindi

$$k = \frac{\sqrt{5}}{\sqrt{2\pi}}.$$

La densità a posteriori ha come grafico la *curva a campana* di Gauss con asse di simmetria $x = \frac{8}{25}$. Il punto di massimo è dunque $\bar{x} = \frac{8}{25}$. Se ne può fare la verifica calcolando la derivata o tracciando il grafico della funzione.

- I parametri della distribuzione a posteriori ci forniscono inoltre
 - la previsione a posteriori

$$\mathbf{P}(\Theta | x_1 = 0.1, x_2 = 2, x_3 = -1, x_4 = 0.5) = \frac{8}{25};$$

2. la varianza a posteriori

$$\sigma^2(\theta | x_1 = 0.1, x_2 = 2, x_3 = -1, x_4 = 0.5) = \frac{1}{5}.$$

□

Esercizio 14.6. I numeri aleatori X_1, X_2, \dots sono stocasticamente indipendenti subordinatamente alla conoscenza del parametro aleatorio Θ con densità subordinata marginale

$$f(x_i|\theta) = \frac{1}{2\sqrt{2\pi}} \exp\left(-\frac{(x_i - \theta)^2}{8}\right), \quad x_i \in \mathbb{R}.$$

La densità a priori di Θ è data da

$$\pi_0(\theta) = \frac{1}{\sqrt{4\pi}} \exp\left(-\frac{(\theta - 1)^2}{4}\right), \quad \theta \in \mathbb{R}.$$

Si osservano i valori dei primi tre esperimenti

$$x_1 = 1, \quad x_2 = 0.5, \quad x_3 = -1.$$

- Calcolare il fattore di verosimiglianza.
- Calcolare la densità a posteriori di Θ .
- Stimare la probabilità a posteriori dell'evento ($\Theta > 1000$).

Soluzione 14.6. a) Il fattore di verosimiglianza è per definizione dato da

$$\begin{aligned} f(x_1, x_2, x_3|\theta) &= \prod_{i=1}^3 f(x_i|\theta) = \\ &= \frac{1}{8\sqrt{(2\pi)^3}} \exp\left(-\frac{\sum_{i=1}^3 (x_i - \theta)^2}{8}\right) = \\ &= \frac{1}{8\sqrt{(2\pi)^3}} \exp\left(-\frac{1}{8}\left(3\theta^2 - \theta + \frac{9}{4}\right)\right). \end{aligned}$$

- Dal calcolo del fattore di verosimiglianza si ricava subito la densità a posteriori nel seguente modo

$$\begin{aligned} \pi_3(\theta | x_1 = 1, x_2 = 0.5, x_3 = -1) &= \\ &= k f(x_1, x_2, x_3, x_4|\theta) \pi_0(\theta) = \\ &= k \exp\left(-\frac{1}{8}\left(3\theta^2 - \theta + \frac{9}{4}\right) - \frac{(\theta - 1)^2}{4}\right) = \\ &= k \exp\left(-\frac{5}{8}\left(\theta - \frac{1}{2}\right)^2\right), \end{aligned}$$

dove si è messo nella costante k tutto ciò che non dipende da θ . Si ottiene che la distribuzione a posteriori è una gaussiana $N\left(\frac{1}{2}, \frac{4}{5}\right)$ con costante di

normalizzazione $k = \frac{\sqrt{5}}{2\sqrt{2\pi}}$.

- Per stimare la probabilità a posteriori dell'evento ($\Theta > 1000$) si ricorre alla stima delle code per la distribuzione gaussiana standard. Prima però

bisogna esprimere Θ come funzione di una variabile aleatoria Y con distribuzione $N(0, 1)$. Poiché la distribuzione a posteriori di Θ è una gaussiana $N(\frac{1}{2}, \frac{4}{5})$, si ha che

$$\Theta = \frac{2\sqrt{5}}{5}Y + \frac{1}{2},$$

dove $Y \sim N(0, 1)$. Quindi

$$\mathbf{P}(\Theta > 1000) = \mathbf{P}\left(\frac{2\sqrt{5}}{5}Y + \frac{1}{2} > 1000\right) = \mathbf{P}\left(Y > \frac{\sqrt{5}}{2} \cdot 999,5\right).$$

Dalla stima delle code per la distribuzione gaussiana standard si ha che

$$\frac{n(x)}{x} - \frac{n(x)}{x^3} < \mathbf{P}(Y > x) < \frac{n(x)}{x},$$

dove $x > 0$, $n(x) = \frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}$. Per ottenere un valore approssimato (per eccesso) di $\mathbf{P}(\Theta > 1000)$ basta dunque calcolare per esempio $\frac{n(x)}{x}$ nel punto

$$x = \frac{\sqrt{5}}{2} \cdot 999,5.$$

□

Esercizio 14.7. I numeri aleatori X_1, X_2, \dots sono stocasticamente indipendenti subordinatamente alla conoscenza del parametro aleatorio Φ con densità subordinata marginale

$$f(x_i|\phi) = \frac{1}{\sqrt{2\pi}}\phi^{\frac{1}{2}} \exp\left(-\frac{\phi(x_i-1)^2}{2}\right), \quad x_i \in \mathbb{R}.$$

La distribuzione a priori di Φ è data da una gamma di parametri $\alpha = 2, \lambda = 1$. Si osservano i valori dei primi tre esperimenti

$$x_1 = 1.5, \quad x_2 = 0.5, \quad x_3 = 2.$$

- Scrivere la densità a priori di Φ .
- Calcolare la densità a posteriori di Φ ed il suo punto di massimo.
- Calcolare previsione e varianza a posteriori di Φ .

Soluzione 14.7. a) La distribuzione a priori di Φ è di tipo $\Gamma(2, 1)$, quindi la sua densità è

$$\pi_0(\phi) = \begin{cases} \phi e^{-\phi} & \phi \geq 0 \\ 0 & \phi < 0. \end{cases}$$

b) La densità a posteriori è data da

$$\begin{aligned} \pi_3(\phi | x_1 = 1.5, x_2 = 0.5, x_3 = 2) &= \\ k f(x_1, x_2, x_3 | \phi) \pi_0(\phi) &= \\ k \phi^{\frac{5}{2}} \exp\left(-\left(\frac{\sum_{i=1}^3 (x_i - 1)^2}{2} + 1\right)\phi\right) &= \\ k \phi^{\frac{5}{2}} e^{-\frac{7}{4}\phi} & \end{aligned}$$

per $\phi \geq 0$, 0 altrimenti. Si noti che si è messo nella costante k tutto ciò che non dipende da ϕ . La distribuzione a posteriori è dunque una gamma $\Gamma(\frac{7}{2}, \frac{7}{4})$ con costante di normalizzazione

$$k = \left(\frac{7}{4}\right)^{\frac{7}{2}} \frac{1}{\Gamma(\frac{7}{2})} = \frac{\sqrt{7^7}}{240\sqrt{\pi}}.$$

Per ottenere il punto di massimo della densità si possono calcolare gli zeri della derivata prima. Si ha che

$$\pi_3(\phi | x_1 = 1.5, x_2 = 0.5, x_3 = 2) = k \phi^{\frac{3}{2}} e^{-\frac{7}{4}\phi} \left(\frac{5}{2} - \frac{7}{4}\phi\right) = 0$$

se $\phi = \frac{10}{7}$. Discutendo il segno della derivata prima si ottiene subito che $\phi = \frac{10}{7}$ è il punto di massimo.

- c) I parametri della distribuzione a posteriori ci forniscono inoltre
1. la previsione a posteriori

$$\mathbf{P}(\Phi | x_1 = 1.5, x_2 = 0.5, x_3 = 2) = 2,$$

2. la varianza a posteriori

$$\sigma^2(\Theta | x_1 = 0.1, x_2 = 2, x_3 = -1, x_4 = 0.5) = \frac{8}{7}.$$

□

Esercizio 14.8. I numeri aleatori X_1, X_2, \dots sono stocasticamente indipendenti subordinatamente alla conoscenza del parametro aleatorio Φ con densità subordinata marginale

$$f(x_i|\phi) = \frac{1}{\sqrt{2\pi}} \phi^{\frac{1}{2}} e^{-\frac{\phi x_i^2}{2}}, \quad x_i \in \mathbb{R}.$$

La distribuzione a priori di Φ è data da una distribuzione esponenziale di parametro $\lambda = 2$. Si osservano i valori dei primi quattro esperimenti

$$x_1 = 1, \quad x_2 = 2, \quad x_3 = 0.5, \quad x_4 = \sqrt{2}.$$

- Scrivere la densità a priori di Φ e la probabilità a priori dell'evento $(\Phi > 2)$.
- Calcolare la densità a posteriori di Φ e la probabilità a posteriori dell'evento $(\Phi > 2)$.
- Calcolare la previsione a posteriori di $Z = \Phi^2$.

Soluzione 14.8. a) La distribuzione a priori di Φ è di tipo $\Gamma(1, 2)$, quindi la sua densità è

$$\pi_0(\phi) = \begin{cases} 2e^{-2\phi} & \phi \geq 0 \\ 0 & \phi < 0. \end{cases}$$

La probabilità a priori dell'evento $(\Phi > 2)$ è dunque

$$\mathbf{P}(\Phi > 2) = \int_2^{+\infty} 2e^{-2\phi} d\phi = e^{-4}.$$

- b) La densità a posteriori è data da

$$\begin{aligned} \pi_4(\phi | x_1 = 1, x_2 = 2, x_3 = 0.5, x_4 = \sqrt{2}) &= \\ &= kf(x_1, x_2, x_3, x_4 | \phi) \pi_0(\phi) = \\ &= k\phi^2 \exp\left(-\left(\frac{\sum_{i=1}^4 x_i^2}{2} + 2\right)\phi\right) = \\ &= k\phi^2 e^{-\frac{45}{8}\phi} \end{aligned}$$

per $\phi \geq 0$, 0 altrimenti. Si noti che si è messo nella costante k tutto ciò che non dipende da ϕ . La distribuzione a posteriori è dunque una gamma $\Gamma(\alpha_4, \lambda_4) = \Gamma(3, \frac{45}{8})$ con costante di normalizzazione

$$k = \left(\frac{45}{8}\right)^3 \frac{1}{\Gamma(3)} = \frac{45^3}{2^{10}}.$$

La probabilità a posteriori dell'evento $(\Phi > 2)$ è data da

$$\begin{aligned}
& \mathbf{P}(\Phi > 2 \mid x_1 = 1, x_2 = 2, x_3 = 0.5, x_4 = \sqrt{2}) = \\
& \int_2^{+\infty} \pi_4(\phi \mid x_1 = 1, x_2 = 2, x_3 = 0.5, x_4 = \sqrt{2}) d\phi = \\
& \qquad \qquad \qquad k \int_2^{+\infty} \phi^2 e^{-\frac{45}{8}\phi} d\phi = \\
& k \frac{8}{45} \left[-(\phi^2 + 2\phi + 2) e^{-\frac{45}{8}\phi} \right]_2^{+\infty} = \frac{3^4 5^3}{2^6} e^{-\frac{45}{4}}.
\end{aligned}$$

c) Per calcolare la previsione a a posteriori di $Z = \Phi^2$ basta osservare che

$$\begin{aligned}
& \mathbf{P}(\Phi^2 \mid x_1 = 1, x_2 = 2, x_3 = 0.5, x_4 = \sqrt{2}) = \\
& \sigma^2(\Phi \mid x_1 = 1, x_2 = 2, x_3 = 0.5, x_4 = \sqrt{2}) + \\
& \mathbf{P}(\Phi \mid x_1 = 1, x_2 = 2, x_3 = 0.5, x_4 = \sqrt{2}) = \\
& \qquad \qquad \qquad \frac{\alpha_4}{\lambda_4^2} + \frac{\alpha_4^2}{\lambda_4^2} = \frac{256}{675}.
\end{aligned}$$

□

Appendici

A

Elementi di calcolo combinatorio

Si consideri un insieme $\Omega = \{a_1, \dots, a_n\}$ contenente n elementi. Si ricorda che il simbolo $\binom{n}{r}$ si dice *coefficiente binomiale* e vale che $\binom{n}{r} = \frac{n!}{r!(n-r)!}$.

A.1 Disposizioni

Si vogliono contare i modi di scegliere r elementi da un insieme di n elementi con ripetizione e tenendo conto dell'ordine, ovvero il numero delle *disposizioni* di r elementi su n . Si hanno

1° elemento $\longrightarrow n$ scelte

2° elemento $\longrightarrow n$ scelte

· ·
· ·
· ·

r° elemento $\longrightarrow n$ scelte.

In totale, le disposizioni sono $n \cdot n \cdots n = n^r$. Esse danno il numero di funzioni da un insieme di r elementi in un insieme di n elementi.

A.2 Disposizioni semplici

Si vogliono contare i modi di scegliere r elementi da un insieme di n elementi senza ripetizione e tenendo conto dell'ordine, ovvero il numero delle *disposizioni semplici* di r elementi su n . Si hanno

1° elemento $\longrightarrow n$ scelte

2° elemento $\longrightarrow (n - 1)$ scelte

3° elemento $\longrightarrow (n - 2)$ scelte

· ·
· ·
· ·

r° elemento $\longrightarrow (n - r + 1)$ scelte.

In totale, le disposizioni semplici sono $n \cdot (n - 1) \cdots (n - r + 1) = \frac{n!}{n - r!}$ e si indicano con il simbolo D_r^n oppure $(n)_r$. Esse danno il numero di funzioni iniettive da un insieme di r elementi in un insieme di n elementi. Se $r = n$, si parla di *permutazioni*.

A.3 Combinazioni semplici

Si vogliono contare i modi di scegliere r elementi da un insieme di n elementi senza ripetizione e senza tener conto dell'ordine, ovvero il numero delle *combinazioni semplici* di r elementi su n . Data una combinazione semplice di r elementi su n , si ottengono $r!$ disposizioni permutando gli r elementi. Il numero delle combinazioni è allora

$$\frac{1}{r!} D_r^n = \frac{n!}{r!n - r!} = \binom{n}{r}.$$

Esse danno il numero di funzioni iniettive da un insieme di r elementi in un insieme di n elementi con immagine diversa.

A.4 Combinazioni

Si vogliono contare i modi di scegliere r elementi da un insieme di n elementi con ripetizione e senza tener conto dell'ordine, ovvero il numero delle *combinazioni* di r elementi su n . Data una combinazione $\{a_1, \dots, a_n\}$, senza perdere di generalità si può supporre $a_1 \leq \dots \leq a_n$. Si costruisce a partire da essa una combinazione semplice di r elementi in $n + r - 1$ elementi nel modo seguente:

$$\begin{aligned}
 b_1 &= a_1 \\
 b_2 &= a_2 + 1 \\
 &\cdot \quad \cdot \\
 &\cdot \quad \cdot \\
 &\cdot \quad \cdot \\
 b_r &= a_r + r - 1.
 \end{aligned}$$

Viceversa, data una combinazione semplice si può ad essa associare una combinazione. Le r -combinazioni sono quindi tante quante le r -combinazioni semplici in $n + r - 1$, ovvero sono $\binom{n+r-1}{r}$.

A.5 Coefficiente multinomiale

Il numero di modi di formare k gruppi di r_1, \dots, r_k elementi ciascuno in modo tale che $r_1 + \dots + r_k = n$ è dato dal *coefficiente multinomiale*

$$\frac{n!}{r_1!r_2!\cdots r_k!}.$$

Per formare il primo gruppo di r_1 elementi, vi sono $\binom{n}{r_1}$ modi. Per il secondo gruppo, esso si può formare scegliendo gli elementi in $\binom{n-r_1}{r_2}$ modi. Si procede analogamente per formare i gruppi restanti. Si ottiene

$$\binom{n}{r_1} \binom{n-r_1}{r_2} \cdots \binom{n-r_1-\cdots-r_{k-1}}{r_k} = \frac{n!}{r_1!r_2!\cdots r_k!}.$$

B

Dalle distribuzioni discrete a quelle assolutamente continue

Di sotto viene riportata una tabella che mette in evidenza le analogie che si possono individuare confrontando la teoria delle distribuzioni discrete con quella delle distribuzioni assolutamente continue.

C. Discreto		C. Ass. Continuo
Probabilità $P(X = x)$	\longrightarrow	Densità $f(x)$
Funzione di $\sum_{i \in I(X), i \leq x} P(X = i)$	ripartizione \longrightarrow	$P(X \leq x)$ $\int_{-\infty}^x f(s) ds$
$\sum_{i \in I(X)} i P(X = i)$	Previsione di X \longrightarrow	$\int_{-\infty}^{+\infty} s f(s) ds$
Previsione di $\sum_{i \in I(X)} \Psi(i) P(X = i)$	$Y = \Psi(X)$ \longrightarrow	$\int_{-\infty}^{+\infty} \Psi(s) f(s) ds$
$\sum_{i \in I(X), i \in A} P(X = i)$	$P(X \in A)$ \longrightarrow	$\int_A f(s) ds$

C

Schema delle principali distribuzioni di
probabilità

C.1 Distribuzioni discrete

Distribuzione	$I(X)$	$P(X = k)$	$P(X)$	$\sigma^2(X)$
Bernoulli p	$\{0, 1\}$	$P(X = 1) = p$	p	$p(1 - p)$
Binomiale $B(n, p)$	$\{0, \dots, n\}$	$\binom{n}{k} p^k (1 - p)^{n-k}$	np	$np(1 - p)$
Geometrica p	$\{1, 2, \dots\}$	$p(1 - p)^{k-1}$	$\frac{1}{p}$	$\frac{1 - p}{p^2}$
Ipergeometrica (n, N, b)	$\{n - n \vee (N - b), \dots, n \wedge b\}$	$\frac{\binom{b}{k} \binom{N - b}{n - k}}{\binom{N}{n}}$	$\frac{b}{nN}$	$n \left(\frac{N - n}{N - 1} \right) \frac{b}{N} \left(1 - \frac{b}{N} \right)$
Poisson λ	\mathbb{N}	$\frac{\lambda^k}{k!} e^{-\lambda}$	λ	λ

C.2 Distribuzioni assolutamente continue

Distribuzione	$I(X)$	Densità	$P(X)$	$\sigma^2(X)$
Uniforme $[a, b]$	$[a, b]$	$\frac{1}{b-a} I_{[a,b]}$	$\frac{a+b}{2}$	$\frac{(b-a)^2}{12}$
Esponenziale λ	\mathbb{R}^+	$\lambda e^{-\lambda x} I_{\{x \geq 0\}}$	$\frac{1}{\lambda}$	$\frac{1}{\lambda^2}$
Normale Std. $N(0, 1)$	\mathbb{R}	$\frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$	0	1
Normale Gen. $N(\mu, \sigma^2)$	\mathbb{R}	$\frac{1}{\sqrt{2\pi}\sigma^2} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$	μ	σ^2
Gamma $\Gamma(\alpha, \beta)$	\mathbb{R}^+	$\frac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\lambda x} I_{\{x \geq 0\}}$	$\frac{\alpha}{\lambda}$	$\frac{\alpha}{\lambda^2}$
Beta $\beta(\alpha, \beta)$	$[0, 1]$	$\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1} (1-x)^{\beta-1}$	$\frac{\alpha}{\alpha+\beta}$	$\frac{\alpha\beta}{(\alpha+\beta)^2 (\alpha+\beta+1)}$

D

La distribuzione normale n -dimensionale

Densità	$f(x_1, \dots, x_n) = k e^{(-\frac{1}{2}Ax \cdot x + b \cdot x)}$ $x = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}, A \in \mathcal{S}(n), b = \begin{pmatrix} b_1 \\ \vdots \\ b_n \end{pmatrix}$
Costante di Normalizzazione	$k = \sqrt{\frac{\det A}{(2\pi)^n}} e^{-\frac{1}{2}A^{-1}b \cdot b}$
Previsione	$\mathbf{P}(X) = A^{-1}b \Rightarrow \mathbf{P}(X_i) = (A^{-1}b)_i$
Matrice di Varianza e Covarianza	$C = A^{-1}$
Distribuzione marginale di X_i ($i = 1, \dots, n$)	$X_i \sim N((A^{-1}b)_i, [A^{-1}]_{ii})$

E

La formula di Stirling

In questa appendice ci occupiamo della *formula di Stirling*, che ci dà il comportamento asintotico di $n!$ al crescere di n . Vale infatti che

$$\text{formula di Stirling: } n! = \sqrt{2\pi} n^{n+\frac{1}{2}} e^{-n} (1 + O(n^{-1})).$$

Per dimostrarla, sono state fornite varie dimostrazioni. Qui ne presentiamo due, quella classica ed una che prova un risultato più generale, di cui la formula di Stirling rappresenta un caso particolare.

Cominciamo con la dimostrazione classica come la si può trovare in [F1] che riportiamo qui per comodità.

E.1 Prima dimostrazione

Questa dimostrazione ottiene la formula di Stirling a meno della costante $(\sqrt{2\pi})$. Questo valore si può ottenere come conseguenza del teorema di De Moivre-Laplace visto nel capitolo 5 approssimando la probabilità che un numero aleatorio con distribuzione binomiale di parametri $2n, \frac{1}{2}$ assuma il valore n .

La formula di Stirling è equivalente ad affermare che

$$\lim_{n \rightarrow \infty} \frac{n!}{\sqrt{2\pi} n^{n+\frac{1}{2}} e^{-n}} = 1.$$

Per capire come calcolare tale limite, si cerca una sorta di stima per il logaritmo del fattoriale

$$\log n! = \log(1 \cdot 2 \cdot \dots \cdot n) = \log 1 + \log 2 + \dots + \log n.$$

Il $\log x$ è una funzione monotona crescente che può essere approssimata nel modo seguente

$$\int_{k-1}^k \log x dx < \log k \cdot 1 < \int_k^{k+1} \log x dx.$$

Quindi, passando alle somme

$$\sum_{k=1}^n \int_{k-1}^k \log x dx < \sum_{k=1}^n \log k < \sum_{k=1}^n \int_k^{k+1} \log x dx,$$

si ha

$$n \log n - n < \log n! < (n+1) \log(n+1) - n.$$

Questa disuguaglianza ci suggerisce di approssimare $\log n!$ con

$$\left(n + \frac{1}{2}\right) \log n - n.$$

Si può infatti pensare che $(n + \frac{1}{2}) \log n$ rappresenti una sorta di media. Ponendo

$$d_n = \log n! - \left(n + \frac{1}{2}\right) \log n - n = \log \left(\frac{n!}{n^{n+\frac{1}{2}} e^{-n}}\right),$$

si ha

$$d_n - d_{n+1} = \left(n + \frac{1}{2}\right) \log \left(\frac{n+1}{n}\right) - 1, \quad (\text{E.1})$$

ma

$$\frac{n+1}{n} = \frac{1 + \frac{1}{2n+1}}{1 - \frac{1}{2n+1}}$$

ed inoltre

$$\log(x+1) = \sum_{n=1}^{\infty} (-1)^{n+1} \frac{x^n}{n}. \quad (\text{E.2})$$

Poiché

$$\log \left(\frac{n+1}{n}\right) = \log \left(\frac{1 + \frac{1}{2n+1}}{1 - \frac{1}{2n+1}}\right) = \log \left(1 + \frac{1}{2n+1}\right) - \log \left(1 - \frac{1}{2n+1}\right),$$

usando lo sviluppo in serie di potenze (E.2) con $x = \pm \frac{1}{2n+1}$ si ottiene

$$\begin{aligned} d_n - d_{n+1} &= \frac{1}{2}(2n+1) \left[\log \left(1 + \frac{1}{2n+1}\right) - \log \left(1 - \frac{1}{2n+1}\right) \right] - 1 = \\ &= \frac{1}{2}(2n+1) \left[\frac{2}{(2n+1)} + \frac{2}{3(2n+1)^3} + \frac{2}{5(2n+1)^5} + \dots \right] - 1 = \\ &= \frac{1}{3(2n+1)^2} + \frac{1}{5(2n+1)^4} + \dots, \end{aligned}$$

da cui

$$d_n - d_{n+1} > 0$$

e quindi d_n risulta decrescente. Ne segue che il limite di d_n esiste (finito od infinito). Per dimostrare che il limite esiste finito, si procede nel seguente modo. Si nota che:

$$0 < d_n - d_{n+1} < \frac{1}{3} \sum_{k=1}^{\infty} \left(\frac{1}{2n+1} \right)^{2k} = \frac{1}{3} \left[\frac{1}{1 - \frac{1}{(2n+1)^2}} - 1 \right]$$

$$= \frac{1}{3} \frac{1}{(2n+1)^2 - 1} = \frac{1}{12n} - \frac{1}{12(n+1)},$$

ovvero la successione

$$a_n = d_n - \frac{1}{12n}$$

risulta crescente. Poiché

$$a_n \leq d_n \quad \forall n \in \mathbb{N}, n \neq 0$$

e vale che

$$\lim_{n \rightarrow \infty} a_n = \lim_{n \rightarrow \infty} \left(d_n - \frac{1}{12n} \right) = \lim_{n \rightarrow \infty} d_n,$$

si ottiene che il limite di d_n esiste finito in quanto le due successioni a_n e d_n si limitano l'un l'altra.

E.2 Dimostrazione tramite la funzione Gamma

Consideriamo la funzione Gamma di Eulero data da

$$\text{funzione Gamma: } \Gamma(\alpha) = \int_0^{+\infty} x^\alpha e^{-x} dx$$

dove $\alpha > 0$. Essa rappresenta una generalizzazione del fattoriale in quanto per ogni $\alpha > 0$ vale che

$$\Gamma(\alpha + 1) = \alpha \Gamma(\alpha),$$

come si verifica facilmente integrando per parti. Se α è un numero naturale, per iterazione si ottiene

$$\Gamma(n + 1) = n!.$$

Per dimostrare la formula di Stirling si prova quindi il risultato più generale

$$\Gamma(\alpha + 1) = \sqrt{2\pi} \alpha^{\alpha + \frac{1}{2}} e^{-\alpha} (1 + O(\alpha^{-1})).$$

Consideriamo il logaritmo della funzione integranda $\phi(x) = \log x^\alpha e^{-x} = \alpha \log x - x$. Facciamo lo sviluppo di Taylor di $\phi(x)$ nel punto di massimo α :

$$\phi(x) = \alpha \log \alpha - \alpha - \frac{1}{2\alpha}(x-\alpha)^2 + \sum_{k=3}^n \frac{(-1)^{k-1}}{k} \frac{(x-\alpha)^k}{\alpha^{k-1}} + \alpha \frac{(-1)^n}{n+1} \frac{(x-\alpha)^{n+1}}{\xi^{n+1}},$$

dove $\xi \in [\alpha, x]$. Effettuiamo nell'integrale il cambio di variabile

$$u = \frac{x-\alpha}{\sqrt{\alpha}}, \quad dx = \sqrt{\alpha} du.$$

Si ottiene

$$\Gamma(\alpha+1) = \alpha^{\alpha+\frac{1}{2}} e^{-\alpha} \int_{-\sqrt{\alpha}}^{+\infty} e^{-\frac{u^2}{2} + \psi(u)} du,$$

dove

$$\psi(u) = \sum_{k=3}^n \frac{(-1)^{k-1}}{k} \frac{u^k}{\alpha^{\frac{k}{2}-1}} + \alpha^{\frac{n+3}{2}} \frac{(-1)^n}{n+1} \frac{u^{n+1}}{(\alpha + \xi\sqrt{\alpha})^{n+1}}$$

con $\xi \in [0, u]$. Dividiamo l'integrale in tre parti:

$$I_1 = [-\sqrt{\alpha}, -\alpha^\delta], \quad I_2 = [-\alpha^\delta, \alpha^\delta], \quad I_3 = [\alpha^\delta, +\infty],$$

dove $\delta > 0$ è una costante opportunamente piccola. Per quanto riguarda I_1, I_3 osserviamo che la funzione $\phi(u)$ è concava e quindi risulta tale anche la funzione

$$\theta(u) = -\frac{u^2}{2} + \psi(u),$$

che si ottiene da ϕ con l'aggiunta di una costante ed un cambiamento lineare di variabile. Per $u \leq -\alpha^\delta$ abbiamo quindi

$$\theta(u) \leq -\frac{u}{\alpha^\delta} \theta(-\alpha^\delta)$$

e per $u \geq \alpha^\delta$

$$\theta(u) \leq \frac{u}{\alpha^\delta} \theta(\alpha^\delta).$$

Dall'espansione di $\psi(u)$ con $n = 2$ vediamo che per α abbastanza grande e $\delta < \frac{1}{6}$ abbiamo $\theta(-\alpha^\delta) < -\frac{\alpha^{2\delta}}{4}$, $\theta(\alpha^\delta) < -\frac{\alpha^{2\delta}}{4}$ e quindi vale che per $|u| \geq \alpha^\delta$

$$\theta(u) \leq -|u| \frac{\alpha^\delta}{4}.$$

Ne segue che

$$\begin{aligned} \int_{I_1} e^{\theta(u)} du + \int_{I_3} e^{\theta(u)} du &\leq \int_{|u| \geq \alpha^\delta} e^{-|u| \frac{\alpha^\delta}{4}} du \\ &= \left[-\frac{8}{\alpha^\delta} e^{-|u| \frac{\alpha^\delta}{4}} \right]_{\alpha^\delta}^{+\infty} = \frac{8}{\alpha^\delta} e^{-\frac{\alpha^{2\delta}}{4}}. \end{aligned}$$

Consideriamo ora I_2 . Se scegliamo $n = 3$ si ottiene:

$$e^{\psi(u)} = \exp\left(\frac{1}{3} \frac{u^3}{\alpha^{\frac{1}{2}}} - \frac{1}{4} \frac{\alpha^3 u^4}{(\alpha + \xi \sqrt{\alpha})^4}\right) = 1 + \frac{1}{3} \frac{u^3}{\alpha^{\frac{1}{2}}} + O\left(\frac{u^4}{\alpha}\right)$$

con $\xi \in [0, u] \subset I_2$ e per $|u| < \alpha^\delta$. Ne segue che

$$\begin{aligned} \int_{I_2} e^{-\frac{u^2}{2} + \psi(u)} du &= \int e^{-\frac{u^2}{2}} du - \int_{I_2^c} e^{-\frac{u^2}{2}} du + \frac{\alpha^{-\frac{1}{2}}}{3} + \int_{I_2} u^3 e^{-\frac{u^2}{2}} du + O(\alpha^{-1}) \\ &= \sqrt{2\pi} + O(\alpha^{-1}) \end{aligned}$$

e quindi

$$\Gamma(\alpha + 1) = \sqrt{2\pi} \alpha^{\alpha + \frac{1}{2}} e^{-\alpha} (1 + O(\alpha^{-1})).$$

F

Richiami di Analisi

In questa appendice si richiamano definizioni e concetti dell'analisi in una variabile necessarie per la comprensione del testo e lo svolgimento degli esercizi.

F.1 Limite di una successione

Sia $(a_n)_{n \in \mathbb{N}}$ una successione di numeri reali. Essa si dice

1. *convergente* se

$$\lim_{n \rightarrow \infty} a_n = L < \infty,$$

ovvero se $\forall \epsilon > 0 \exists N = N(\epsilon) \mid \forall n > N$

$$|a_n - L| < \epsilon.$$

2. *divergente* se

$$\lim_{n \rightarrow \infty} a_n = +\infty,$$

ovvero se $\forall M > 0 \exists N = N(M) \mid \forall n > N$

$$a_n > M.$$

Una successione può anche non essere né convergente né divergente. Per esempio, la successione $a_n = (-1)^n$ oscilla fra 1 e -1 .

F.2 Limiti per le funzioni

Una funzione $f : \mathbb{R} \rightarrow \mathbb{R}$ ha

1. *limite finito* in x

$$\lim_{y \rightarrow x} f(y) = L < \infty.$$

Tale scrittura ha il seguente significato: $\forall \epsilon > 0 \exists \delta = \delta(\epsilon) \mid \forall y, |y-x| < \delta$

$$|f(y) - L| < \epsilon.$$

2. *limite infinito* in x

$$\lim_{y \rightarrow x} f(y) = +\infty.$$

Tale scrittura ha il seguente significato: $\forall M > 0 \exists \delta = \delta(M) \mid \forall y, |y-x| < \delta$

$$f(y) > M.$$

F.3 Limiti notevoli

Si ricordano i seguenti limiti notevoli:

1.

$$\lim_{n \rightarrow \infty} \left(1 + \frac{1}{n}\right)^n = e.$$

2. $\forall x \in \mathbb{R}$

$$\lim_{n \rightarrow \infty} \left(1 + \frac{x}{n}\right)^n = e^x.$$

3.

$$\lim_{x \rightarrow 0} \frac{\log(1+x)}{x} = 1.$$

F.4 Serie notevoli

Si ricordano le seguenti serie notevoli:

1. La serie *geometrica*

$$\sum_{n=0}^{\infty} x^n = \frac{1}{1-x}$$

per ogni $|x| < 1$.

2. La serie “derivata” della serie geometrica

$$\sum_{n=1}^{\infty} nx^{n-1} = \frac{1}{(1-x)^2}$$

per ogni $|x| < 1$.

3. La serie *esponenziale*

$$\sum_{n=0}^{\infty} \frac{x^n}{n!} = e^x$$

per ogni $x \in \mathbb{R}$.

F.5 Continuità

Una funzione si dice *continua* in un punto x_0 se

$$\lim_{x \rightarrow x_0^-} f(x) = \lim_{x \rightarrow x_0^+} f(x) = f(x_0),$$

dove $\lim_{x \rightarrow x_0^-} f(x)$, $\lim_{x \rightarrow x_0^+} f(x)$ si dicono rispettivamente *limite sinistro* e *limite destro* di f in quanto il primo di essi viene fatto per gli $x < x_0$, il secondo per gli $x > x_0$.

F.6 Le derivate

Funzione $f(x)$	Derivata $f'(x)$
x^n	$n x^{n-1}$
e^x	e^x
$\log x$	$\frac{1}{x}$
$\sin x$	$\cos x$
$\cos x$	$-\sin x$
$e^{-\frac{x^2}{2}}$	$-x e^{-\frac{x^2}{2}}$

F.7 Tabella delle principali regole di derivazione

$\frac{d}{dx} [f(x) + g(x)]$	$f'(x) + g'(x)$
$\frac{d}{dx} [f(x)g(x)]$	$f'(x)g(x) + f(x)g'(x)$
$\frac{d}{dx} \left[\frac{f(x)}{g(x)} \right]$	$\frac{f'(x)g(x) - f(x)g'(x)}{g^2(x)}$
$\frac{d}{dx} [f(g(x))]$	$f'(g(x)) \cdot g'(x)$

F.8 Gli integrali**1. Formula di integrazione per parti**

$$\int_a^b f(x)g'(x) dx = [f(x)g(x)]_a^b - \int_a^b f'(x)g(x) dx.$$

2. Cambio di variabili

$$x = g(y) \Rightarrow dx = g'(y) dy$$

$$\int_a^b f(x) dx = \int_{g^{-1}(a)}^{g^{-1}(b)} f(g(y))g'(y) dy.$$

G

Integrali bidimensionali

In questa appendice si richiamano in breve alcune nozioni di analisi delle funzioni di più variabili necessarie per la comprensione del testo e lo svolgimento degli esercizi.

G.1 Area delle figure bidimensionali

Sia A una regione del piano; la sua area è data da

$$\text{area } A = \int \int_A dx dy,$$

in modo analogo al caso unidimensionale, in cui la lunghezza del segmento $[a, b]$ è data da

$$l([a, b]) = \int_a^b dx.$$

G.2 Integrale delle funzioni in due variabili

Sia $f : \mathbb{R}^2 \rightarrow \mathbb{R}$ ovvero $z = f(x, y)$. Una funzione in due variabili descrive una superficie in \mathbb{R}^3 di coordinate $(x, y, f(x, y))$. Si vuole calcolare il volume compreso fra la superficie descritta dalla funzione e il piano xy . Tale volume è dato dall' integrale doppio

$$\int \int_{\mathbb{R}^2} f(x, y) dx dy.$$

Si può anche considerare l'integrale su una regione $\Omega \subset \mathbb{R}^2$; intuitivamente, l'idea è che il volume del solido di base Ω ed altezza data dalla funzione $f(x, y)$ sia decomponibile in volumi infinitesimi tali che:

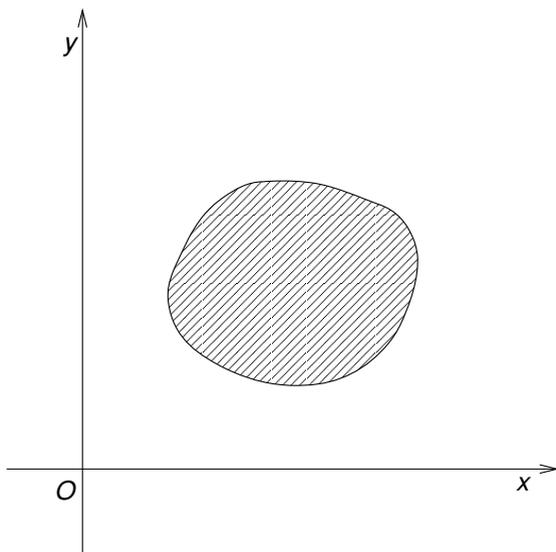


Figura G.1. Una generica regione del piano.

$$\text{volume } \Omega = \sum f(x, y) \Delta x \Delta y.$$

L'integrale è il limite per Δx e Δy che tendono a 0 di tali somme.

$$\sum \inf_{\Delta x \Delta y} f \Delta x \Delta y \leq I(f)_\Omega \leq \sum \sup_{\Delta x \Delta y} f \Delta x \Delta y.$$

In pratica, per calcolare gli integrali doppi si usa il *teorema di Fubini-Tonelli* che ci permette di calcolarli come due integrali in una dimensione annidati l'uno dentro l'altro.

Esempio G.2.1. Sia $A = \{1 < x < 2, 3 < y < 4\}$. Calcoliamo il seguente integrale su A .

$$\begin{aligned} \int \int_A x^2 y \, dx dy &= \int_1^2 dx \underbrace{\int_3^4 x^2 y \, dy}_{x \text{ è un parametro!}} \\ &= \int_1^2 x^2 \left(\int_3^4 y \, dy \right) dx \\ &= \int_1^2 x^2 \left[\frac{1}{2} y^2 \right]_3^4 dx \\ &= \frac{7}{2} \int_1^2 x^2 dx \\ &= \frac{49}{6}. \end{aligned}$$

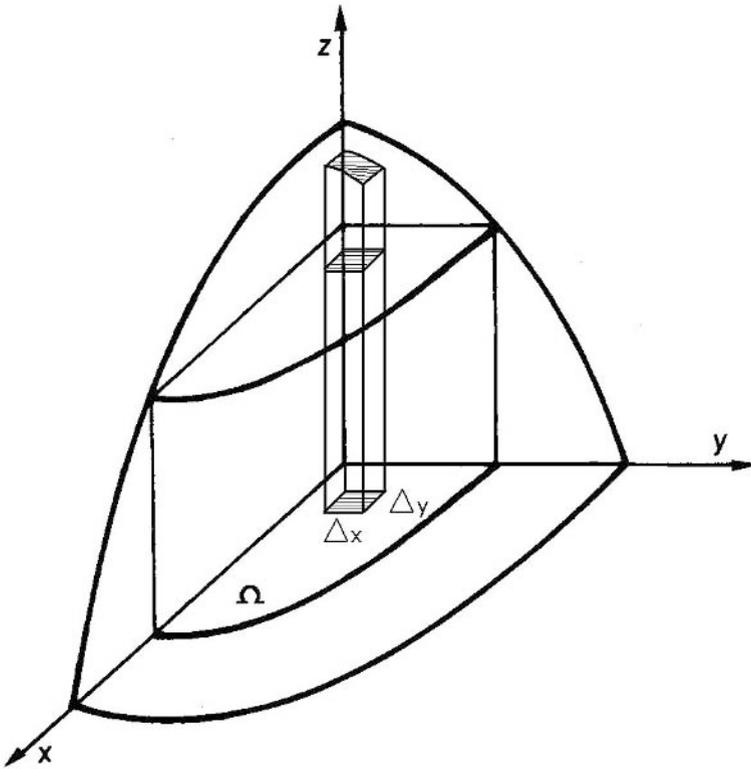


Figura G.2. Una funzione in due variabili integrata sull'area Ω del piano xy .

Esempio G.2.2. Sia $B = \{0 < x < 1, x - 1 < y < x + 1\}$ rappresentato in figura G.3.

Calcoliamo il seguente integrale su B .

$$\begin{aligned}
 \iint_B e^{-y} dx dy &= \int_0^1 dx \int_{x-1}^{x+1} e^{-y} dy \\
 &= \int_0^1 [e^{-y}]_{x-1}^{x+1} dx \\
 &= \int_0^1 (e^{-(x+1)} - e^{-(x-1)}) dx \\
 &= \int_0^1 (e^{-1} - e) e^{-x} dx \\
 &= (e^{-1} - e) (1 - e^{-1}) .
 \end{aligned}$$

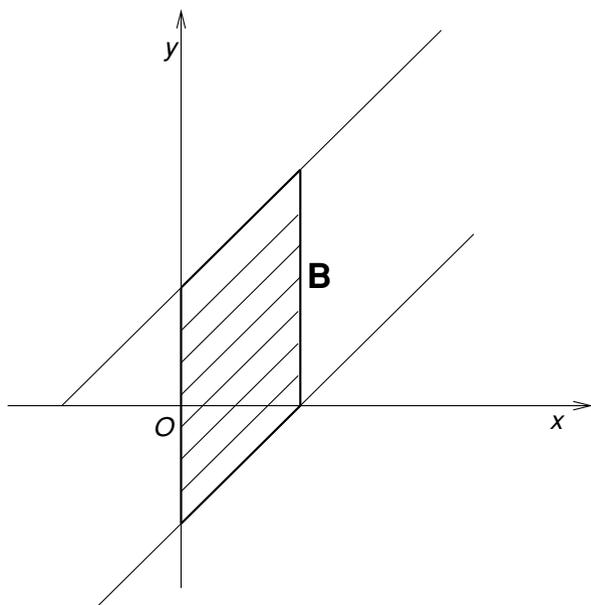


Figura G.3. La regione del piano individuata da B .

Esempio G.2.3. Talvolta il problema è quello di suddividere opportunamente il dominio. Cambiare l'ordine di integrazione, ovvero integrare in un ordine diverso da quello con cui sono specificati i differenziali delle variabili, non cambia il risultato dell'integrale e può rendere più semplice il calcolo dell'integrale, come nell'esempio della figura G.4, dove $D = \{0 < y < 1, y-1 < x < -y+1\}$.

Calcoliamo l'integrale di una generica funzione $f(x, y)$ su D .

$$\begin{aligned} \iint_D f(x, y) \, dx dy &= \int_0^1 dy \int_{y-1}^{-y+1} f(x, y) \, dx \\ &= \int_0^1 dx \int_0^{-x+1} f(x, y) \, dy + \int_{-1}^0 dx \int_0^{x+1} f(x, y) \, dy. \end{aligned}$$

Al primo passaggio, gli estremi di integrazione si trovano tracciando le parallele all'asse x e trovando i punti di intersezione di queste con il bordo della regione D . Al secondo passaggio invece, l'integrale è stato spezzato in due parti e gli estremi sono stati trovati con lo stesso metodo di prima tracciando però le parallele rispetto all'asse delle y .

G.3 Derivate parziali rispetto ad una variabile

Sia $f : \mathbb{R}^2 \rightarrow \mathbb{R}$, $z = f(x, y)$. Si dice *derivata parziale* della funzione f rispetto alla variabile x e si scrive

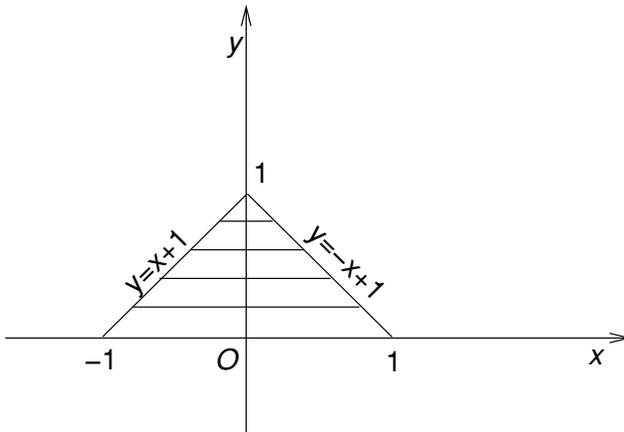


Figura G.4. Dominio individuato da D

$$\frac{\partial f}{\partial x}$$

la derivata di f ottenuta considerando la funzione come dipendente solo dalla variabile x considerando le altre variabili come se fossero parametri. In modo analogo si definiscono le derivate parziali di una funzione rispetto alle altre variabili.

Esempio G.3.1 (Derivate parziali).

1. $f(x, y) = x^2y$

$$\frac{\partial f}{\partial x} = 2xy \quad \frac{\partial f}{\partial y} = x^2.$$

2. $f(x, y) = \log(xy)$

$$\frac{\partial f}{\partial x} = \frac{1}{x} \quad \frac{\partial f}{\partial y} = \frac{1}{y}.$$

G.4 Cambio di variabili

Definite le derivate parziali di una funzione rispetto ad una variabile, è ora possibile definire il cambio di variabili per le funzioni in due dimensioni.

Sia $\Psi : \mathbb{R}^2 \rightarrow \mathbb{R}^2$, $(x, y) = (\Psi_1(x, y), \Psi_2(x, y))$; si dice *Jacobiano* della funzione di trasformazione Ψ e si indica con la notazione J_Ψ la matrice così definita:

$$J_\Psi = \begin{pmatrix} \frac{\partial \Psi_1}{\partial x} & \frac{\partial \Psi_1}{\partial y} \\ \frac{\partial \Psi_2}{\partial x} & \frac{\partial \Psi_2}{\partial y} \end{pmatrix}.$$

Un cambiamento di coordinate in \mathbb{R}^2 è dato da una funzione

$$\Psi : \mathbb{R}^2 \rightarrow \mathbb{R}^2$$

$$(u, v) \longmapsto (x, y)$$

con determinate proprietà di regolarità (diffeomorfismo). Per cambiare le coordinate negli integrali, si usa la regola:

$$\int \int_A f(x, y) \, dx dy = \int \int_{\Psi^{-1}(A)} f(\Psi(u, v)) |\det J_\Psi| \, dx dy$$

aiutandosi con il seguente grafico

$$\begin{array}{ccc} \mathbb{R}^2_{(u,v)} & \xrightarrow{\Psi} & \mathbb{R}^2_{(x,y)} \\ & \searrow f \circ \Psi & \downarrow f \\ & & \mathbb{R} \end{array}$$

Esempio G.4.1. In questo esempio si richiama il calcolo della costante di normalizzazione della distribuzione normale standard in due dimensioni per portare un esempio del cambio di coordinate nella risoluzione di un integrale.

$$\int \int_{\mathbb{R}^2} e^{-\frac{1}{2}(x^2+y^2)} \, dx dy.$$

Per calcolare questo integrale è necessario passare alle coordinate polari:

$$x = \rho \cos \theta, \quad y = \rho \sin \theta$$

$$(\theta, \rho) \xrightarrow{\Psi} (x, y) = (\rho \cos \theta, \rho \sin \theta).$$

La matrice Jacobiana della trasformazione è

$$J_\Psi = \begin{pmatrix} \frac{\partial}{\partial \theta} \rho \cos \theta & \frac{\partial}{\partial \rho} \rho \cos \theta \\ \frac{\partial}{\partial \theta} \rho \sin \theta & \frac{\partial}{\partial \rho} \rho \sin \theta \end{pmatrix} = \begin{pmatrix} -\rho \sin \theta & \cos \theta \\ \rho \cos \theta & \sin \theta \end{pmatrix}.$$

Calcolando il determinante Jacobiano si ottiene

$$\det J_\Psi = -\rho(\sin^2 \theta + \cos^2 \theta) = -\rho,$$

ovvero

$$|\det J_\Psi| = \rho.$$

Ne segue che

$$\begin{aligned}
\int \int_{\mathbb{R}^2} e^{-\frac{1}{2}(x^2+y^2)} dx dy &= \int_0^{+\infty} d\rho \int_0^{2\pi} \rho e^{-\frac{1}{2}\rho^2} d\theta \\
&= \int_0^{+\infty} \rho e^{-\frac{1}{2}\rho^2} d\rho \int_0^{2\pi} d\theta \\
&= 2\pi \underbrace{\left[-e^{-\frac{1}{2}\rho^2} \right]_0^{+\infty}}_1 \\
&= 2\pi.
\end{aligned}$$

e quindi

$$\begin{aligned}
\left(\int_{-\infty}^{+\infty} e^{-\frac{x^2}{2}} dx \right)^2 &= \int_{-\infty}^{+\infty} e^{-\frac{x^2}{2}} dx \int_{-\infty}^{+\infty} e^{-\frac{y^2}{2}} dy \\
&= \int \int_{\mathbb{R}^2} e^{-\frac{1}{2}(x^2+y^2)} dx dy = 2\pi.
\end{aligned}$$

In conclusione

$$\int_{-\infty}^{+\infty} e^{-\frac{x^2}{2}} dx = \sqrt{2\pi}.$$

G.5 Sintesi dell'appendice

1. Teorema di *Fubini-Tonelli*

$$\int dx \int f(x, y) dy = \int dy \int f(x, y) dx = \int \int f(x, y) dx dy.$$

2. Cambio di variabili

Dato il cambio di variabili $x_i = f_i(y_1, \dots, y_n)$ $i = 1, \dots, n$ corrispondente Jacobiano

$$J_f = \det \begin{pmatrix} \frac{\partial f_1}{\partial x_1} & \dots & \frac{\partial f_1}{\partial x_n} \\ \vdots & & \vdots \\ \frac{\partial f_n}{\partial x_1} & \dots & \frac{\partial f_n}{\partial x_n} \end{pmatrix}.$$

Se $A \subset \mathbb{R}^n$, vale che

$$\int \int_A \Psi(x_1, \dots, x_n) dx_1 \dots dx_n = \int \int_{f^{-1}(A)} \Psi(f(y)) |J_f| dy.$$

Inoltre, per determinare i nuovi estremi di integrazione, si utilizza il *metodo delle rette normali*, alcuni esempi del quale si possono osservare nelle figure G.5 e G.6.

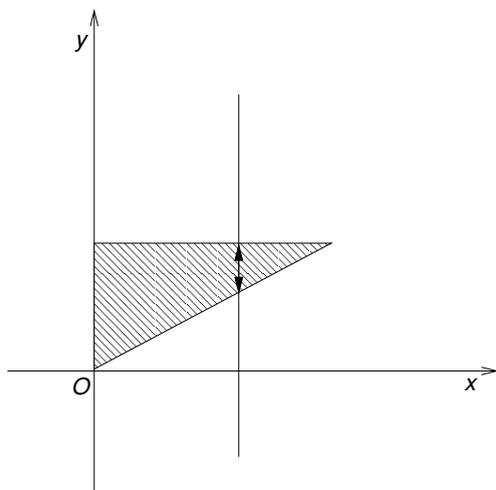


Figura G.5. Metodo delle rette normali quando varia la x .

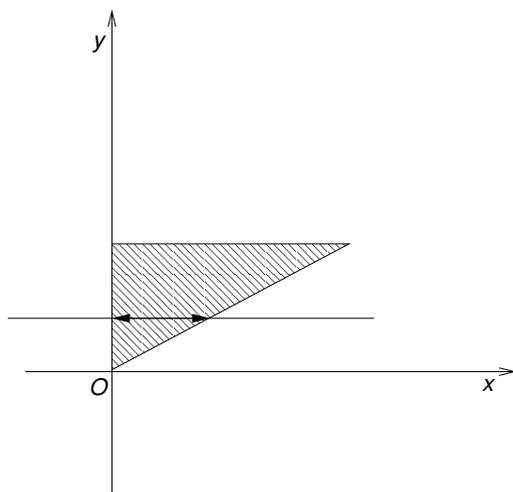


Figura G.6. Metodo delle rette normali quando varia la y .

Riferimenti bibliografici

- [B] P. Baldi. Calcolo delle Probabilità e Statistica. Mc Graw-Hill, 1992.
- [Br] L. Breiman. Probability. Addison-Wesley, 1968.
- [Da] L. Daboni. Calcolo delle Probabilità ed Elementi di Statistica. Utet, 1992.
- [dF] B. de Finetti. Teoria delle Probabilità. Einaudi, Torino, 1970.
- [F1] W. Feller. An Introduction to Probability Theory and Its Applications. Vol. I. Wiley, New York, 1957.
- [F2] W. Feller. An Introduction to Probability Theory and Its Applications. Vol. II. Wiley, New York, 1968.
- [FF] D. Foatà, A. Fuchs. Calcul des probabilités, Dunod, 2nda edizione, 1998.
- [G] B. Gnedenko. The Theory of Probability. Gordon and Braech Science Publishers, 1997.
- [HT] R. V. Hogg, E. A. Tanis. Probability and Statistical Inference. Prentice Hall, 2001.
- [JP] J. Jacod, P. Protter. Probability Essentials, Springer, 2003.
- [K] L. Kleinrock. Queueing Systems 1, John Wiley & Sons, 1975.
- [K2] L. Kleinrock, R. Gail. Queueing Systems: Problems and Solutions. John Wiley & Sons, 1996.
- [Lee] P.M. Lee, Bayesian Statistics: An Introduction. Edward Arnold, 1994.
- [Le] G. Letta. Probabilità Elementare : Compendio di Teoria, Problemi Risolti. Zanichelli, 1993.
- [L] D. V. Lindley. Introduction to Probability and Statistics. Vol. 1 e 2., Cambridge University Press, 1965.
- [R] S. M. Ross. Probabilità e Statistica per l'Ingegneria e le Scienze, Apogeo, 2003.

Indice analitico

- area, 241
- assenza di memoria, 27

- cambio di variabili, 245
- catena di Markov omogenea, 77
- cliente, 85
- coefficiente
 - binomiale, 219
 - multinomiale, 221
- coefficiente di correlazione, 21
- combinazioni, 220
- combinazioni semplici, 220
- complementare, 5
- condizioni al bordo assorbenti, 78
- continuità, 239
- convergenza per successioni di funzioni,
 - 69
- correlati
 - negativamente, 19
 - positivamente, 19
- costituente, 6
- covarianza, 19
- criterio di coerenza, 8

- densità
 - a posteriori, 100
 - congiunta, 55
 - di probabilità, 40
 - marginale, 55
 - subordinata, 99
- derivata, 239
- derivate parziali, 244
- disposizioni, 219
- disposizioni semplici, 219

- distribuzione
 - a priori, 100
 - assolutamente continua, 40
 - beta, 58
 - chi-quadro, 50
 - congiunta, 31
 - di Poisson, 28
 - di Student, 59
 - esponenziale, 44
 - gamma, 47
 - gaussiana n -dimensionale, 61
 - iniziale, 77
 - ipergeometrica, 29
 - multinomiale, 30
 - normale, 45
 - stazionaria, 92
 - uniforme, 42
 - binomiale, 26
 - di Cauchy, 51
 - discreta, 25
 - geometrica, 27
 - marginale, 32

- equazioni
 - di Chapman-Kolmogorov, 86
 - di Kolmogorov in avanti, 87
- esaustività, 6
- evento, 5
- evento logicamente
 - dipendente, 7
 - indipendente, 7
 - semidipendente, 7

- formula

- delle probabilità composte, 13
- di Bayes, 14
- di Stirling, 231
- funzione
 - di ripartizione, 39
 - di ripartizione congiunta, 53
 - di ripartizione marginale, 54
 - generatrice, 36
- incompatibilità, 6
- indipendenza logica, 4
- induzione statistica, 102
- inferiormente limitato, 3
- integrale doppio, 241
- intervalli di fiducia, 105
- Jacobiano, 245
- legge dei grandi numeri, 23
- limitato, 4
- limite, 237
- linearità, 8
- matrice di transizione, 77
- media campionaria, 23
- monotonia, 8
- non correlati, 19
- numero aleatorio, 3
- partizione, 6
- passeggiata aleatoria, 78
- penalità, 8
- permutazioni, 220
- plurievento, 29
- precisione, 102
- previsione, 7
- previsione subordinata, 13
- probabilità, 9
- probabilità di transizione, 78
- probabilità subordinata, 13
- processo
 - di Poisson, 87
 - stocastico, 85
- prodotto logico, 5
- regime stazionario, 96
- relazione di Little, 97
- schema di Bernoulli, 25
- scommessa, 7
- serie, 238
- sistema a coda, 85
- somma logica, 5
- spazio degli stati, 77
- sportello, 85
- stocasticamente indipendenti, 15
- successi, 25
- superiormente limitato, 3
- tempi di servizio, 85
- teorema
 - di De Moivre-Laplace, 72
 - di Fubini-Tonelli, 247
- valori possibili, 3
- vettore aleatorio, 31
- volume, 241

Springer - Collana Unitext

a cura di

Franco Brezzi
Ciro Ciliberto
Bruno Codenotti
Mario Pulvirenti
Alfio Quarteroni

Volumi pubblicati

A. Bernasconi, B. Codenotti
Introduzione alla complessità computazionale
1998, X+260 pp, ISBN 88-470-0020-3

A. Bernasconi, B. Codenotti, G. Resta
Metodi matematici in complessità computazionale
1999, X+364 pp, ISBN 88-470-0060-2

E. Salinelli, F. Tomarelli
Modelli dinamici discreti
2002, XII+354 pp, ISBN 88-470-0187-0

A. Quarteroni
Modellistica numerica per problemi differenziali (2a Ed.)
2003, XII+334 pp, ISBN 88-470-0203-6
(1a edizione 2000, ISBN 88-470-0108-0)

S. Bosch
Algebra
2003, VIII+380 pp, ISBN 88-470-0221-4

S. Graffi, M. Degli Esposti
Fisica matematica discreta
2003, X+248 pp, ISBN 88-470-0212-5

S. Margarita, E. Salinelli
MultiMath - Matematica Multimediale per l'Università
2004, XX+270 pp, ISBN 88-470-0228-1

A. Quarteroni, R. Sacco, F. Saleri
Matematica numerica (2a Ed.)
2000, XIV+448 pp, ISBN 88-470-0077-7
2002, 2004 ristampa riveduta e corretta
(1a edizione 1998, ISBN 88-470-0010-6)

A partire dal 2004, i volumi della serie sono contrassegnati da un numero di identificazione

13. A. Quarteroni, F. Saleri
Introduzione al Calcolo Scientifico (2a Ed.)
2004, X+262 pp, ISBN 88-470-0256-7
(1a edizione 2002, ISBN 88-470-0149-8)
14. S. Salsa
Equazioni a derivate parziali - Metodi, modelli e applicazioni
2004, XII+426 pp, ISBN 88-470-0259-1
15. G. Riccardi
Calcolo differenziale ed integrale
2004, XII+314 pp, ISBN 88-470-0285-0
16. M. Impedovo
Matematica generale con il calcolatore
2005, X+526 pp, ISBN 88-470-0258-3
17. L. Formaggia, F. Saleri, A. Veneziani
Applicazioni ed esercizi di modellistica numerica
per problemi differenziali
2005, VIII+396 pp, ISBN 88-470-0257-5
18. S. Salsa, G. Verzini
Equazioni a derivate parziali - Complementi ed esercizi
2005, VIII+406 pp, ISBN 88-470-0260-5
19. C. Canuto, A. Tabacco
Analisi Matematica I (2a Ed.)
2005, XII+448 pp, ISBN 88-470-0337-7
(1a edizione, 2003, XII+376 pp, ISBN 88-470-0220-6)
20. F. Biagini, M. Campanino
Elementi di Probabilità e Statistica
2006, XII+236 pp, ISBN 88-470-0330-X