Lecture Notes in Economics and Mathematical Systems 651

Thomas Wensing

Periodic Review Inventory Systems

Springer

Lecture Notes in Economics and Mathematical Systems 651

For further volumes:
http://www.springer.com/series/300

Thomas Wensing

Periodic Review Inventory Systems

Springer

Dr. Thomas Wensing
Catholic University Eichstätt-Ingolstadt
Chair of Production and Operations Management
Auf der Schanz 49
85049 Ingolstadt
Germany
thomas.wensing@ku-eichstaett.de

ISSN 0075-8442
ISBN 978-3-642-20478-4 e-ISBN 978-3-642-20479-1
DOI 10.1007/978-3-642-20479-1
Springer Heidelberg Dordrecht London New York

Library of Congress Control Number: 2011931729

Cover design: eStudio Calamar S.L.

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

Meinen Taufpaten Hans-Werner und Marianne Backhaus gewidmet.

Preface

This work arose during my occupation at the Chair of Production and Operations Management at the Catholic University Eichstätt–Ingolstadt, where it was accepted as my doctoral thesis. Professor Heinrich Kuhn served as the first and Professor Ulrich Küsters as the second reviewer. This thesis started off as my attempt to solve all the issues related to multi-level inventory management in discrete time once and for all. In retrospect, I am very glad that I took on the challenge. It taught me humility. Much as I wished to resolve everything in one fell swoop, I am now proud to be able to say that this is it: this is what is possible at the present time. We – the many researchers that came before me, and now myself – gave it our very best. I collated all the knowledge that was available, and hope to have driven the topic forwards a little. I admit the going was extremely tough, and wish my successors in this particular field of inventory theory courage and tenacity.

I am so very grateful to the many friends and colleagues who accompanied me on this path. It would be hard to pick out individuals – I would run the risk of leaving out far too many. Instead, I will not mention anyone specifically at all, and would like to simply express my heartfelt thanks to everyone who shared this journey with me, or contributed to it – I am sure you all recognize yourselves, and will understand.

I would, however, like to single out two names that I have already mentioned: Professor Heinrich Kuhn and Professor Ulrich Küsters. I owe the deepest appreciation to Professor Ulrich Küsters for his open and always constructive criticism. Readers should be thankful to him, too: his comments have spared them many of my excesses. And to more than anyone else, I am enormously grateful to my doctoral supervisor, Professor Heinrich Kuhn, for his guidance and latitude during the six years that I worked with him at his chair. His untiring commitment to teaching and research is a continuous inspiration. I am indebted to him as steadfast mentor, and know he will always remain true to himself. But now curtains up on the models, methods and approaches to the management of periodically reviewed inventory systems that await you on the following pages. I hope I have managed to clearly express my thoughts and ideas, and that I can perhaps encourage you, the reader, to work with them, and – even more importantly – to further refine them.

Ingolstadt *Thomas Wensing*
February 2011

Contents

List of Tables

Chapter 1
Introduction

1.1 Subject Matter

Controlling and maintaining inventories of physical goods has always been an issue for enterprises that consume materials in their goods or service production processes. While in the past this field has often been treated as a subordinate, merely operational task, recent developments in supply chain management have given rise to a strong demand for more profound methods for inventory management. The increasing division of labor coupled with the application of multi-national sourcing strategies have led to the formation of complex supply networks that need protection from random disturbances. Protection – but not at any price. The requirements that inventory management methods need to fulfil are thus easily defined: safeguard supply at an acceptable cost level.

Even the general problem is easily defined. As pointed out by Hadley and Whitin in their early and influential book, all theory on inventory management is dedicated to answering two fundamental questions: when to replenish the inventory, and how much replenishment to order (Hadley and Whitin 1963, p. 1). How difficult it is to give an appropriate answer to those two questions, however, differs from one type of inventory system to another. Another issue is that the performance of existing solution methods vitally depends on the nature of the inventory system being examined. It is not unusual for methods that are promising in one type of system to perform badly in another, or to even be completely useless. In contemporary inventory theory we therefore observe a focus on the development of multiple specific approaches to different representative systems, rather than on the search for one single approach that would comply with any type of inventory system. It is common believe in the research community nowadays that the (individual) nature of inventory systems is too complex to allow a unified, generic approach.

T. Wensing, *Periodic Review Inventory Systems*, Lecture Notes in Economics and Mathematical Systems 651, DOI 10.1007/978-3-642-20479-1_1,

1.2 Purpose and Problem Definition

The aim of this study is to assist the reader in making decisions on inventory management problems within supply chain structures. As outlined in the previous section, however, it is practically impossible to solve all potential problems in advance and create a multi-purpose toolbox for all needs. Instead of trying that, we have chosen a systematic modular approach to introduce a general way of thinking about problems in inventory theory and provide an understanding of structural coherence as well as causes and consequences. The reader will undoubtedly find a lot of formulae and specialized algorithms on specific problems relating to closely defined inventory systems. However, the structure and representation of these technical aspects is meant to provide an open framework that should allow modification or recombination depending on the specific needs of the analysis of comparable or even entirely different systems.

To be more specific, the core of this study is the analysis and optimization of a periodically reviewed inventory system in discrete time. Customer demand arrives in stochastic batches, and we observe stochastic replenishment lead times. The overall model describes a basic system type that is typically found within multi-level production or distribution systems. Reflecting the basic idea of a modular structure, the demand generating process substantially complies with the demand process that the supplier observes from our inventory system. The same holds – with limitations – for the replenishment process, which may exhibit varying delivery times. This variation may be caused by possible material unavailability, where the lead time delay may either be zero or – due to the underlying periodic time axis – an integer. Admittedly, the replenishment process does not perfectly match the customer waiting time process. Following almost the entire literature so far, we assume here that replenishment lead times are independent of the amounts ordered, where the customer waiting time clearly depends on how much is ordered from an inventory system.

Technically speaking, we examine a single-level inventory model that is operated on a discrete time axis according to a periodic review order-up-to policy, i.e., replenishment orders are regularly placed at fixed intervals and the order amount equals consumption since the last order was issued. Customer demand occurs in independent and identically distributed batches once in a constant demand interval. We assume full backordering, i.e., the customer never cancels an order, regardless of the waiting time observed. Replenishment lead times are discretely distributed, where we consider both the case of independent and dependent distributions.

In order to illustrate the analytical needs of the specific characteristics, we examine and implicitly contrast the two mentioned types of dependent and independent lead time processes on the supplier's side, as well as two demand delivery modes on the customer's side. Besides analyzing the specific system variants, we provide the reader with additional background on the analysis of inventory systems in discrete time. We include a broad introduction to the literature on single-level inventory models as well as a brief aside on some statistical and computer science methods that are helpful for analyzing inventory systems in discrete time. We also discuss basic aspects of modeling replenishment processes to provide further insights into thinking about inventory systems in a modular way.

The contribution of this work is threefold. First, we try to give access to the inner logic of the class of systems that we examine, and that we consider highly relevant for understanding and configuring inventory systems within multi-level supply chains. Second, we aim to provide ready-to-implement formulae and computational approaches that can help to improve decisions on practical inventory management wherever the system being considered is close enough or even identical to the one we analyze. Third, we contribute to inventory theory by presenting a comprehensive analysis of the system class considered. In that context, we summarize the research that has already been done on comparable systems and extend the available approaches or develop new ones wherever there is no existing solution method. The individual research issues thus emerge automatically from the attempt to conduct a systematic analysis of the inventory system variants examined. To our knowledge, analysis of order-oriented customer waiting times and the entire independent lead time case has not previously been performed for this system class. Furthermore, a comprehensive discussion of one family of inventory systems has never been conducted in this way before.

1.3 Outline

The outline of this study is as follows.

Chapter 2 provides an introduction to inventory systems, where the emphasis is on classifying different possible characteristics and describing cost and service metrics that are commonly used to evaluate the performance of inventory systems. This chapter maps out the field of our interest and defines the key indicators that we will use to describe the performance of an inventory system.

Chapter 3 contains an overview of the literature that we consider relevant for the understanding, analysis and optimization of systems of our type. We have chosen a rather wide scope, and give a comprehensive overview of important work on single-level inventory systems overall. This is primarily due to the fact that we have drawn some important ideas for our analysis from work on systems quite different to ours. One general observation, for example, is that several important findings on periodic review systems were first examined based on continuous review systems. We therefore consider it worthwhile to give at least a rough idea of the entire field of research on single-level inventory systems.

Chapter 4 presents some basic stochastic analysis methods and proceeds that are highly relevant for the analysis and optimization of inventory systems in discrete time. This chapter is meant to give guidance for implementing the basic analytical methods in an appropriate programming language such as Java, C++, C# or even Visual Basic. We also give insights into the corresponding computational complexity that is required for runtime examinations in the chapters after that.

Chapter 5 discusses different approaches to modeling the process of generating replenishment lead times. We examine the basic cases of dependent and independent lead times as well as two rather academic special forms. The latter may be especially helpful for testing new ideas where one may wish to exclude specific problems that emerge from a generally defined lead time process.

Chapter 6 constitutes the main part of this study. Here, we systematically discuss the analysis and optimization of the basic type of inventory system outlined above for different specifications relating the lead time and order fulfillment process.

Chapter 7 provides a conclusion and outlook.

Chapter 2
Concepts and Definitions

We first give a brief introduction to the objects and phenomena that we will examine in this study. The key object of our interest is an inventory system within a non-changing, static environment. In order to make any statements about this inventory system, we will first define how to describe the *state* of such a system in terms of the relevant characteristics. From an epistemological point of view, we will introduce the vocabulary here that enables us to talk about the object being examined. Next, we propose a framework to classify the structure and relevant elements of an inventory system within a static environment. This should help us to later on declare the specific nature of the system being analyzed and its variants. To evaluate the specific configuration of a system and allow for a preferential order with alternative configurations, the last section summarizes common metrics for evaluating the performance of inventory systems.

2.1 State of an Inventory System

From the pure perspective of keeping inventory, the following seven basic concepts are generally used in the literature to characterize the state of an inventory system at a certain moment t in time. In discrete time, these states commonly refer to the end of each period. Instead of inventory, a significant fraction of the literature uses the term *stock*, at least in some of the definitions. In the following, we use the term inventory to describe amounts of the goods in question, whereas stock refers to the installation (e.g., a warehouse) in which the goods are kept.

- *Physical inventory.* Physical inventory (I_t) is the amount of inventory that is immediately available to satisfy incoming demand. I_t is sometimes also referred to as inventory on hand.
- *Number of backorders.* The number of backorders (B_t) or simply backorder(s) is the volume of customer orders that currently cannot be satisfied due to inventory unavailability, and that are not lost. Backorders occur whenever customers are

T. Wensing, *Periodic Review Inventory Systems*, Lecture Notes in Economics and Mathematical Systems 651, DOI 10.1007/978-3-642-20479-1_2,

willing to wait for their orders, and the physical inventory is insufficient to cover their demand. The term *backlog* is often used as synonym for B_t in the literature.
- *Outstanding orders.* The (number of) outstanding orders K_t is the random number of orders that have been issued before time t, and have not yet arrived.
- *Inventory on order.* Inventory on order (IO_t) is the amount of material that has been ordered from the supplier but has not yet arrived. In other words, it is the total amount of material covered by outstanding orders.
- *Inventory shortfall.* The inventory shortfall (SF_t) is the difference between planned and the actual inventory levels. It is equal to the inventory on order plus the demand that has occurred since the last replenishment order was placed (see also Sect. 5.2.2.).
- *Net inventory.* Net inventory (NI_t) is defined as physical inventory minus backorders ($NI_t = I_t - B_t$).
- *Inventory position.* The inventory position (IP_t) is defined as physical inventory minus backorders plus inventory on order ($IP_t = I_t - B_t + IO_t$).

The state of an inventory system in terms of the concepts defined above may be changed by three events: order initiation, order arrival and demand occurrence. We thus have the following three concepts to describe the change of an inventory system at moment t in time, as compared to the previous moment. In our case of a discrete time axis, the previous moment is $t - 1$, so we have chosen to illustrate the definitions according to this paradigm:

- *Demand.* Demand (D_t) is the customer demand that is requested from the system in t. The demand may effect all of the state concepts described above.
- *Issued Inventory.* Issued inventory (II_t) is the amount of inventory that is ordered from suppliers in t. The issued inventory may effect the outstanding orders, the inventory on order and the inventory position (e.g. $IP_t = IP_{t-1} + II_t - D_t$).
- *Arriving Inventory.* Arriving inventory (AI_t) is the amount of inventory on order that finally arrives in t. The arriving inventory may effect all of the state concepts described above except for the inventory position (e.g., $IO_t = IO_{t-1} + II_t - AI_t$, $I_t = max\{0, NI_{t-1} + II_t - D_t\}$).

Finally, we have the concept of replenishment lead time which somewhat connects the issued and arriving inventory.

- *Replenishment lead time.* Replenishment lead time (L_k) is the time that passes between issue and arrival of an order indexed with k.

The concept of *safety stocks* (SS) is also frequently used in inventory management. It is commonly referred to as the amount of physical inventory that would never be undershot if replenishment lead times and demand were deterministic according to their expected values. In other words, *safety stock* is the amount of stock that is kept beyond expected requirements. From the perspective of analyzing a policy, SS is more of a performance indicator, even if it is often perceived as a parameter for configuring an inventory system. We do not adhere to this perspective in the following, instead, we will always define the configuration of an inventory system by declaring when and how much to order.

2.2 Classification

In the literature on inventory theory, various different systems are analyzed using diverse methods. Comparing the approaches on different systems reveals that even minor changes in the assumptions may lead to dramatic changes in systems behavior. It is therefore essential for the analysis to carefully define the nature of the system being examined.

This section presents a classification scheme for inventory systems with special consideration of the systems we analyze in Chap. 6. The organization of the scheme basically follows the proposal of Hollier and Vrat (1978). On the top level, we distinguish between the system structure, its environmental parameters and the replenishment policy that is applied. Similar schemes with less aggregation are given by Aggarwal (1974), Silver (1981) and Silver (2008). An even more detailed scheme is proposed by Prasad (1994). Deviating from Hollier and Vrat, we do not consider inventory-related costs as part of the system classification. The performance aspects (including costs) are therefore treated separately in Sect. 2.3.

We will briefly indicate for each item to what extent it is considered in the analysis in Chap. 6. These model limitations are also summarized in Sect. 2.2.4.

2.2.1 *Structure*

In terms of the overall structure, we primarily distinguish between *single-level* (or single-echelon) and *multi-level* models, where the latter may have a linear, converging, diverging or general structure. In a linear structure, each stock may have one predecessor (supplier) and one successor (receiver) at most. In converging systems, each stock may have multiple predecessors but only one successor at the most, whereas this is vice versa in diverging systems. Finally, a system is said to have a general structure if a stock may have multiple predecessors as well as multiple successors.

Furthermore, inventory systems may hold a *single item* or *multiple items*. In the event of multiple predecessors, the inventories of certain items may be replenished from a *single source* or from *multiple sources*.

In the following we will limit the analysis to the single-item and single-source case.

2.2.2 *Environmental Parameters*

Every inventory system is emptied by a demand process and refilled by a replenishment process, both of which significantly influence the systems behavior. Furthermore, the nature of the stored item may be of importance. The remainder of this section takes a closer look at these three aspects.

2.2.2.1 Demand Process and Customer Order Fulfilment

An inventory system can either face *stationary* or *dynamic* (time-varying) demand. Furthermore, the *interarrival time* between two demand occurrences as well as the *volume* of a single demand occurrence can be either *deterministic* or *stochastic*. In case of stochastic interarrival times and/or demand volumes, the underlying distribution may be fully known, only be known in type but not in parameterization or be completely unknown. Both, interarrival times and demand volumes may be *independent* from or *dependent* on the previous occurrences.

In the event that demand exceeds the physical inventory, customers are either utterly willing to wait (*full backordering*), may or may not be willing to wait (*partial backordering*), or may immediately lose interest (*lost sales*). If backordering is possible and physical inventory is only sufficient to partly cover the demand, the system may allow for *split deliveries* to the customer, or may only fulfill customer orders with *full deliveries*.

In the following, we will only consider stationary demand occurring with a constant interarrival time. Demand may be deterministic or stochastic, with the distribution completely known and independent from previous occurrences. Demand may always be backordered without limitation, and we address the case of split deliveries as well as that of full deliveries only.

2.2.2.2 Replenishment Process

Analogously to demand arrivals, the replenishments arrive after a certain (lead) time and consist of a certain amount. The lead time may be *stationary* or *dynamic*. It may furthermore be *deterministic* or *stochastic*, where in the latter case the distribution may be known, be known in type but not in parameterization, or may be completely unknown. Stochastic lead times may be considered *independent* from or *dependent* on the previous occurrence or even a history of occurrences.

Replenishment deliveries may either arrive in the exact amount as issued (full reliability), or may deviate with known, partially known or unknown distribution. Finally, there may or may not be dependencies between lead times and order amounts.

Please also refer to Chap. 5 for further distinctions concerning replenishment processes.

In the following we will assume that replenishment lead times are stationary, independent from order volumes and either deterministic or stochastic with known distributions. Amounts delivered may not deviate from what has been ordered. We also consider independent lead times as well as a case of lead time dependencies.

2.2.2.3 Stored Goods

Stored goods may be considered *non-perishable* or *perishable*. Perishable products may either be completely lost or decrease in value or usefulness, if kept in stock for too long.

We will always assume non-perishable goods in the following.

2.2.3 Replenishment Policies

In the introduction we outlined that inventory management always revolves around the two fundamental questions of when to replenish the inventory and how much to order. In the literature we commonly find two decision parameters for each of the two questions that may be combined to form an inventory policy. On behalf of the first question, we can either place an order every fixed period r or as soon as the inventory position falls below a particular value s. The order volume may either be a fixed quantity q or may be determined as the difference between a value S (the so-called order-up-to-level) and the inventory position.

Considering the task of determining current inventory levels, we may add the further decision of how often the inventory status should be determined. (See Silver et al. 1998, for example.) In that context, one may either want to establish a fixed review period t or continuously review the inventory levels ($t \to 0$).

From the combination of these possible decision parameters, we may derive five useful inventory policies as boldly highlighted in Table 2.1. Note that the possible (t, r, S) policy is not explicitly regarded in the literature as it would be useless in terms of the replenishment rule to review the inventory levels without having the option to place an order. Thus, $t := r$ is commonly assumed, where this (special) case is covered by the (r, S) policy. Furthermore, a non-adaptive policy such as (t, r, q) or (r, q) is not appropriate in a stochastic environment.

Thus, we derive the five basic decision rules printed in bold in Table 2.1, two of which apply continuous review and three of which apply periodic review.

The replenishment doctrines highlighted in bold can be described as follows. Remember that IP_t denotes the inventory position at time t.

- *(s,S)*: Order a variable amount of $Q_t = S - IP_t$ as soon as IP_t falls below s. This policy demands high standards of the supply and inventory review process. Changes in the IP must instantly be monitored, and it must be possible to order any amount in Q_t at any time. A special case of this policy, with $s := S - 1$ is discussed as base stock policy. Here, an order is placed as soon as one unit or more is taken from stock. Requirements of the supply and inventory review process are more or less equivalent to those of the (s, S) doctrine.

Table 2.1 Elementary inventory policies

No.	Monitoring t or 0	Impulse r or s	Quantity q or S	Policy
1	t	r	q	(t,r,q)
2	0	r	q	(r,q)
3	t	s	q	**(t,s,q)**
4	0	s	q	**(s,q)**
5	t	r	S	(t,r,S)
6	0	r	S	**(r,S)**
7	t	s	S	**(t,s,S)**
8	0	s	S	**(s,S)**

- *(s,q)*: Order a fixed quantity q as soon as IP_t falls below s.
  This policy reflects that the supplier may only offer certain discrete order batches of size q or an integer multiple of q. As for the (s, S) policy, instant review as well as the possibility that an order could always be placed is required for the application of this policy.
- *(t,s,S)*: Every period t, review IP_t and order a variable amount $Q_t = S - IP_t$ if IP_t is below s.
  This periodic review equivalent to the (s, S) policy requires that the supplier may deliver any amount in Q_t. Review effort is reduced to the possible order periods, where the main benefit for the process is the possibility of coordinating orders over time.
- *(t,s,q)*: Every period t, review IP_t and order the smallest multiple of the fixed quantity q that raises IP_t above s. This policy is also discussed as $(t, s, n \cdot q)$ policy to indicate that a multiple of q may be necessary to raise IP_t sufficiently. This periodic review equivalent to the (s, q) policy offers the possibility of coordinating order processes regarding time due to the fixed review interval t as well as regarding volume due to the fixed q.
- *(r,S)*: Every period r, order a variable amount of $Q_t = S - IP$.
  The characteristics of this policy are more or less equal to those of the (t, s, S) policy. In contrast, it lacks the possibility of skipping an order as a reaction to low demand occurrences between two order periods.
  This is the decision rule that we will examine in this study.

Comparing these policies, the (t, s, q) and (r, S) policy appear to be the most advantageous with respect to the supply and review processes. In turn, however, we are most likely observing the highest costs for operating those policies in terms of costs for inventory holding and backordering or lost sales.

2.2.4 Summary of Model Assumptions

This study considers a single-level, single-item and single-source inventory system with non-perishable goods. Replenishment orders are placed according to the periodic review order-up-to (r, S) doctrine.

From the customer side, our system observes stationary (static) stochastic or deterministic demand batches with a constant interarrival time. The distribution of the demand batches is known, and each realization is independent of previous occurrences. We assume full backordering, where two alternative delivery modes are addressed in the event of material insufficiency: we examine the option of split deliveries where a customer order may be delivered in two or more instalments as well as the restriction to the delivery of complete orders only.

From the supplier side, we observe stationary (static) stochastic or deterministic lead times that are independent of the volumes that we order. The arriving order amounts may not deviate from what has been ordered. We examine an independent lead time process as well as the case of interdependent lead time occurrences.

2.3 Performance Indicators

2.3.1 Costs

Assuming that the operating tasks of an inventory system range from material replenishment to customer order delivery, we may identify five types of relevant costs: the purchase costs of the replenishment material, the inventory holding costs, the costs for fulfilling customer demand, the costs that occur as the result of a stockout situation, and the costs of operating the inventory system itself, e.g., the effort required for data gathering and control procedures (Hadley and Whitin 1963, Chap. 1). A detailed description of the drivers of inventory system costs is given by Brooking (1987), for example.

In the more recent literature, we usually find the order fulfillment task reduced to holding the inventory available, where the actual delivery costs are not taken into account (Silver et al. 1998, Chap. 3).

For our scope, we will also neglect the costs of operating the system itself, as we will examine different parameterizations of the same basic system, where we assume identical operating costs. We thus retain the three *classical* types of costs that are considered for evaluating inventory system policies, namely purchasing costs, costs for inventory keeping, and costs that are incurred due to stockout situations.

2.3.1.1 Purchasing Costs

Purchasing costs may be incurred per unit and/or for an entire order, while both may depend on the size of the corresponding order. Dependence occurs for example, if the supplier offers volume discounts. In the following however, we will assume that costs per unit as well as costs for full orders are independent of order sizes, because this aspect is not what we want to focus our model on. In our case, acquisition costs cannot be influenced by the parameterization of an inventory system, and are thus irrelevant for the decision on order quantities. Nonetheless, we define them here because we need them for the proper calculation of inventory holding costs.

Definition 1 (Fixed acquisition price per unit). We define p as the fixed acquisition price per unit that is kept in stock.

The total order-related costs of replenishing inventory may well be influenced by the parameterization of an inventory system. These costs are obviously higher the more often replenishment orders are placed.

Definition 2 (Fixed costs per order). We define c_1 as fixed costs that are incurred for placing an order of arbitrary size.

2.3.1.2 Inventory Holding Costs

Definition 3 (Inventory holding costs). Let p be the acquisition price (Definition 1), i the interest rate per period, and h the costs of keeping one unit in stock for one period that do not include costs of tied capital. Then

$$c_2 = p \cdot i + h \tag{2.1}$$

are the costs that are incurred for keeping one unit of stock for one period.

The quantification of these costs for an inventory system requires computation of average stock levels for a certain time span that is fully representative for the system behavior. In our case of the (r, S) replenishment doctrine, this time span is one complete order cycle of length r.

From an accounting perspective, one might argue that the fixed order costs c_1 must also be considered proportionally within the expected purchase price, and thus influence c_1. Let Q be the stochastic order size that is implied by the application of a stock policy. Then we may state $c_2(.)$ depending on the order quantity as follows:

$$c_2(E[Q]) = \left(p + \frac{c_1}{E[Q]}\right) \cdot i + h. \tag{2.2}$$

Whenever the average inventory develops proportionally to the average order size, i.e., $I^{\o}(E[k \cdot Q]) = k \cdot I^{\o}(E[Q])$, we can easily show that the tied capital associated with the fixed order costs is independent of the order size Q:

$$\begin{aligned} c_2(E[k \cdot Q]) \cdot I^{\o}(E[k \cdot Q]) &= \left(p \cdot i + \frac{c_1 \cdot i}{E[k \cdot Q]} + h\right) \cdot I^{\o}(E[k \cdot q]) \\ &= \frac{c_1 \cdot I^{\o}(E[k \cdot Q]) \cdot i}{E[k \cdot Q]} + \left(p \cdot i + h\right) \cdot I^{\o}(E[k \cdot Q]) \\ &= \frac{c_1 \cdot k \cdot I^{\o}(E[Q]) \cdot i}{k \cdot E[Q]} + \left(p \cdot i + h\right) \cdot I^{\o}(E[k \cdot Q]) \\ &= \frac{c_1 \cdot I^{\o}(E[Q]) \cdot i}{E[Q]} + \left(p \cdot i + h\right) \cdot I^{\o}(E[k \cdot Q]) \end{aligned} \tag{2.3}$$

However, we cannot generally assume that $I^{\o}(E[k \cdot Q]) = k \cdot I^{\o}(E[Q])$ holds. Using a constant c_2 therefore must be considered an approximation under general conditions.

2.3.1.3 Costs of Stockout

When an inventory system fails to provide a required amount of material in time, some negative consequences have to be expected, otherwise the need for the inventory system is called into question. In the literature, three basic types of costs are discussed to estimate the consequences of unavailability of material. See Schneider (1981) for example.

Definition 4 (Costs of being in stockout state). We define c_{31} as the costs that are incurred if an inventory system is unable to provide material in the corresponding period, regardless of the amount of material missing and the previous duration of the stockout.

Relatively irrelevant from the perspective of the inventory system's customers, costs of type c_{31} may be incurred in connection with general arrangements to overcome the stockout situation. This could be the exceptional start of a production process, for example, or the request for an expensive emergency delivery, where the actual amount of missing units is secondary for the calculation of total costs. Even if the stockout does not induce any emergency activities, costs of type c_{31} may reflect a loss of customer goodwill as a reaction to the news that the system is in trouble.

Definition 5 (Costs per missing unit). We define c_{32} as the costs that are incurred per unit of material that cannot be provided in time.

Costs of type c_{32} are typically incurred if potential sales are lost in the event that immediate delivery is not possible.

Definition 6 (Costs per missing unit and time). We define c_{33} as the costs that are incurred if the delivery of one unit of material is delayed for one (further) period.

In contrast with c_{32}, c_{33} is only observed when unsatisfied demand can be delivered with a delay (backorder case).

It will generally be difficult to give reasonable estimates for all three types of stockout costs. This problem especially emerges from the wide range of possible customer reactions when demands cannot be satisfied. At one end of the scale, customers may be willing to wait without any compensation, while at the other, they may forever be lost as business partners. To circumvent the problem of estimating stockout costs, a common approach is to define service metrics, where the inventory system must fulfill a certain prescribed level.

2.3.2 Service Metrics

The pertinent literature discusses a variety of different elementary and compound service metrics. See Zinn et al. (2002) for an illustrative overview and Boylan and Johnston (1994) for insights into service measurement when orders may include multiple items.

We focus our scope on non-compound service metrics for single items. Following the logic that underlies our definition of stockout costs in Sect. 2.3.1.3, we consider one analogous service metric for each of the three types of costs that we have defined above. See Ronen (1982), for example, for a similar scope.

We have to distinguish two perspectives for determining the values of certain service metrics. On the one hand, we can obtain these values a posteriori from the observed history of a real or simulated inventory system. In this case, we may speak of the random final state of a stochastic process that has run for a certain time. On the other hand, we can try to estimate the expected values a priori from the parametrization of a modeled system. This is the scope of Chap. 6.

In our view, it is important to understand that the metrics defined in the following sections describe either observed or expected relative frequencies. They may be *interpreted* as probabilities if it can reasonably be assumed that the system will behave in the same manner in the future as it did in the past. See Hacking (2001), p. 127–139, on the different conceptions of probability, and Hacking (2006) for an overview of the historic development that also illustrates the problems of the different conceptions.

2.3.2.1 Ready Rate

Definition 7 (Ready rate). Let IS be an inventory system that is observed in a time interval of length τ. Let A^{τ} be a discretely distributed random variable with two possible states, where $A^{\tau} = 1$ means that the system had material on hand throughout the full time interval, and $A^{\tau} = 0$ means that the system was out of stock for at least a fraction of the time interval regarded. Then we define

$$\alpha^{\tau} = E[A^{\tau}] \tag{2.4}$$

as the ready rate for a basic interval of length τ, $\alpha^{\tau} \in [0, 1]$.

As indicated by the symbol, this event-oriented service level is often referred to as α-service level, especially in the European literature.

Let IS^* be a real or simulated inventory system with an underlying and probably unknown A^{τ}. Let $a^{\tau,i*}$ be the recorded system behavior in one interval i of length τ, where $a^{\tau,i*} = 1$ if all demands in i could immediately be fulfilled and $a^{\tau,i*} = 0$ otherwise, and finally let $\mathcal{I}$ be a set of observed time intervals. Then

$$a^{\tau*} = \frac{1}{|\mathcal{I}|} \sum_{i \in \mathcal{I}} a^{\tau,i*} \tag{2.5}$$

is an unbiased estimator for α^{τ}. We can interpret each $a^{\tau,i*}$ as the random draw from an underlying and probably unknown A^{τ}, while $a^{\tau*}$ is obviously the fraction of all observed periods in which no stockout has occurred. In the literature, $a^{\tau*}$ is sometimes referred to as *realized* ready rate or *realized* α service level, respectively. See for example Suchanek (1996) in Chap. 4 and Tempelmeier (2006) in Chap. B.3.

Various values for τ may apply. Especially in practical applications, τ is often chosen as equivalent to some period length critical for the planning (i.e., 1 day, week, month, ...). In this case, α^{τ} is referred to as the *periodic* ready rate. Two values for τ that are more inventory system-related are the replenishment lead time and the order interarrival rate. The service level corresponding to the first value is then referred to as *cyclic* ready rate, while the one corresponding to the latter value is equivalent to the probability that a customer order can be completely fulfilled without any delay.

If τ refers to a time unit smaller than an order cycle, we may observe different values for the *micro* periods of a cycle. For example, if replenishment orders always arrive in the second period of an order cycle, then the stockout probability for this period is most likely lower than it is for the previous period in which the replenishment material is still outstanding. It may therefore be helpful to differentiate the corresponding service levels for the periods of a subdivided order cycle.

For values of τ that are larger than the customer order interarrival rate, one or more stockout incidents might be considered as one event. Under special circumstances, the result may be influenced by the pattern we choose to jointly consider the periods of an order cycle. For example, let $SEQ_1 = 0000111100001111$ and $SEQ_2 = 1000011110000111$ be two sequences of inventory states that have been observed for two inventory systems, where 0 denotes unavailability and 1 availability of material. In the event that we consider four consecutive periods as one cycle, we observe $\alpha^{\tau=4} = 0.5$ for SEQ_1 and $\alpha^{\tau=4} = 0.0$ for SEQ_2 for the four fully observable cycles if we start with the first period and vice versa $\alpha^{\tau=4} = 0.0$ for SEQ_1 and $\alpha^{\tau=4} = 0.5$ for SEQ_2 if we start with the second period. This effect may only be observed if replenishment lead times are longer than the order cycle and demand is highly volatile. Nonetheless, it raises some doubt as to the expressiveness of the commonly used ready rate per replenishment cycle.

In the literature, the ready rate is also defined as the fraction of time for which the net inventory levels are positive, see Axsäter (2006) for example. However, we do not consider this definition appropriate for a discrete time axis. In this instance, we may consider time units τ in which more than one event can possibly alter the inventory levels, so that we may observe both positive and negative stock levels in the same underlying period.

For a discussion of fiscal-period-based versus lead-time-based measurement of the ready rate, see Haehling von Lanzenauer and Hamidi-Noori (1986). Insights on mathematical properties of a special ready rate metric are studied by Zipkin (1986a), for example.

2.3.2.2 Fill Rate

Definition 8 (Fill rate). Let IS be an inventory system that observes random demand D^τ in time intervals of length τ. Let $D^{\tau,p}$ be the random demand in τ that could be served by IS without delay. Then we define

$$\beta^\tau = \frac{E[D^{\tau,p}]}{E[D^\tau]} \tag{2.6}$$

as fill rate, $\beta^\tau \in [0, 1]$.

Let $D^{\tau,f} = D^{\tau} - D^{\tau,p}$ be the demand in periods of length τ that could not have been served in time. Then obviously

$$\beta^{\tau} = 1 - \frac{E[D^{\tau,f}]}{E[D^{\tau}]} \tag{2.7}$$

is equivalent to (2.6).

Proposition 1. *Let τ_b be a basic demand arrival interval observing i.i.d. demand occurrences of size D. Then (2.6) has an identical result for any value τ that is an integer multiple of τ_b, and $\beta = \beta^{\tau}$ for all $\tau = c \cdot \tau_b, c \in \mathbb{Z}^+$.*

Proof. Since the D are i.i.d. by definition, $E[D^{\tau}] = \tau \cdot E[D]$ holds for any $\tau = c \cdot \tau_b, c \in \mathbb{Z}^+$. The demand D^p that is satisfied in the basic period τ_b clearly depends on D. Furthermore, inventory systems in static environments typically observe a cyclic change of state, meaning that D^p also depends on the observed micro period of the characteristic cycle. Hence, the D^p are neither identically nor independently distributed. We may, however, always find a *macro* τ_m for which $D^{\tau_m,p}$ are i.i.d. due to the cyclic structure of the inventory system's behavior. Therefore, $E[D^{\tau,p}] = \tau \cdot E[D^p]$ also holds for any $\tau = c \cdot \tau_b, c \in \mathbb{Z}^+$ as long as the intervals of length τ do not systematically begin in specific micro periods, and thus equally represent all subperiods of τ_m. □

Thus, we have

$$\beta = \frac{E[D^{\tau,p}]}{E[D^{\tau}]} = \frac{\tau \cdot E[D^p]}{\tau \cdot E[D]} = \frac{E[D^p]}{E[D]}$$

for any $\tau = c \cdot \tau_b, c \in \mathbb{Z}^+$.

Note that the unit-oriented service fill rate is frequently referred to as β-service level, especially in the European literature.

Analogous to the ready rate that we discussed in the previous section, we also distinguish between the theoretical mean fill rate β described above and the *realized* fill rate β^*. For this purpose, let $IS^{\tau *}$ be a real or simulated inventory system that observes demand occurrences $d^{\tau,i*} \in D^*$ and fulfils demand $d^{\tau,p,i*} \in D^{\tau,p*}$ over time. Furthermore, let $\mathcal{I}$ be a set of observed time intervals. Then

$$\beta^* = \frac{\mu_{D^{\tau,p*}}}{\mu_{D^{\tau*}}}, \quad \mu_{D^{\tau,p*}} = \frac{1}{|\mathcal{I}|}\sum_{i \in \mathcal{I}} d^{\tau,p,i*}, \quad \mu_{D^{\tau*}} = \frac{1}{|\mathcal{I}|}\sum_{i \in \mathcal{I}} d^{\tau,i*} \tag{2.8}$$

is an unbiased estimator for β.

Besides the fill rate as defined here, the order fill rate is discussed in the literature as the percentage of customer orders that have been served in time. As already mentioned in Sect. 2.3.2.1, we consider this a special case of the ready rate, where τ is the replenishment order interarrival time and thus may also be a random variable.

Larsen and Thorstenson (2008) compare the latter service metric (defined as order fill rate) and the (volume) fill rate.

Just like the ready rate, the average fill rate may differ for certain subperiods of the review or replenishment cycle. In this context, Haehling von Lanzenauer and Hamidi-Noori (1986) and Haehling von Lanzenauer (1988) discuss the ready rate based on fiscal periods and propose a model that requires meeting the corresponding service level in every period of their fiscal-based scheme, not just the periods' average.

In this context, please also refer to the study of Zipkin (1986a) for a discussion of mathematical properties of the average number of backorders.

2.3.2.3 Time-Weighted Fill Rate

Definition 9 (Time-weighted fill rate). Let IS be an inventory system that observes random demand D^τ in time intervals of length τ (as in Definition 8), and let $D^{\tau,b}$ be a random variable denoting the customer demand that is outstanding due to insufficient stock (number of backorders). Then

$$\gamma = \gamma^\tau \quad 1 - \frac{E[D^{\tau,b}]}{E[D^\tau]} \tag{2.9}$$

is called the time-weighted fill rate, $\gamma \in [-\infty, 1]$.

This unit- and time-oriented service level is often referred to as γ-service level, especially in the European literature. Under the same assumptions as for β and using analogous argumentation, we can consider γ as independent from the underlying time interval τ. As for the ready rate and (classical) fill rate, we distinguish the theoretically expected time-weighted fill rate and the *realized* time-weighted fill rate. The latter is defined in the same manner as the fill rate, where we regard the number of backorders instead of the delayed or non-delayed demand. We will therefore not go into a detailed description here.

The interpretation of this metric is difficult because of two properties. First, γ is not bound to the left and may therefore not be interpreted as a fraction of any kind. Second, two performance aspects of an inventory system are inseparably mixed: the fill rate, and the customer waiting time.

As an illustration of the first problem, consider the following example. Let IS^* be an assemble-to-order system with fixed demand D, reorder cycle r, order quantity $r \cdot D$ and lead time $l = 0$ for simplicity. Then in each micro period of the order cycle, an amount D is backlogged and finally fulfilled in the order period. The expected backlog per period is thus given as:

$$E[D^b] = \frac{1}{r} \cdot \sum_{i=1}^{r} i \cdot D = \frac{D}{r} \cdot \sum_{i=1}^{r} i = \frac{D}{r} \cdot \frac{r \cdot (r+1)}{2} = \frac{D \cdot (r+1)}{2}.$$

The time-weighted fill rate for this system is given as follows:

$$\gamma = 1 - \frac{\frac{D\cdot(r+1)}{2}}{r \cdot D} = \frac{1-r}{2}$$

Thus, we have $\lim_{r\to\infty} \gamma = -\infty$, or more generally, γ converges with $-\infty$ in the (theoretical) case that an inventory system observes positive demand and does not deliver at all. Clearly, the time-weighted fill rate does not describe a relative frequency and may not be interpreted as probability. The metric must therefore be viewed critically, and we recommend considering the customer waiting times to analyze the time span that an item or order has to wait until it can be delivered.

2.3.2.4 Customer Waiting Times

Definition 10 (Customer waiting times per order). Let IS be an inventory system that completely serves a random number of orders O^τ in an interval τ, and let O_w^τ be the (complying) random number of orders that are completely served by IS after a delay of exactly w periods. Then we define

$$W^O: \quad P\{W^O = w\} = p_w^O = \frac{E[O_w^\tau]}{E[O^\tau]} \quad \forall \quad w \in W^O. \tag{2.10}$$

as the customer waiting time distribution for complete order fulfillment, where we follow the concept of frequency probability.

Let IS^* be a real or simulated inventory system with an underlying and probably unknown W^O. Let $o_w^{\tau *}$ be the number of orders in time interval τ that have been served with a delay of w periods. Then we have

$$p_w^{O*} = \frac{o_w^{\tau *}}{\sum_{w=0}^{\infty} o_w^{\tau *}}$$

as unbiased estimator for p_w^O, where p_w^{O*} describes the *realized* waiting time frequencies.

Analogously to the waiting times per customer order, we define the waiting times per order unit.

Definition 11 (Customer waiting times per part). Let IS be an inventory system that serves a random number of order units (parts) V^τ in an interval τ and let V_w^τ be the (complying) random number of parts that are completely served by IS after a delay of exactly w periods. Then we define

$$W^V: \quad P\{W^V = w\} = p_w^V = \frac{E[V_w^\tau]}{E[V^\tau]} \quad \forall \quad w \in W^V. \tag{2.11}$$

as the customer waiting time distribution per part delivered.

The corresponding *realized* waiting time frequencies p_w^{V*} are determined in the same manner as described for p_w^{O*}, where $v_w^{\tau*}$, the number of parts delivered with delay w in time interval τ, replaces the number of orders $o_w^{\tau*}$.

Chapter 3
Literature Overview

In the previous chapter we have defined the vocabulary that we will use to talk about inventory systems (Sect. 2.1), what structure and elements may constitute them (Sect. 2.2), and which performance metrics may apply to decide in the end which system configuration we prefer to the other (Sect. 2.3). We are therefore now ready to take a systematic look at the literature that appears relevant to our study.

Regarding the classification scheme specified in Sect. 2.2, we notice that many different inventory systems can be derived from the combination of possible alternatives for the various aspects. Unfortunately every single system usually comes with certain special characteristics that typically require specific analytical treatment. Research literature so far merely deals with this problem by elaborating individual approaches for each of the various systems one may reasonably think of. Particular systems are typically selected for examination because they either exhibit some feature of general interest, or because they are regarded as typical for a certain kind of practical application.

Considering this nature of the research landscape, it is relatively difficult to tightly define a rule to decide if work on a certain type of system is relevant for a particular research project or not. It is not necessarily the case that the work on systems most closely related to the one being examined contains the most helpful ideas for one's own research. Especially when following a modular approach that combines the solutions of a certain set of subproblems, one may find several different traits of research useful for solving the overall problem.

We therefore choose to give a broad overview of the literature on single-level inventory systems where we focus on the single-item case, and only marginally mention some relevant papers on multi-item systems. Although we consider a periodic review model with backorders, we also discuss the continuous review case and briefly reference to the lost sales case as well. We decided to cover the continuous review paradigm because many phenomena on inventory systems were first examined against the backdrop of continuously reviewed systems. The corresponding basic findings, especially in the field of optimality conditions, have been transferred later on to the periodic review case where some arguments apply

T. Wensing, *Periodic Review Inventory Systems*, Lecture Notes in Economics and Mathematical Systems 651, DOI 10.1007/978-3-642-20479-1_3,

to both review paradigms. We also briefly treat lost sales to facilitate the possible extension of our methods to this case.

Apart from this general mapping of the basic research landscape, we go into more depth on three selected papers that we consider particularly important for our study in Chap. 6. The first two of these papers stand for a particularly relevant school of thought that we will follow in the analysis later on. They include some important basic decomposition ideas that also underlie the analysis that we conduct.

In the following, we will first deal with systems that continuously review inventory levels (Sect. 3.1), and afterwards consider the case of periodic reviews (Sect. 3.2). Where the latter are concerned, we do not follow a frequently found distinction between systems that are reviewed on the basis of an exogenously determined schedule, and systems for which the review interval is also considered an optimizable variable. The basic analysis is more or less identical for both cases, so the main difference lies merely in the presence or absence of an additional optimizing principle.

Besides the fundamental form of inventory review, we will review approaches that solely consider costs separately from those that focus on determining performance criteria such as the fill rate or the customer waiting time. When treating the pure cost view, we firstly refer to studies that analyze the conditions under which the given form of the considered replenishment doctrine is optimal, and then review studies that regard the problem of optimally configuring a given type of policy.

Section 3.3 goes into more detail on the three papers mentioned above that are concerned with systems and problems similar to those that we regard in this study.

For a general overview on the genesis of inventory models, we recommend Lee and Nahmias (1993).

3.1 Continuous Review Models

If inventory levels are continuously monitored, two basic replenishment policies are possible, namely the (s, q) and (s, S) policies, as described in Sect. 2.2.3. A further special case of the (s, S) policy with $s := S - 1$ is discussed as base-stock-policy in the literature. Note that in reality we find transactions reporting rather than systems that truly review their inventory levels continuously, i.e., by registering every transaction, the system behaves as if the inventory were being continuously monitored (Hadley and Whitin 1963).

3.1.1 Pure Cost View

3.1.1.1 (s, q) Policy

Optimality Conditions. It is shown by Axsäter (2006) in Chap. 6 that the continuous review (s, q) type policy is optimal for a system with constant lead times and a compound Poisson demand process if costs for inventory holding and backordering are constant, and order sizes are a multiple of q. The proof follows an approach by Chen (2000) on a multi-echelon system with periodical demand occurrences.

Policy Configuration. The seminal paper on the (s, q) inventory policy is by Galliher et al. (1959). They present an approach for minimizing the total costs for two classes of inventory models with random demand, where the lead times are fixed in the first model and exponentially distributed in the second. A similar model is intensively discussed in Chap. 4 of Hadley and Whitin (1963). Hadley and Whitin primarily develop heuristic proceeds to determine near-optimum configurations for both the backorder and lost sales case with deterministic lead times, where fixed order costs, inventory holding costs and costs for backordering and lost sales are considered. The basic observation underlying the solution principle is that the optimum values for s and q can easily be determined if the other parameter is fixed. Thus, an iterative solution approach is proposed that repeatedly fixes one and optimizes the other parameter until they have sufficiently converged.

Hadley and Whitin furthermore develop exact formulations for the assumption of one-unit demands arriving with Poisson-distributed interarrival times and deterministic replenishment lead times. Based on these formulations, two possibilities are indicated to significantly reduce the computational effort required at the price of losing general optimality. The first simplification is to assume that replenishment orders will always be sufficient to remove all backorders, i.e., backorders will not be carried over from one replenishment cycle to the next. The higher the average availability of the system is, the lower the error due to this relaxation will be. The second proposed simplification is to approximate the lead time demand by a normal distribution. Due to its tempting computational properties, this simplification has been widely used in various inventory management approaches. However, the error may be significant if the shape of the true distribution deviates from the shape of the normal distribution in the regions of interest (Eppen and Martin 1988). For more specific approaches to approximate lead time demand distributions, see Ord and Bagchi (1983), Bagchi et al. (1984) and Dominey and Hill (2004), for example. Browne and Zipkin (1991) examine a more elaborate demand model where the amounts are driven by an exogenous stochastic process in continuous time.

Finally, Hadley and Whitin briefly discuss the implications of stochastic lead times for the system being examined, where they explicitly indicate that replenishment orders may cross over in time if lead times are variable. (Two orders are said to cross over if they arrive in a different sequence to that in which they were issued; see Sect. 5.2 for details.) For a recent study on the impact of lead time and lead time demand on the optimum configuration of (s, q) systems with very similar properties, see Song et al. (2010), for example.

It is a common observation that the computational effort required to determine the optimum configuration of an (s, q) inventory system sincerely depends on the specific characteristics of the demand and replenishment processes. Federgruen and Zheng (1992) outline conditions for the overall cost function that allow the computation of optimal configurations with very limited effort. The authors also give an example of the general algorithmic approach.

Besides optimization procedures and the heuristic mentioned above, various other heuristics have been proposed, where (besides computational effort) the ease of implementation is considered crucial for successful application. See for example Yano (1985), Zhang et al. (2001), Gallego (1998) and the referenced literature. A model with stochastic dependent lead times is examined by Heuts and de Klein (1995) for example.

All the papers mentioned above consider the full backorder case. Please refer to Kim and Park (1985), for example, for a heuristic treatment of the lost sales and limited backorder case.

3.1.1.2 (s, S) Policy

Optimality Conditions. With regard to the optimality of continuous review (s, S) policies, the first important findings are given by Scarf (1960) on a periodically reviewed ($t = 1$) inventory system with constant lead times, arbitrarily distributed period demand and a finite time horizon. Although Scarf considers periodic review, we will discuss his findings here, as they provide the grounds for the corresponding continuous review analysis. His work extends the studies of Arrow et al. (1951), Dvoretzky et al. (1953) and Karlin (1958). Scarf proves that the optimum policy is a sequence of period-dependent pairs of (s_n, S_n) when the relevant cost function consisting of fixed ordering costs (K) and holding and shortage costs $(f(x))$ satisfy the so-called K-convexity property (see Scarf 1960, p.199).

Definition 12 (K-convexity). Let $K \geq 0$, and let $f(x)$ be a differentiable function. Then $f(x)$ is K-convex if $f(a + x) - f(x) - a \cdot f'(x) \geq -K$.

The fixed order costs K introduce some sort of *tolerance* compared to strict convexity. Note that the definition implies that $f(x)$ needs to be strictly convex if the fixed order costs K are zero.

Note that we use K to denote the number of outstanding orders throughout this study *except for* this passage on the K-convexity. Unfortunately, the letter K is commonly used for both concepts in the literature, so that we consider it even more confusing to speak of the M-convexity here, for example.

For the continuous review problem with fixed lead times, and a demand process with arbitrarily distributed amounts and interarrival times that may be fully backordered, Beckmann (1962) shows that the above findings also apply for an infinite time horizon, where stationary parameterization (s, S) is optimal here. Bather (1966) gives analogous findings for a system with constant lead times and demand following a Wiener Process. Hordijk and van der Duyn Schouten (1986) model the demand process as superposition of a compound Poisson process with arbitrarily distributed amounts and a deterministic continuous process that can also be interpreted as depletion rate. A proof of optimality for a similar system is given by Bensoussan et al. (2006).

According to our knowledge, the optimality of $(S - 1, S)$ has not been examined in a dedicated study for the backorder case. However, it is easy to show that

$(S-1, S)$ is optimal for conditions under which (s, S) is optimal when the fixed costs of ordering are zero. Furthermore, it is shown by Hill (1999) that the $(S-1, S)$ policy is generally not optimal in the lost sales case even if fixed ordering costs are zero.

Policy Configuration. Archibald and Silver (1978) develop an algorithm to determine the optimal parameterization for a model that was first considered by Beckmann (1962). The main assumptions are i.i.d. demand interarrival times and i.i.d. demand volumes that may depend on the preceding interarrival time, constant lead times, fixed ordering costs and linear inventory holding and shortage costs. Archibald and Silver adapt a general approach introduced by Veinott and Wagner (1965) for the corresponding periodic review model. Unfortunately, the cost function may have several local optima for general assumptions. The basic approach is therefore to apply a full enumeration of possibly optimal pairs of s and S, where the search may be accelerated by specifying lower and upper bounds and using specific properties of the objective function to direct the search. For a discussion of these properties, see Sahin (1982) and Zheng and Federgruen (1991), for example. The latter propose an algorithm whose computational complexity to determine the optimum configuration is reported to be only 2.4 times that required to evaluate a single policy parameterization. For a more recent discussion of the approach and an improved algorithm, see Feng and Xiao (2000). According to our knowledge, an exact approach for the continuous review model with stochastic lead times does not exist. Lee (1995) reports on an exact approach for considering stochastic delays between reaching the reorder level and actually issuing the order. The original model was first examined by Weiss (1988).

Besides exact approaches, many approximations and heuristics can be found in the literature, where it is also possible to consider stochastic lead times to a certain extent. We will discuss these approaches in Sect. 3.2.1, as they are mostly formulated for the (t = 1) periodic review (s, S) policy. Nonetheless, these methods may also provide useful solutions for continuous review systems. Studies on approximations that explicitly consider the continuous model with constant lead times are given by Sahin and others (Sahin and Kilari 1984; Sahin and Sinha 1987; Sahin 1988). An approximation method for exponentially distributed lead time is given by Wijngaard and van Winkel (1979).

While the above papers consider the unlimited backorder case, Archibald (1981) examines a model with lost sales. A limited backorder model is studied by Krishnamoorthy and Islam (2004). Instead of considering an estimate of the customer demand process, Iyer and Schrage (1992) determine the parameterization that would have been optimal for a historical demand data stream.

The $(S-1, S)$ policy with i.i.d. Poisson demand arrivals, arbitrary i.i.d. lead times and lost sales is analyzed by Smith (1977). The author gives insights into optimum inventory levels and develops a heuristic approach for determining near-optimum solutions.

Furthermore, Feeney and Sherbrooke (1966) examine a continuous review base-stock policy with a compound Poisson demand process and constant lead times.

Song et al. (2010) examines the effect of lead time variability on the optimum reorder point for a base-stock policy.

3.1.2 Performance View

3.1.2.1 (s, q) Policy

We have already discussed the difficulties of monetarily quantifying the costs of a stockout in Sect. 2.3.1.3. The common alternative to purely counting costs against costs is to minimize the more traceable costs of ordering and holding inventory subject to a certain service criterion. In order to do this, the primary requirement to analytical approaches is the capability to evaluate a parameterized system according to the performance criterion considered.

Where this problem is concerned, Hadley and Whitin propose an appraoch based on Lagrangean relaxation to minimize the average costs of ordering and inventory holding subject to a ready rate constraint (Hadley and Whitin 1963, Sect. 4–16). Both parameters are simultaneously determined by an iterative procedure. The underlying inventory system is the same as described above for the pure cost perspective.

The problem of configuring an (s, q) policy subject to a fill rate criterion is addressed by Yano (1985). The author gives insights on properties of the optimum solution and develops a heuristic approach. A practice-driven approach to configure the reorder level of an (s,q) policy subject to a minimum fill rate can be found in van der Veen (1981). Approximate results are given for normal and Gamma-distributed lead time demand. In van der Veen (1984), related findings are given for more general assumptions on lead time demand distribution. The expected duration of a stockout is examined as well in that paper. In van der Veen (1986), the ready rate (per replenishment cycle) is considered instead of the fill rate. Alternatively to considering the fill rate, Boyaci and Gallego (2001) describe an approach to minimize the total holding and ordering costs subject to a limit on the expected number of backorders. In a more recent paper, Agrawal and Seshadrin (2000) discuss general limits for the fill rate constrained problem as well as convexity conditions for the corresponding Lagrangian relaxation.

Eppen and Martin (1988) point out that approximating the lead time demand by a normal distribution can lead to severe errors in the estimate of safety stock requirements. They emphasize that modeling lead time demand as mixed distribution allows for a precise safety stock adjustment in discrete-time (s, q) systems if q is preselected.

Akinniyi and Silver (1981) examine the problem of determining the reorder point subject to a maximum expected stockout duration for the case of constant lead time and normal or Poisson demand. They point out that the customer waiting times is probably an even more interesting metric in the case of backordered demands. Boyaci and Guillermo (2002) address the problem of finding the optimum (s, q)

configuration subject to an upper bound on the expected and maximum waiting time per volume for those demand units that are actually backordered. The authors consider the class of logconcave continuous lead time demand distributions. The approach could easily be extended to all (backordered and immediately delivered) demand units.

Heuristic appraoches for characterizing customer waiting times in continuous review (s, q) systems are given by van Beek (1981) as well as Svoronos and Zipkin (1988), both in a multi-echelon environment.

3.1.2.2 (s, S) Policy

On behalf of the continuous review (s, S) model, Sahin (1982) gives a comprehensive analysis of elementary characteristics, such as inventory position, on-hand inventory and performance metrics. Expressions are obtained for generally i.i.d. demand interarrival times, generally i.i.d. demand volumes and constant lead times. A study on configuring continuous review (s, S) systems subject to a fill rate and two ready rate criterions is given by Tijms and Groenevelt (1984). Chen and Krass (2001) present an adaptation of the cost-based optimization procedure of Zheng and Federgruen (1991) to minimize costs subject to the three types of service criteria discussed in Sects. 2.3.2.1–2.3.2.3. More recently, Larsen and Thorstenson (2008) compare the classical volume fill rate (as defined in Sect. 2.3.2.2) and the fraction of orders that are delivered in time as performance indicators for an $(S - 1, S)$ base-stock inventory policy. The ready rate and fill rate are also considered in the paper by Feeney and Sherbrooke (1966) that we have already mentioned above.

A special performance metric, customer waiting times, has been intensively studied relating to continuous review (s, S) policies. In fact, the first studies on customer waiting times in inventory systems consider this type of replenishment doctrine. The corresponding papers are therefore also important for the analysis of other types of systems as they introduce the general proceed. To begin with the first paper, Higa et al. (1975) develop approximate results for the distribution of waiting times per customer order batch in an $(S - 1, S)$ system with compound Poisson demand arrivals and negative exponential lead times. For the corresponding system with constant lead times, Sherbrooke (1975) derives exact formulations. Finally, Kruse (1980) gives an exact formulation for the continuous review $(S-1, S)$ system with compound Poisson demands and arbitrarily distributed lead times. This study also provides a numerical comparison of the exact results with the approximation by Higa et al. Kruse (1981) later considers the waiting time per demand unit for general (s, S) systems with arbitrarily distributed demand interarrival times and volumes and constant lead times. This study is based on distribution of the inventory position, and this makes it methodically comparable with the approaches we will develop in Chap. 6. (See also Lee and Nahmias 1993 on the genesis of the approaches mentioned.) Das (1977) considers customer waiting times for an $(S-1, S)$ inventory system against the backdrop of customer demands that may only be backordered for a certain limited time.

For a multi-item production system, Federgruen and Katalan (1994) give approximations for the customer waiting time if all items are replenished from the same capacity-restricted source according to a base-stock policy.

Continuous review inventory models with perishable products are discussed in Weiss (1980), Berk and Gürler (2008) and the literature that is referenced in these two papers. Liu (1990) and Kalpakam and Sapna (1994), for example, focus particularly on the (s, S) policy.

3.2 Periodic Review Models

The general alternative to continuously monitoring inventory levels is the use of a regular review interval. This interval can either be externally determined or itself be a parameter of the replenishment policy, which may be changed. According to our classification scheme (see Sect. 2.2.3), the three relevant policies are (t, s, q), (t, s, S), and (r, S), where we assume that all given parameters may be adjusted. In the event that parameter t of the first two policies is fixed to the underlying base period $(t = 1)$, the corresponding systems are commonly denoted as *periodic review* (s, q) and (s, S) systems in the literature. Following our classification, these policies would be equivalent to an $(t = 1, s, q)$ and $(t = 1, s, S)$ doctrine.

Note that the works presented in this section are selected according to their general value for understanding and analyzing periodic review systems. We therefore did not consider it helpful to treat models with a fixed review interval separately from those that may freely choose t.

As in Sect. 3.1, we will first distinguish between cost and performance view.

3.2.1 Pure Cost View

3.2.1.1 (t, s, S) Policy

Optimality Conditions. It is proven in several papers that the (s, S) replenishment policy is the optimum doctrine to minimize costs if the inventory is reviewed every period $(t = 1)$, and certain conditions apply to the cost function. The individual proofs differ mainly in terms of the planning horizon considered and the assumptions concerning the cost function. The first relevant paper on the periodic review system is found in Scarf (1960). It was already mentioned in Sect. 3.1.1 as it is seminal to both the analysis of the continuous review and periodic review case. Based on Scarf's work, Iglehart (1963) proves that the analogous findings apply for the infinite time case, where the optimum policy is then of a stationary (s, S) type. Please also refer Veinott and Wagner (1965) for a summary of the proof. The general cost structure is the same as for the continuous review model, i.e., we observe fixed costs for issuing an order and additional costs for inventory holding and shortfall compensation that are proportional to the corresponding amounts. Some models also consider variable ordering costs (apart from the fixed costs for ordering), which does not greatly complicate the analysis. Veinott (1966) as well as Johnson (1968) extend these results to more general properties of the cost function. More compact proofs that directly focus on the finite time horizon are provided by Zheng (1991) and Benkherouf (2008).

An interesting extension to the demand process is examined by Sethi and Cheng (1997). The authors indicate properties of the cost function that result in the optimality of the (s, S) doctrine if the demand in successive periods depends on the previous occurrence, and may be modeled as a Markov chain.

For further discussion on the optimality of (s, S) type policies with a finite planning horizon, see Porteus (1971), Porteus (1972) and Schäl (1976). Similar proofs for multi-item systems are elaborated by Johnson (1967) and Kalin (1980).

Optimality conditions for (s, S) systems with stochastic lead times can be found in Kaplan (1970) for a finite planning horizon and in Ehrhardt (1984) for an infinite planning horizon.

An important proof for the periodic review $(S-1, S)$ base-stock policy is given by Clark and Scarf (1960). They show that the base-stock policy is optimal even in a multi-echelon environment, when holding and shortage costs are convex and there is no capacity limit on order quantities, which in this model is equivalent to zero fixed ordering costs. Considering that $(t = 1, s, S)$ is optimal in the presence of fixed order costs, the result is rather intuitive: clearly there is no reason to wait with an order if the ordering costs for two split orders are the same as for one combined order. While Clark and Scarf consider a finite time horizon, Federgruen and Zipkin (1984a) show that their results also apply to the infinite time horizon. Proofs of optimality for the base-stock policy type when demands are non-stationary are given by Karlin (1960) and Morton (1978).

In all of the above studies, full backordering is assumed. To our knowledge, no proof of optimality conditions for the lost sales case exists.

Policy Configuration. For the problem of finding optimal (s, S) configurations, the genesis of approaches is quite similar to the continuous review model. In fact, the general algorithmic approach can basically be applied to both problem types, where the differences emerge in evaluating single solutions and using specific properties to direct the search. As mentioned above, Veinott and Wagner (1965) introduce a basic algorithmic approach for that problem. For the periodic review model, improvements within the general framework are proposed by Johnson (1968), Bell (1970) and Archibald and Silver (1978), for example. For a different approach based on Markov decision models, see Küenle and Küenle (1977) and Federgruen and Zipkin (1984b). Even for the periodic review model, the algorithms of Zheng and Federgruen (1991) and Feng and Xiao (2000), respectively, claim to find optimal (s, S) configurations most quickly.

An approach for determining optimum solutions for systems with stochastic lead times has been developed by Ehrhardt (1984) for the infinite planning horizon.

Besides the optimal approach, numerous approximate methods are proposed. Overviews on the literature are given by Porteus (1985) along with an extensive numerical comparison of various methods and Sani and Kingsman (1997).

A first straightforward approach in that context is using the approximation of Hadley and Whitin for the continuous review (s, q) system and set $S := s + q$, where the deviation from the optimum solution may be rather large. See Porteus (1985) for a further refinement of this method.

Several studies explore different approaches to simplifying the relevant cost functions by replacing those parts that are difficult to evaluate by simpler approximations. A prominent method of this category is the approximation of the lead time demand by normal distribution that has already been mentioned. An example of this strategy is given by Freeland and Porteus (1980), who elaborate an idea due to Norman and White (1968). They first consider the deterministic problem based on expected values that is easy to solve. This initial solution is then improved based on the adjusted cost function. Further possible simplifications to the resulting cost function that may still be difficult to handle are given by Porteus (1985). Related approaches are proposed by Wagner et al. (1965) and Naddor (1975).

Ehrhardt (1979) derives a closed approximation formula by regression analysis on the basis of 288 different instances of (s, S) configuration problems with fixed lead times and period demand following Poisson and negative binomial distributions. See also Ehrhardt and Mosier (1984).

For the stochastic lead time case, we again refer to Ehrhardt (1984), who also presents an approximate method.

3.2.1.2 (t, s, q) Policy

The (s, S) policy may only be applied if it is possible to order arbitrary amounts of the considered goods. In the event that the orders must correspond to certain batch sizes, the periodic review (s, q) policy is the appropriate decision rule. Note that in the periodic review context, the (s, q) policy is to be understood as $(t, s, n \cdot q)$, where the replenishment rule is as follows: every period t, review the inventory position and order the smallest multiple of q that will raise the inventory position above s.
Optimality Conditions. In this context, Veinott (1965) indicates conditions under which (s, q) is the optimum type of replenishment rule. Proof is given for linear ordering costs and unimodal (i.e., convex) costs for inventory holding and backordering. In the case of fixed order costs, the (s, q) type is not optimal for general conditions.
Policy Configuration. Zheng and Chen (1992) propose an algorithm to find the values for s and q that minimize the average costs of ordering, inventory holding and backordering. They examine a system with i.i.d. demand per period, constant lead times, fixed order costs and unimodular holding and backordering costs. The analysis is based on findings given by Hadley and Whitin (1962) and Hadley and Whitin (1963), Sect. 5–3 on distribution of inventory position in systems of the corresponding type. The algorithm sequentially optimizes first s and then q. The authors suggest that the approach may easily be adjusted to stochastic lead times

if these follow the model proposed by Zipkin (1986b). (See also Sect. 5.3 of this study.) More recently, Larson and Kiesmüller (2007) have obtained a closed-form cost expression when lead times are constant and the demand process is compound generalized Erlang.

Early heuristic approaches to the determination of s and q are given by Hadley and Whitin (1962), Hadley and Whitin (1963), Wagner et al. (1965) and Naddor (1975), where all these approaches aim for the (globally) optimum solution, but cannot guarantee not to end up with a local optimum. Approximation formulae for the case of Gamma-distributed lead time demand are derived by Das (1976).

While all papers above assume full backordering, Johansen and Thorstenson (1993) and Johansen and Thorstenson (1996) consider the lost sales case. Heuristic and optimum solutions are given, where the latter depend on the assumption that only one replenishment order is outstanding. See also Johansen and Hill (2000) for a similar study. The case of time-weighted backorders (backorder costs that depend on amount and duration) is examined by Das (1983). Kim and Park (1985) study the mixed case of lost demand and time-weighted backorders.

3.2.1.3 (r, S) Policy

The (r, S) policy may be regarded as a special case of both the (t, s, S) and (t, s, q) policy. The (t, s, S) policy is equivalent to (r, S) if we set $t = r$ and $s = S$. The $(t, s, n \cdot q)$ policy behaves like (r, S) if $t = r$, $s = S$ and $q = d_{min}$, where d_{min} is the minimum positive customer order amount that the system may observe. It is probably due to this observation that the literature on the (r, S) doctrine is relatively sparse. It is reasonable for researchers to try to examine a system under the most general assumptions that still allow for the implications considered.

Optimality Conditions. Being only a special case of (t, s, S), the (r, S) policy type is not optimal for the same general conditions as discussed above. This is mainly due to the fact that the latter policy clearly has less potential to issue orders of similar sizes. While orders are at least of size $S - s$ for the (t, s, S) type, they may be arbitrarily small for the (r, S) type. In the absence of fixed order costs, the (r, S) type may well be optimal due to being the equivalent to the periodic review base-stock policy. Then clearly $r = 1$ will always hold for optimal parameterizations. An explicit proof of the optimality of the (r, S) policy type when the review interval is fixed and no fixed order costs occur is given by Chiang (2006), see also Chiang (2007).

In this context, Rao (2003) discusses theoretical properties of the periodic review (r, S) policy with constant lead times and a continuous stochastic demand process. Specifically, he examines a Brownian motion model and a compound Poisson distribution to generate demands. Rao shows for this model that the costs of the optimum (r, S) policy may be around a maximum of 42% higher than for the corresponding optimum (s, q) policy. Furthermore, conditions for the joint convexity of the cost function in r and S are given that allow a fast enumeration of optimum policy configurations.

Optimality conditions for the periodic review $(S-1, S)$ base-stock model (which is equivalent to the $(r = 1, S)$ model) are given by Feng et al. (2006a,b), where the authors consider the possibility of choosing from multiple delivery modes for each replenishment order.

Optimality conditions. Considering the optimum policy configuration, Hadley and Whitin (1963) point out that it is easy to find if lead times are constant and unit demands arrive on a continuous time axis according to a Poisson distribution. If that is the case, the (r, S) policy is a special case of the $(r, s, n \cdot q)$ policy with $s = S - 1$ and $q = 1$, and the problem may be solved with the same methods that are at hand for the latter policy.

Besides the optimization approach, Hadley and Whitin also describe an approximation that has been rather influential on later inventory theory. The authors consider Poisson-distributed demand unit arrivals and arbitrarily distributed lead times, where the probability that orders cross over, i.e., that they arrive in a different sequence than they were issued, is assumed to be negligible. The approximation is based on the observation that it is much easier to calculate the mean net inventory than the physical inventory. Thus, the mean net inventory is used to approximate the physical inventory, which is obviously the better the lower the stock-out probability is. Hadley and Whitin also develop the corresponding approach for the lost sales case.

Due to the regular replenishment interval, the (r, S) policy is well suited for coordinated joint replenishment in multi-item inventory systems. A problem of this type is examined by Fung et al. (2001), for example, for a system with fixed item-specific lead times and a compound Poisson demand process. As a continuous review counterpart, the *can order* (s, c, S) policy is discussed in the literature, which means that an order is placed whenever the inventory position of any item drops below the specific s. The order then raises the inventory positions of all those items up to the corresponding S whose inventory position is below c. A comparison of both policies is given by Atkins and Iyogun (1988), for example. The authors assume constant lead times and Poisson demand-unit arrivals.

3.2.2 *Performance View*

3.2.2.1 (t, s, q) Policy

The literature that directly considers the performance issues of the (t, s, q) policy is rather sparse, which may be due to the substantial analysis that was already done early on the influential book of Hadley and Whitin (1963). Although not developed directly, the expressions for the fill rate and time-weighted fill rate may quite easily be derived from their analysis of the mean backorder amount, for example. Nonetheless, some further work on this inventory model should be mentioned.

Janssen et al. (1998) approximate the optimum (t, s, q) configuration subject to a fill rate constraint. They consider a system with dependent stochastic lead times (no crossover) and demand following a compound Bernoulli process, i.e., demand in a period is either zero with a certain probability, or follows an arbitrary continuous distribution. Their work extends an earlier practice-driven study of Dunsmuir and Snyder (1989).

Fischer (2008) gives approximations for the mean inventory, fill rate and time-weighted fill rate for the case of discrete distributed lead times and periodical demand that may be normally or Gamma-distributed.

The customer waiting time per order is approximated by Kiesmüller and de Kok (2006) for a (t, s, q) system with discrete distributed lead times and compound Poisson demand occurrences. Please also refer to Tempelmeier (1985) for a related approach with periodical customer order arrivals. The customer order waiting time per part is approximated by Tempelmeier and Fischer (2010) for a (t, s, q) system with discretely distributed demand and normally or Gamma-distributed demand per period.

3.2.2.2 (t, s, S) Policy

Basic insights on the performance analysis of (t, s, S) policies are given by Tijms (1994) for the backorder and lost sales case. The author gives approximations for the fill rate under general assumptions and outlines a simple proceed to parameterize an (t, s, S) policy subject to a fill rate criterion. More detailed studies on similar problems are given by Schneider (1978), Schneider (1981) and Schneider and Ringuest (1990) for the case of constant lead times, whereas stochastic lead times are considered by Tijms and Groenevelt (1984) and Bashyam and Fu (1998).

Exact formulations to determine the fill rate when demand is Gamma-distributed and lead times are constant are given by Moors and Strijbosch (2002).

Recently, Silver et al. (2009) consider a (t, s, S) system with arbitrary independent demand and deterministic lead times. They develop an approach to approximate the optimum (t, s, S) configuration subject to a fill rate criterion and the average time between two consecutive replenishment orders.

Multi-item problems with service constraints are addressed by Cohen et al. (1989) and Cohen et al. (1992), for example.

3.2.2.3 (r, S) Policy

A basic performance analysis of the (r, S) policy is given in Zipkin (2000) for Poisson demand arrivals and constant lead times. The author analyzes the mean physical inventory levels, the ready rate per order cycle and the mean backorder amount that directly implies the fill rate. Also see van der Heijden and de Kok (1998) on mean physical inventory levels and on fill rate. Cardós et al. (2006) develop an exact approach to determine the ready rate per replenishment cycle for a system with constant lead times and arbitrarily distributed lead time demand. Exact formulations for the fill rate are given by Zhang and Zhang (2007) for constant lead times and generally distributed demand per period, where closed expressions are given for normally distributed period demand. Also see Teunter (2009). A similar study on the fill rate as well as a discussion of commonly used approximations can be found in Johnson et al. (1995). A further brief general discussion of the (r, S) policy is given in Silver et al. (1998).

Two similar studies on computing the customer waiting times per order are conducted by van der Heijden and de Kok (1992) and Chen and Zheng (1992) for continuously distributed stochastic lead times and compound Poisson demand arrivals, where the latter use slightly more general assumptions. In both papers, exact non-closed expressions are derived that may be enumerated to any desired accuracy. See Sect. 3.3 for details. Tempelmeier (2000) proposes a heuristic approach for discretely distributed lead times and normal or Gamma-distributed demand per period that is reported to work well for high service levels. This paper is also briefly addressed in Sect. 3.3. A similar study for a multi-item system with fixed lead times is conducted by Hausman et al. (1998).

If customer orders arrive on a periodical basis, the customer waiting time per unit can easily be computed based on development of the average backorder amount from one period to the other. (See Fischer 2008 for example.)

Finally, we would like to mention three papers that discuss special issues related to modeling demand using the example of an (r, s) policy. Strijbosch and Moors (2006) question the widespread approach of using the normal distribution to model demand even if a significant probability mass lies on negative values. Alternatively they propose to either concentrate the mass of negative values on zero or to use the corresponding truncated distribution for values greater than or equal to zero. Strijbosch and Moors (2005) also discuss the problem that demand parameters are usually unknown or uncertain in practical applications. Charnes et al. (1995) drop the assumption of i.i.d. demands and suggest an approach to determine safety stock levels in an (r, S) policy when period demand is serially correlated.

3.3 Selected Studies

In this section, we go into more depth on three papers that analyze the customer waiting time per order for (r, S) systems, each under slightly different assumptions. We pay special attention to this metric here as we consider the corresponding analysis to have the most differentiated structure of those approaches that we will describe in Chap. 6. We select the papers, as these three are the only ones to our knowledge that regard the customer waiting time per order in the context of a single-level, single-item (r, S) policy. The two first studies presented in Sects. 3.3.1 and 3.3.2 are particularly relevant for our analysis in Sect. 6 for two reasons. First, they are closely related to the analysis that we conduct in Sect. 6.3.1.2 and provide valuable insights into the underlying problem structure of determining waiting times. Second, we consider these two studies to be generally very useful for understanding basic concepts of analyzing periodic review inventory systems that go beyond the *classic* analysis of inventory and backorder amounts. The third study is interesting because it more or less considers the same system that we examine in Sect. 6.3.1. However, it uses a rather different approximate approach, while our approach is exact.

3.3.1 *Van der Heijden and De Kok (1992)*

Customer waiting times per order are analyzed by van der Heijden and de Kok (1992) for a periodic review order-up-to (r, S) inventory system with customer orders arriving at the system according to a compound renewal process. Demand per customer order is arbitrary i.i.d. and will be backlogged if stock on hand is insufficient. Replenishment orders take arbitrarily continuous distributed lead times and arrive in the same sequence as they were issued (i.e., order crossover is ruled out). A lead time process that satisfies the assumption is not specified, but it is noted that lead times are not necessarily independent. Customer waiting times are analyzed for complete order fulfilment, i.e., the time between the arrival of an order and the delivery of the last outstanding part.

Compared with the approach of Chen and Zheng (1992) that is described in the following section, the model framework can be regarded as a specialization. While the former regard an arbitrary S, van der Heijden and de Kok assume $S > 0$. Furthermore, Chen and Zheng give a more detailed specification of the lead time generating process. A comparison of the model frameworks is given at the end of Sect. 3.3.2.

Developing their analytical model, van der Heijden and de Kok imagine an (r, S) inventory system starting with an initial net inventory level of S and no outstanding orders. Clearly, the inventory position then also equals S, and there is nothing on backorder. From this initial state, the authors consider the probability of an order arriving at time t within the m-th order cycle according to (3.1), where $Q(t + m \cdot r + w)$ is the number of replenishment orders that have arrived in $[0, t + m \cdot r + w]$, $D(t + (m - i) \cdot r)$ is the demand in $[t + (m - i) \cdot r, t + m \cdot r]$, and D is the demand of the customer order arriving in t:

$$
\begin{aligned}
&P\{W_{t+m\cdot r} > w\} \\
&\quad = \sum_{i=0}^{m} P\{Q(t + m \cdot r + w) = i\} \cdot P\{D(t + (m - i) \cdot r) + D > S\} \qquad (3.1)
\end{aligned}
$$

To understand (3.1), let us first consider the probability $P\{W_{t+m\cdot r} > 0\}$. Regarding a specific amount of orders i, the i-th element of the sum reads as the probability that the i-th replenishment order was the last one that arrived, while the demand since its release date exceeds S. For lead times $w > 0$, the time span being observed for replenishment orders to arrive is simply extended. While the customer order is waiting, it becomes more and more likely that a sufficient replenishment order will arrive to finally cover the demand.

Using the property that orders cannot cross over in time (i.e., if the i-th order has arrived, we know that all $i - 1$ preceding orders have arrived as well) $P\{Q(t + m \cdot r) \geq i\}$ can be derived from the cumulative density function of the lead time L (3.2–3.5):

$$P\{Q(x) \geq i\} = P\{i \cdot r + L_i \leq x\} = \quad P(L_i \leq x - i \cdot r) \quad \forall i \geq 1 \tag{3.2}$$

$$P\{Q(x) \geq 0\} = 1 \tag{3.3}$$

$$P\{Q(x) = i\} = P\{Q(x) \geq i\} - P\{Q(x) \geq i + 1\} \quad \forall i \geq 0 \tag{3.4}$$

$$P\{Q(x) = 0\} = 1 - P\{Q(x) \geq 1\} \tag{3.5}$$

Before we continue with the subsequent steps, please note that the following expositions deviate from the formulations in the original paper for the sake of improving comparability with other approaches. However, only plain mathematics was used, and no further ideas were added.

Inserting (3.4) into (3.1), we derive (3.6) by changing the order of summation (i:=m-i). Without loss of generality, let us assume $P\{Q(t + m \cdot r) = 0\} = 0$, which saves us the need to specially consider the corresponding case (3.5). Note that $Q(x)$ is strictly greater than 0 as soon as the first replenishment order arrives, so we disregard just a finite runtime span of the examined system, which is irrelevant considering the limit $m \to \infty$ later on.

$$\begin{aligned} P\{W_{t+m\cdot r} > w\} &= \sum_{i=0}^{m} \Big(P\{L_i \leq t + (m-i) \cdot r + w\} \\ &- P\{L_i \leq t + (m-i-1) \cdot r + w\} \Big) \cdot P\{D(t + (m-i) \cdot r) + D > S\} \\ &= \sum_{i=0}^{m} \Big(P\{L_i \leq t + i \cdot r + w\} - P\{L_i \leq t + (i-1) \cdot r + w\} \Big) \\ &\cdot P\{D(t + i \cdot r) + D > S\} \end{aligned} \tag{3.6}$$

From the probability that a customer order arriving at time $t + m \cdot r$ will undergo a delay of w or more time units (3.6), the corresponding probability of a customer order arriving at any time in the m-th order cycle can be derived using the property that order arrival probabilities are uniformly distributed within an order cycle. (See Sect. 3.3.2 for details concerning this property.) This implies (3.7):

$$P\{W_m > w\} = \frac{1}{r} \cdot \int_0^r P\{W_{t+m\cdot r} > w\}\, dt \tag{3.7}$$

Let $P\{W_m > w\} := \frac{1}{r} \cdot \int_0^r f(i \cdot r + t)$ and consider Beppo Levi's Theorem ($\int_0^\infty \sum_{i=1}^\infty f_i(t)\, dt = \sum_{i=1}^\infty \int_0^\infty f_i(t)\, dt$, if $f_n(t) \geq 0 \ \forall n, n \in \mathbb{Z}$). Then (3.8) helps us to derive a shorter formulation:

$$\begin{aligned} &\int_0^r \sum_{i=0}^{m} f(i \cdot r + t)\, dt \\ &\quad = \int_0^r f(t)\, dt + \int_0^r f(r + t)\, dt + \cdots + \int_0^r f(m \cdot r + t)\, dt \end{aligned}$$

$$= \int_0^r f(t)\,dt + \int_r^{2\cdot r} f(t)\,dt + \cdots + \int_{(m-1)r}^{m\cdot r} f(t)\,dt$$

$$= \int_0^m f(t)\,dt \tag{3.8}$$

Thus, (3.9) describes the overall probability of a customer order observing a waiting time of w or longer:

$$P\{W > w\} = \lim_{m\to\infty} P\{W_m > w\}$$

$$= \frac{1}{r} \cdot \int_0^\infty [P\{L \le t + w\} - P\{L \le t - r + w\}] \cdot P\{D(t) + D > S\}$$

$$(=) \quad \frac{1}{r} \cdot \int_0^\infty P\{t - r + w \le L \le t + w\} \cdot P\{D(t) + D > S\} \tag{3.9}$$

Evaluating (3.9) with specific distribution assumption may, however, be a difficult task. In the first place, the distribution of $D(t) + D$ for an arbitrary $t > 0$ cannot generally be derived in closed form. Even if that succeeds, the overall integral over t remains to be solved.

Van der Heijden and de Kok develop (3.9) for Gamma-distributed demand, exponentially distributed interarrival times and arbitrarily distributed lead times. It is well known that the number of arrivals in T time units follows a Poisson distribution, with $\mu = \frac{T}{\lambda}$ if the interarrival times are exponentially distributed with $\mu = \lambda$. Thus, $P\{D(t) + D > x\}$ can be calculated according to (3.10), where $D_i(x)$ denotes the i-time convolution of D and λ^* is the interarrival time of customer orders:

$$P\{D(t) + D > x\} = \sum_{i=0}^{\infty} \frac{(\lambda \cdot t)^i}{i!} \cdot e^{\lambda \cdot t} \cdot (1 - CDF_{D_{i+1}}(x)) \quad \forall x \ge 0 \tag{3.10}$$

$$\lambda := \frac{1}{\lambda^*}$$

Inserting (3.10) into (3.9), we derive (3.11), applying Beppo Levi's Theorem:

$$P\{W > w\} = \int_0^\infty [P\{L \le t + w\} - P\{L \le t - r + w\}]$$
$$\cdot \sum_{i=0}^{\infty} \frac{(\lambda \cdot t)^i}{i!} \cdot e^{\lambda \cdot t} \cdot (1 - CDF_{D_{i+1}}(S))\,dt$$

$$= \frac{1}{r} \cdot \Bigg[\int_0^\infty [P\{L \leq t + w\} - P\{L \leq t - r + w\}]\, dt$$
$$- \sum_{i=0}^{\infty} \frac{CDF_{D_{i+1}}(S)}{i!} \cdot \int_0^\infty [P\{L \leq t - r + w\}$$
$$- P\{L \leq t + w\}] \cdot (\lambda \cdot t)^i \cdot e^{-\lambda \cdot t}\, dt \Bigg] \tag{3.11}$$

On the basis of (3.11), van der Heijden and de Kok give ready-to-compute expressions for deterministic, hyperexponential and Erlang distributed lead times, using integral approximation methods when closed forms are unknown. The authors report accurate results for example problem instances using both deterministic and Erlang distributed lead times.

3.3.2 Chen and Zheng (1992)

Chen and Zheng (1992) proceed from a system framework very similar to that of van der Heijden an de Kok. Customer orders arrive according to a compound renewal process, with continuously i.i.d. interarrival times and demand sizes. By contrast, the demand sizes may depend on the corresponding interarrival time, i.e., the distance versus the preceding customer demand arrival order. Furthermore, S may be chosen arbitrarily, while van der Heijden an de Kok assume $S > 0$. Finally, Chen and Zheng are more specific on their replenishment lead times, which are said to emerge from a random process according to the model introduced by Zipkin (1986b), see Sect. 5.3.

Nonetheless, the systems considered are fairly similar, and it is interesting to see that the two approaches start with different views on the problem, but conclude with basic formulas that are more or less equivalent.

At first, Chen and Zheng subdivide the problem into analysis of the arrival process of customer orders within a review cycle and analysis of the probability that a customer order will arrive at some certain time within a review cycle, observing a delay of more than x periods. The further analysis is based on the following observations:

(1) The arrival time of a customer order within the review interval $t \in [0, r]$ is uniformly distributed ($p(t) = \frac{1}{r}$). To understand this property, imagine a sample path of customer order interarrival times (F_n) wrapped around a circle of constant length r. Unless the interarrival times describe fixed cyclic patterns, there will be no accumulation at certain prominent points of the time scale.
(2) The inventory position is S at the beginning of each period. This is the only deterministic assertion we can make concerning inventory levels.
(3) A customer order arriving at time t observes a waiting time $w > x$ if the backorder in $t + x$, i.e., x time units ahead, is higher than the demand between

t and $t + x$.
I.e., $P(W_t > x) = P\{IL(t + x) < -D(t, t + x)\}$

Proceeding from (2) and (3), the authors derive (3.12) as the probability that a customer order will arrive at time t of a review cycle to observe a delay of w time units or higher, where they use the convention that $D(a, b) = -D(b, a)$ if $b \leq a$:

$$P\{W_t > w\} = \sum_{k:kT \leq t+w} P\{S < D(kT, t)\} \cdot P\{t + w - (k + 1)T \leq L_t \leq t + w - kT\} \tag{3.12}$$

According to (3.12), $P\{W_t > w\}$ is developed as the sum from $k = -\infty$ to $\lceil \frac{T}{t+x} \rceil$, enumerating the corresponding demand occurrence and lead time probabilities. For the demand side, the possible cumulative occurrences since the order was issued that will arrive in $t + w$ or later are considered. For the lead time side, the probabilities are accounted for that the lead time falls within the corresponding interval of time, the length of which induces the demand considered. Note that w sets a boundary to the right ($t + w$), while the maximum lead time effectively sets a boundary to the left. If a maximum lead time l_{max} could be specified, the sum could be focused on $k = \lfloor \frac{T}{t+x-l_{max}} \rfloor$ to $\lceil \frac{T}{t+x} \rceil$.

Combined with (3.3.2), the overall probability of a lead time of w or higher is described by (3.13):

$$P\{W > w\} = \frac{1}{T} \int_0^T P(W_t > w)\, dt \tag{3.13}$$

However, it is rather unlikely that closed expressions for both (3.12) and (3.13) will be derived for specific distribution assumptions. Two main difficulties arise here: Considering the demand in $kT - t$ time units, the convolution of N random variables has to be determined, where N is itself a discretely distributed random variable with integer states. Even if this succeeds, determining the integral of (3.12) remains a challenging task.

The authors specify (3.12) and (3.13) for a system with constant lead times and compound Poisson demands, where the interarrival times are i.i.d. exponential random variables and independent of the demand sizes. Demand sizes are arbitrarily i.i.d. For this system, an infinite sum is derived with the summands converging to zero. The resulting formula is ready to compute with an accuracy-controlled breakup criterion. Numerical results are not given.

Let us finally have a closer look at (3.12) and (3.13) to understand that they indeed cover (3.11) of van der Heijden and de Kok. Actually, there are two differences between the approaches of van der Heijden and de Kok and Chen and Zheng that are of relevance here. Firstly, van der Heijden and De Kok separately display the demand portion D that arrives at the specific time $t + m \cdot r$, and secondly, they assume $S > 0$, while Chen and Zheng regard an arbitrary S.

Assuming $S > 0$, obviously $P\{S < D(kT, t)\}$ equals zero for all $k > 0$. Thus, we can reformulate (3.12) according to (3.14) by substituting $D(kT, t) := D(t - kT)$ and setting $k := -k$. Note that for $m \to \infty$, the resulting formulation is already equal to (3.7):

$$\begin{aligned} &P\{W_t > w\} \\ &\quad = \sum_{k=-\infty}^{0} P\{S < D(t - kT)\} \cdot P\{t + w - (k+1)T \leq L_t \leq t + w - kT\} \\ &\quad = \sum_{k=0}^{\infty} P\{S < D(t + kT)\} \cdot P\{t + w + (k-1)T \leq L_t \leq t + w + kT\} \end{aligned} \tag{3.14}$$

Inserting (3.14) into (3.13), we obtain (3.15), which is equal to (3.11) if we remember that van der Heijden and de Kok separately model the demand arriving in t, while Chen and Zheng incorporate it into the demand that has arrived between a certain point in time and t:

$$\begin{aligned} P\{W > w\} = \frac{1}{T} \int_0^T (&\sum_{k=0}^{\infty} P\{S < D(t + kT)\} \\ &\cdot P\{t + w + (k-1)T \leq L_t \leq t + w + kT\})\, dt \\ = \frac{1}{T} \int_0^{\infty} &P\{S < D(t)\} \cdot P\{t + w - T \leq L_t \leq t + w\})\, dt \end{aligned} \tag{3.15}$$

3.3.3 Tempelmeier (2000)

Tempelmeier (2000) analyzes customer waiting times in a periodic review order-up-to (r, S) inventory system, with customer demand arriving according to a compound renewal process with fixed interarrival rate and continuously i.i.d. demand amounts. The replenishment lead time is deterministic, i.e., order crossover cannot occur.

The author basically uses the discrete distribution of the demand coverage to approximate the discrete probability distribution of the customer waiting time for both the full and part delivery case. The main idea is to derive the expected number of orders $H(w) = r \cdot P\{W = w\}$ that observe a delay of $w = 0, 1, \ldots, w_{max}$ periods using the distribution of demand coverage. Knowing $H(w)$, the distribution of waiting times can obviously be calculated according to (3.16):

$$P\{W = w\} = \frac{H(w)}{r} \quad \forall w \tag{3.16}$$

The model was implemented and tested with normal and Gamma-distributed customer demand. A simulation study revealed accurate results for instances with low stockout probabilities, even for high demand variation. With lower S-levels, the accuracy tends to decline.

Chapter 4
Basic Methods

In Chap. 2 we stated how we will be talking about inventory systems, what structure and elements may be considered, and which performance metrics may be used for an evaluation. Following on that, we gave an overview of single-level static stochastic inventory systems and problem aspects that have already been studied scientifically (Chap. 3). We were primarily trying to give an introduction to the general way of thinking about managing inventories that underlies this present study. We also indicated studies that exhibit similar problem scopes to those we are concerned with, discussing two closely related papers in greater depth.

In this chapter we will introduce some basic methods and approaches that we consider helpful for evaluating and optimizing periodic review inventory systems. We will address the following issues. Section 4.1 describes a simple approach to approximate the quantile function of a distribution if the closed form is unknown. Section 4.2 is concerned with the convolution of random variables, and particularly examines the computational complexity of convolving discrete distributions. In Sect. 4.3 we develop a closed expression for the expected value of conditioned normal distributions. Unfortunately, there is no general approach known for this problem, and we therefore need to exploit the specific properties of the normal distribution to carry out the analytical steps required. Nonetheless, the example of normal distribution may indicate what generally needs to be done to derive the corresponding characteristics. Finally, we introduce the general frameworks of truncated distributions (Sect. 4.4) and mixed distributions (Sect. 4.5), two powerful concepts modeling periodic review inventory systems.

4.1 Approximation of the Quantile Function

A standard problem in inventory theory is to determine the minimum x that satisfies $P(X \leq x) = p$, i.e., given a random variable X, we ask for the value x for which p is the probability that random draws from X would fall below x.

T. Wensing, *Periodic Review Inventory Systems*, Lecture Notes in Economics and Mathematical Systems 651, DOI 10.1007/978-3-642-20479-1_4,

Definition 13 (Quantile function). Let X be an arbitrarily distributed random variable with cumulative density $CDF_X(.)$. Then

$$CDF_X^{-1}(p) = \min\{x : CDF_X(x) = p\}$$

is the quantile function or inverse cumulative density function.

There is no general closed expression for $CDF_X^{-1}(p)$, and the closed form is unknown for many distributions. However, from

$$CDF_X(x) = \int_{-\infty}^{x} PDF_X(z)\, dz, \quad PDF_X(z) \geq 0 \quad \forall \quad z \in \mathbb{R} \tag{4.1}$$

it follows that $CDF_X(x)$ is monotonic and we can approximate x using the bisection method. (See Burden and Faires (2005), Chap. 1 and Knuth (1997), Chap. 6 for example.)

Assuming that we can limit the distribution of X to a relevant interval $[x_{min}, \ldots, x_{max}]$, with $CDF_X(x_{min}) \approx 0$ and $CDF_X(x_{max}) \approx 1$, we may generally apply Algorithm 1 to approximate $CDF_X^{-1}(p)$ within a given tolerance. Defining this and all other algorithms, we use the following notation: $x = y$ means that variable x is set to the value of y, whereas $x == y$ checks whether x and y are set to equal values.

Algorithm 1: Approximation of the quantile function for general assumptions

Input: Distribution X, probability p, tolerance t_x
Output: Quantile x
$x_{lb} = x_{min}$;
$x_{ub} = x_{max}$;
$x = \frac{1}{2} \cdot (x_{lb} + x_{ub})$;
$q = CDF_X(x)$;
while $|p - q| > t_x$ **do**
  if $p < q$ **then**
    $x_{lb} = x$;
  else
    $x_{ub} = x$;
  end
  $x = \frac{1}{2} \cdot (x_{lb} + x_{ub})$;
  $q = CDF_X(x)$;
end

Algorithm 1 works correctly, even if the cumulative density function is not strongly monotonic, i.e., $PDF_X(x) \geq 0\ \forall x \in [x_{min}, x_{max}]$ holds, but $PDF_X(x) > 0\ \forall x \in [x_{min}, x_{max}]$ does not hold. If our required value lies on a plateau of weakly monotonic cumulative density on which all values x have the same cumulative density, the search interval will reduce around the leftmost point of the plateau, as we only set $x_{lb} = x$ if $CDF_X(x)$ is *truly* smaller than the required value, and $x_{ub} = x$ otherwise.

Let us now examine the computational complexity Algorithm 1. Note that each step of the *while* loop halves the width of the interval $[x_{min}, \ldots, x_{max}]$. The interval's width after i steps is therefore equal to $\frac{x_{max}-x_{min}}{2^i}$. The algorithm definitively terminates if the interval is smaller than the given fault tolerance t_x. We thus observe the following maximum number of steps i:

$$\frac{x_{max} - x_{min}}{2^i} \leq t_x$$
$$\Leftrightarrow i \leq \log_2\left(\frac{x_{max} - x_{min}}{t_x}\right)$$

Let CC_{CDFX} be the runtime effort to compute the cumulative density of X for any value x. We can then denote the computational complexity for Algorithm 1 (CC_{a1}) as follows:

$$\begin{aligned} CC_{a1} &\leq i \cdot CC_{CDFX} \\ &\leq \log_2\left(\frac{x_{max} - x_{min}}{t_x}\right) \cdot CC_{CDFX} \end{aligned} \tag{4.2}$$

Let $x_{max} - x_{min} \leq 10^k$ and $t_x \geq 10^{-l}$, then we have (4.3):

$$\begin{aligned} CC_{a1} &\leq \log_2\left(\frac{10^k}{10^{-l}}\right) \cdot CC_{CDFX} \\ &\leq \log_2(10) \cdot (k + l) \cdot CC_{CDFX} \\ &= O((k + l) \cdot CC_{CDFX}) \end{aligned} \tag{4.3}$$

4.2 Convolution of Random Variables

In the following, we will repeatedly encounter the problem of analytically handling the sum of two or more random variables, i.e., given two random variables X and Y, we ask for the distribution Z of the sum $z = x + y$ of random draws from X and Y.

We therefore define the *convolution* of random variables in this section for the two cases of continuously and discretely distributed random variables. In preparation of runtime examinations in Chap. 5 that are also relevant for Chap. 6, we furthermore give a detailed runtime analysis on two algorithms to determine the convolution of discretely distributed random variables.

Definition 14 (Convolution of continuous random variables). Let X and Y be two independent continuous random variables, with density functions PDF_X and PDF_Y. Let Z be a random variable given by PDF_Z:

$$\begin{aligned} PDF_Z(z) &= (PDF_X * PDF_Y)(z) \\ &= \int_{-\infty}^{\infty} PDF_X(a) \cdot PDF_Y(z-a)\, da \\ &= \int_{-\infty}^{\infty} PDF_X(z-a) \cdot PDF_Y(a)\, da \end{aligned} \tag{4.4}$$

Then Z is called the convolution or faltung of X and Y.

Definition 15 (Convolution of discrete random variables). Let X and Y be two independent discretely distributed random variables. The convolution of their density functions is then given as follows:

$$\begin{aligned} PDF_Z(z) &= \sum_{a=-\infty}^{\infty} PDF_X(a) \cdot PDF_Y(z-a) \\ &= \sum_{a=-\infty}^{\infty} PDF_X(z-a) \cdot PDF_Y(a) \end{aligned} \tag{4.5}$$

For further details, please refer to Ross (2006) or Strichartz (2003), for example.

If the sum of more than two random variables needs to be determined, one may exploit the fact that the convolution of functions is generally associative, i.e.

$$f_1 * (f_2 * f_3) = (f_1 * f_2) * f_3$$

holds, where f_1, f_2 and f_3 are arbitrary functions.

4.2.1 Continuous Distributions

For continuously distributed random variables, the integral in (4.4) may not generally be solved in closed form. Under some assumptions, however, we may easily calculate the convolution of two density functions. If both X and Y are normally distributed, $Z = X + Y$ is also normally distributed and expression (4.6) applies, i.e., Z is parameterized by the sum of X and Y's mean value and variance:

$$Z \sim NORM(\mu_Z = \mu_X + \mu_Y, \sigma_Z^2 = \sigma_X^2 + \sigma_Y^2) \tag{4.6}$$

As the convolution is associative, we can derive the convolution of n normally distributed variables $\mathcal{X}$ directly from (4.6):

$$Z \sim NORM(\mu_Z, \sigma_Z)$$

$$\mu_Z = \sum_{X_i \in \mathcal{X}} \mu_{X_i}$$
$$\sigma_Z^2 = \sum_{X_i \in \mathcal{X}} \sigma_{X_i}^2 \tag{4.7}$$

For the special case of n normally distributed variables $\mathcal{X}$ following identical independent distributions, obviously (4.8) holds:

$$Z \sim NORM(n \cdot \mu_X, n \cdot \sigma_X) \tag{4.8}$$

The same findings apply for the convolution of (independently) Gamma-distributed and exponentially distributed random variables. In the event that (4.4) is not given in closed form, one of the numerous methods of numerical integration may be applied. See Davis and Rabinowitz (1984), for example.

4.2.2 Discrete Distributions

Let us have a closer look at (4.5) to estimate the computational effort for evaluating the denoted sum. Assuming that both X and Y only have non-zero probabilities for certain (ordered) values $[x_{min}, \ldots, x_{max}]$ and $[y_{min}, \ldots, y_{max}]$, respectively, Z may obviously only have positive probabilities for values in $[z_{min} = x_{min} + y_{min}, \ldots, z_{max} = x_{max} + y_{max}]$. Thus, probabilities for a total of $z_{max} - z_{min} = x_{max} - x_{min} + y_{max} - y_{min}$ values have to be calculated to determine the distribution of Z.

Let us now consider the effort required to compute the probability for *one* such value. For this purpose, let X and Y be restricted to values of two finite ordered sets so that we are able to identify a lower (lb) and upper bound (ub) for any given z that limit the summation in (4.5), leading to (4.9):

$$(PDF_X * PDF_Y)(z) = \sum_{a=lb(z)}^{ub(z)} PDF_X(a) \cdot PDF_Y(z-a) \tag{4.9}$$

We can determine $lb(z)$ and $ub(z)$ regarding the overlap of the two ordered sets, $[x_{min}, \ldots, x_{max}]$ and $[z - y_{max}, \ldots, z - y_{min}]$. In the event that a does not fall into both sets, clearly $PDF_X(a) \cdot PDF_Y(z-a) = 0$ holds. Regarding z_{min} and z_{max}, the two sets always overlap.

Proof. The sets always overlap if $x_{min} \leq z - y_{min}$ and $z - y_{max} \leq x_{max}$ for all z. This can easily be shown as follows:

$$x_{min} \leq z - y_{min} \quad \forall \quad z$$
$$\Leftrightarrow \quad x_{min} \leq z_{min} - y_{min}$$

$$\begin{aligned}
&\Leftrightarrow \ x_{min} \leq x_{min} + y_{min} - y_{min} \\
&\Leftrightarrow \ x_{min} \leq x_{min} \qquad \surd && (4.10) \\
&\quad z - y_{max} \leq x_{max} \quad \forall \quad z \\
&\Leftrightarrow \ z_{max} - y_{max} \leq x_{max} \\
&\Leftrightarrow \ x_{max} + y_{max} - y_{max} \leq x_{max} \\
&\Leftrightarrow \ x_{max} \leq x_{max} \qquad \surd && (4.11)
\end{aligned}$$

□

Thus, we can state $lb(z)$ and $ub(z)$:

$$lb(z) = \max\{x_{min}, z - y_{max}\} \tag{4.12}$$

$$ub(z) = \min\{x_{max}, z - y_{min}\} \tag{4.13}$$

According to (4.9), the computation of $(PDF_X * PDF_Y)(z)$ takes $ub(z) - lb(z) + 1$ steps. Thus, the computational complexity CC_{DCS} to determine one single value of the convolution of two discretely distributed random variables is given by (4.14):

$$CC_{DCS} = \sum_{z=z_{min}}^{z_{max}} ub(z) - lb(z) + 1 \tag{4.14}$$

For clarity, let $X_{span} = x_{max} - x_{min}$ be the value span of the distribution of X. Without loss of generality, we assume $X_{span} \leq Y_{span}$, i.e., we consider two discrete distributions, where we associate the one with the smaller value span with X, and the other with Y. Let $z := x_{min} + y_{min} + a, a \in \{0, 1, \ldots, Z_{span}\}$. We then receive the following piecewise defined functions for $lb(a)$ and $ub(a)$:

$$lb(z) = \max\{x_{min}, x_{min} - Y_{span} + a\} = \begin{cases} x_{min} & \text{if } a \leq Y_{span} \\ a - Y_{span} & \text{if } a \geq Y_{span} \end{cases} \tag{4.15}$$

$$ub(z) = \min\{x_{max}, x_{min} + a\} = \begin{cases} x_{min} + a & \text{if } a \leq X_{span} \\ x_{max} & \text{if } a \geq X_{span} \end{cases} \tag{4.16}$$

This directly implies the following piecewise defined function for $ub(z) - lb(z)$:

$$ub(z) - lb(z) = \begin{cases} a & \text{if } < a \leq X_{span} \\ X_{span} & \text{if } Y_{span} \leq a \leq X_{span} \\ X_{span} - b & \text{if } a \geq Y_{span} \end{cases} \tag{4.17}$$

$$b = a - Y_{span}$$

We thus derive the computational complexity of determining the convolution of two discrete distributions depending on the number of non-zero probability values of the latter according to (4.18):

$$
\begin{aligned}
CC_{DC} &= \sum_{z=z_{min}}^{z_{max}} ub(z) - lb(z) + 1 \\
&= 1 + 2 + \cdots + X_{span} + (X_{span} + 1) + \cdots + (X_{span} + 1) + X_{span} \\
&\quad + \cdots + 2 + 1 \\
&= \frac{X_{span} \cdot (X_{span} + 1)}{2} + (X_{span} + 1) \cdot (Y_{span} - X_{span} + 1) \\
&\quad + \frac{X_{span} \cdot (X_{span} + 1)}{2} \\
&= X_{span} \cdot (X_{span} + 1) + (X_{span} + 1) \cdot (Y_{span} - X_{span} + 1) \\
&= (X_{span} + 1) \cdot (Y_{span} + 1)
\end{aligned}
\tag{4.18}
$$

To understand (4.18), note that the number of combinations to consider for values in $[z_{min}, \ldots, z_{max}]$ is pyramid-shaped. We have one combination for z_{min} and again for z_{max} on the other side, two for $(z_{min}+1)$ and $(z_{max}-1)$, and so on, until the maximum number of combinations $(X_{span} + 1)$ is reached (remember that we assume $X_{span} \leq Y_{span}$). Clearly we have $(X_{span} + 1)$ values that may be generated by $(X_{span} + 1)$ combinations of values from X and Y.

Because of the associativity, the convolution of multiple random variables can be determined by multiple application of (4.9). Algorithm 2 illustrates the approach.

Algorithm 2: Convolution of n discrete distributions

Input: Array of discrete distributions $DD[1, \ldots, n]$
Output: Convolution con of DD
$con = DD[1]$;
for $i = 2$ *to* n **do**
| $con = con * DD[i]$
end

Let X_{span} now be the greatest span of all random variables in DD. We can then determine the computational complexity CC_{a2} of Algorithm 2 as follows:

$$
\begin{aligned}
CC_{a2} &\leq \sum_{i=2}^{n} ((i-1) \cdot X_{span} + 1) \cdot (X_{span} + 1) \\
&= (X_{span} + 1) \cdot \sum_{i=1}^{n-1} (i \cdot X_{span} + 1) \\
&= (X_{span} + 1) \cdot \left[\frac{(n-1) \cdot n}{2} \cdot X_{span} + n - 1 \right]
\end{aligned}
$$

$$
\begin{aligned}
&= (X_{span} + 1) \cdot \left[\frac{(n-1) \cdot n}{2} \cdot (X_{span} + 1) - \frac{(n-1) \cdot n - 2 \cdot n + 2}{2} \right] \\
&= \frac{(n-1) \cdot n}{2} \cdot (X_{span} + 1)^2 - \frac{(n-3) \cdot n + 2}{2} \cdot (X_{span} + 1) \\
&= O(n^2 \cdot X_{span}^2)
\end{aligned}
\tag{4.19}
$$

If all random variables in DD have the same distribution, we can apply Algorithm 3, which runs faster than Algorithm 2 in this special case. The idea is as follows. In a first loop, the convolutions of $1, 2, 4, 8, \ldots, 2^i$ distributions are determined, where $i = \lfloor \log_2(n) \rfloor$, i.e., the greatest exponent i for which 2^i is smaller than or equal to n. The required convolution is then constructed using the corresponding smaller convolutions. E.g., let $X^{[n]}$ denote the convolution of n random variables with independent identical distributions. $X^{[14]}$ is then constructed as $X^{[14]} = X^{[8]} * X^{[4]} * X^{[2]}$. As there obviously always exists an $\overrightarrow{x} \in \{0, 1\}^{\lfloor \log_2(n) \rfloor}$ so that $n = \sum_{x_i \in \overrightarrow{x}} x_i \cdot 2^i, n \leq 2^i$, the approach is correct.

Algorithm 3: Convolution of n i.i.d. random variables

Input: Discrete distribution dd, Integer n
Output: Convolution con of n random variables having the distribution dd
$CONS$ = new DiscreteDistribution[$\lfloor \log_2(n) \rfloor$];
$CONS[1]$ = dd;
for $i = 2$ *to* $\lfloor \log_2(n) \rfloor$ **do**
  $CONS[i] = CONS[i-1] * CONS[i-1]$;
end
$con = CONS[\lfloor \log_2(n) \rfloor]$;
$n = n - 2^{(\lfloor \log_2(n) \rfloor)}$;
for $i = \lfloor \log_2(n) \rfloor - 1$ *to* 1 **do**
  if $2^i \leq n$ **then**
    $con = con * CONS[i]$;
    $n = n - 2^i$;
  end
end

The computational complexity is driven by the two *for*-loops. For clarity, let us first consider them separately, and then combine the result. In each step of the first loop, two distributions with equal span are convolved, leading to a complexity of $\sum_{i=1}^{\lfloor \log_2(n) \rfloor} (2^{(i-1)} \cdot X_{span} + 1)^2$. This can be simplified as follows:

$$
\begin{aligned}
CC_{a3_{l1}} &= \sum_{i=1}^{\lfloor \log_2(n) \rfloor} (2^{(i-1)} \cdot X_{span} + 1)^2 \\
&= \sum_{i=1}^{\lfloor \log_2(n) \rfloor} 4^{(i-1)} \cdot X_{span}^2 + 2 \cdot 2^{i-1} \cdot X_{span} + 1
\end{aligned}
$$

$$= \left[\frac{4^{\lfloor \log_2(n) \rfloor}}{3} - 1\right] \cdot X_{span}^2 + 2 \cdot (2^{\lfloor \log_2(n) \rfloor} - 1) \cdot X_{span} + \lfloor \log_2(n) \rfloor \quad (4.20)$$

$$\leq \frac{n^2 - 1}{3} \cdot X_{span}^2 + 2 \cdot (n-1) \cdot X_{span} + \log_2(n) \quad (4.21)$$

Note that (4.20) is equal to (4.21) if $\lfloor \log_2(n) \rfloor = \log_2(n)$, i.e., for all $n = 2^i, i \in \mathbb{Z}$.

The second loop is trickier. Let us consider only the worst case here, where we have to convolve all distributions that we have prepared in the array $cons$. This is the case for $n = 2^i - 1, i \in \mathbb{Z}$. We then experience the following computational complexity of the second loop:

$$CC_{a3_{l2}}$$

$$= \sum_{i=1}^{\lfloor \log_2(n) \rfloor} \left[2^{\lfloor \log_2(n) \rfloor - i} \cdot X_{span} + 1\right] \cdot \left[\left(\sum_{j=1}^{i} 2^{\lfloor \log_2(n) \rfloor + 1 - j}\right) \cdot X_{span} + 1\right]$$

$$= \sum_{i=1}^{\lfloor \log_2(n) \rfloor} \left[2^{\lfloor \log_2(n) \rfloor - i} \cdot X_{span} + 1\right]$$

$$\cdot \left[(2^{\lfloor \log_2(n) \rfloor + 1} - 2^{\lfloor \log_2(n) \rfloor + 1 - i}) \cdot X_{span} + 1\right]$$

$$= \sum_{i=1}^{\lfloor \log_2(n) \rfloor} \left[2^{-(i+1)} \cdot 2^{\lfloor \log_2(n) \rfloor + 1} \cdot X_{span} + 1\right]$$

$$\cdot \left[(1 - 2^{-i}) \cdot 2^{\lfloor \log_2(n) \rfloor + 1} \cdot X_{span} + 1\right]$$

$$= \sum_{i=1}^{\lfloor \log_2(n) \rfloor} (2^{-(i+1)} - 2^{-(2 \cdot i + 1)}) \cdot 4^{\lfloor \log_2(n) \rfloor + 1} \cdot X_{span}^2$$

$$+ \sum_{i=1}^{\lfloor \log_2(n) \rfloor} (1 - 2^{-(i+1)}) \cdot 2^{\lfloor \log_2(n) \rfloor + 1} \cdot X_{span} + 1$$

$$= 4^{\lfloor \log_2(n) \rfloor + 1} \cdot X_{span}^2 \cdot \frac{1}{2} \cdot \left[\sum_{i=1}^{\lfloor \log_2(n) \rfloor} 2^{-i} - 4^{-i}\right]$$

$$+ 2^{\lfloor \log_2(n) \rfloor + 1} \cdot X_{span} \cdot \left[\lfloor \log_2(n) \rfloor - \frac{1}{2} \cdot \sum_{i=1}^{\lfloor \log_2(n) \rfloor} 2^{-i}\right] + \sum_{i=1}^{\lfloor \log_2(n) \rfloor} 1$$

$$= 4^{\lfloor \log_2(n) \rfloor + 1} \cdot X_{span}^2 \cdot \frac{1}{2} \cdot \left[1 - 2^{-\lfloor \log_2(n) \rfloor} - \frac{1 - 4^{-\lfloor \log_2(n) \rfloor}}{3} \right]$$

$$+ 2^{\lfloor \log_2(n) \rfloor + 1} \cdot X_{span} \cdot \left[\lfloor \log_2(n) \rfloor - \frac{1}{2} \cdot (1 - 2^{-\lfloor \log_2(n) \rfloor}) \right] + \lfloor \log_2(n) \rfloor \tag{4.22}$$

$$\leq 4 \cdot n^2 \cdot X_{span}^2 \cdot \frac{1}{2} \cdot (1) + 2 \cdot n \cdot X_{span} \cdot (\log_2(n) - \frac{1}{2}) + \log_2(n)$$

$$(=)\ 2 \cdot n^2 \cdot X_{span}^2 + 2 \cdot n \cdot \log_2(n) \cdot X_{span} - n \cdot X_{span} + \log_2(n) \tag{4.23}$$

Combining these findings, we can approximate the complexity for the whole algorithm by adding (4.21) and (4.23):

$$\begin{aligned} CC_{a3} &= CC_{a3_{l1}} + CC_{a3_{l2}} \\ &\leq \frac{7 \cdot n^2 - 1}{3} \cdot X_{span}^2 + 2 \cdot n \cdot \log_2(n) \cdot X_{span} + (n-2) \cdot X_{span} \\ &\quad + 2 \cdot \log_2(n) \qquad (4.24) \\ &= O(n^2 \cdot X_{span}^2) \end{aligned}$$

Although Algorithms 2 and 3 fall into the same complexity class in terms of the O-Notation, we observe a significantly lower runtime of Algorithm 3. For example, for $x_{span} = 20$, $N = 2, \ldots, 100$, Algorithm 3 is faster in every instance where the middle ratio is $1 : 0.74$.

4.3 Mass Integral of the Normal Distribution

The mean value of a continuously distributed random variable X is defined as follows:

$$\mu = \int_{-\infty}^{\infty} PDF_X(x) \cdot x \, dx \tag{4.25}$$

For the analysis in the following sections, we will repeatedly consider not the full integral, but only a fraction between some certain bounds a and b:

$$m(a, b) = \int_a^b PDF_X(x) \cdot x \, dx \tag{4.26}$$

The integral may not be solved into a closed form for general distribution assumptions. In the event that X is normally distributed, we may solve instances of (4.26) using the probability density and cumulative density function. We will show the corresponding reformulations in the following:

$$
\begin{aligned}
m(a,b) &= \int_a^b PDF_X(x) \cdot x \, dx \\
&= \int_a^b \frac{1}{\sigma_X \sqrt{2 \cdot \pi}} \cdot e^{\left(-\frac{x-\mu_X}{2 \cdot \sigma_X}\right)^2} \cdot x \, dx \\
&= \int_a^b \frac{1}{\sigma_X} \cdot \frac{1}{\sqrt{2 \cdot \pi}} \cdot e^{\left(-\frac{x-\mu_X}{2 \cdot \sigma_X}\right)^2} \cdot \left(\frac{x-\mu_X}{\sigma_X} \cdot \sigma_X + \mu_X\right) dx \\
&= \int_a^b \left(\frac{\sigma_X}{\sigma_X} \cdot \frac{1}{\sqrt{2 \cdot \pi}} \cdot e^{\left(-\frac{x-\mu_X}{2 \cdot \sigma_X}\right)^2} \cdot \left(\frac{x-\mu_X}{\sigma_X}\right)\right. \\
&\quad \left. + \frac{1}{\sigma_X} \cdot \frac{1}{\sqrt{2 \cdot \pi}} \cdot e^{\left(-\frac{x-\mu_X}{2 \cdot \sigma_X}\right)^2} \cdot \mu_X \right) dx \\
&= \int_a^b \left(\frac{1}{\sqrt{2 \cdot \pi}} \cdot e^{\left(-\frac{x-\mu_X}{2 \cdot \sigma_X}\right)^2} \cdot \left(\frac{x-\mu_X}{\sigma_X}\right)\right. \\
&\quad \left. + \frac{1}{\sigma_X} \cdot \frac{1}{\sqrt{2 \cdot \pi}} \cdot e^{\left(-\frac{x-\mu_X}{2 \cdot \sigma_X}\right)^2} \cdot \mu_X \right) dx
\end{aligned} \tag{4.27}
$$

Applying integration by substitution (4.28), we derive (4.29) and (4.30):

$$
\int_a^b u(v(x)) \, dx = \int_a^b \frac{1}{v'(x)} \cdot U(v(x)) \, dx \tag{4.28}
$$

$$
\begin{aligned}
\int_a^b \frac{1}{\sqrt{2 \cdot \pi}} \cdot e^{\left(-\frac{x-\mu_X}{2 \cdot \sigma_X}\right)^2} \cdot \left(\frac{x-\mu_X}{\sigma_X}\right) dx &= -\sigma_X \cdot \frac{1}{\sqrt{2 \cdot \pi}} \cdot e^{\left(-\frac{x-\mu_X}{2 \cdot \sigma_X}\right)^2} \Bigg|_a^b \\
&= -\sigma_X^2 \cdot PDF_X(x) \Bigg|_a^b
\end{aligned} \tag{4.29}
$$

$$
\begin{aligned}
\int_a^b \frac{1}{\sigma_X} \cdot \frac{1}{\sqrt{2 \cdot \pi}} \cdot e^{\left(-\frac{x-\mu_X}{2 \cdot \sigma_X}\right)^2} \cdot \mu_X &= \int_a^b PDF_X(x) \cdot \mu_X \, dx \\
&= \mu_X \cdot CDF_X(x) \Bigg|_a^b
\end{aligned} \tag{4.30}
$$

Thus, we derive (4.31):

$$
\begin{aligned}
&\int_a^b PDF_X(x) \cdot x \, dx \\
&\quad = \sigma_X^2 \cdot (PDF_X(a) - PDF_X(b)) + \mu_X \cdot (CDF_X(b) - CDF_X(a))
\end{aligned} \tag{4.31}
$$

We can derive the following more general expression from (4.31):

$$
\begin{aligned}
& mr(a,b,c) \\
&= \int_a^b PDF_X(x) \cdot (x-c)\, dx && (4.32) \\
&= \int_a^b PDF_X(x) \cdot x\, dx - \int_a^b PDF_X(x) \cdot c\, dx \\
&= \sigma^2 \cdot (PDF_X(a) - PDF_X(b)) + \mu_X \cdot (CDF_X(b) - CDF_X(a)) \\
&\quad - c \cdot (CDF_X(b) - CDF_X(a)) \\
&= \sigma_X^2 \cdot (PDF_X(a) - PDF_X(b)) + (\mu_X - c) \cdot (CDF_X(b) - CDF_X(a)) && (4.33)
\end{aligned}
$$

If the integral is unbounded on the right side ($b \rightarrow \infty$), (4.33) is simplified according to (4.34):

$$mr(a, b \rightarrow \infty, c) = \sigma_X^2 \cdot PDF_X(a) + (\mu_X - c) \cdot (1 - CDF_X(a)) \tag{4.34}$$

4.4 Truncated Distributions

A truncated distribution may be described as the conditional equivalent of an arbitrary distribution, whose values are bound on a given interval $[a, b]$. Let X be an arbitrarily distributed random variable, and let $x_1, x_2, \ldots$ be random draws from X, where we ignore all draws that do not lie in $[a, \ldots b]$. In that case, the random draws we account for follow the corresponding truncated distribution of X.

Definition 16 (Truncated distribution). Let X be an arbitrary random variable with arbitrary support, density PDF_X and cumulative density CDF_X. Furthermore, let $X_{(a,b)}$ be a random variable with a density function $PDF_{X_{(a,b)}}$, given by (4.35) and (4.36):

$$PDF_{X_{(a,b)}}(x) \notin [a, b] = 0 \tag{4.35}$$

$$PDF_{X_{(a,b)}}(x) \in [a, b] = \frac{PDF_X(x)}{CDF_X(b) - CDF_X(a)} \tag{4.36}$$

Then the distribution of $X_{(a,b)}$ is called the truncated distribution of X with its support restricted to $x \in [a, b]$.

The cumulative density function can be derived as follows:

$$CDF_{X_{(a,b)}}(x) | x \leq a = 0 \tag{4.37}$$

$$\begin{aligned}
CDF_{X_{(a,b)}}(x)|x \in [a,b] &= \int_a^x PDF_{X_{(a,b)}}(z)\,dz \\
&= \int_a^x \frac{PDF_X(z)}{CDF_X(b) - CDF_X(a)}\,dz \\
&= \frac{1}{CDF_X(b) - CDF_X(a)} \cdot \int_a^x PDF_X(z)\,dz \\
&= \frac{CDF_X(x) - CDF_X(a)}{CDF_X(b) - CDF_X(a)} \qquad (4.38)
\end{aligned}$$

$$CDF_{X_{(a,b)}}(x)|x \geq b = 1 \qquad (4.39)$$

For further details on truncated distributions, please refer to Johnson et al. (1994), for example.

4.5 Mixed Distributions

A mixed distribution is the weighted arithmetical mean of at least two different arbitrary distributions. In the following, we will examine inventory models for which the lead time demand follows a mixed distribution. A formal definition is given in Titterington et al. (1985):

Definition 17 (Mixed distribution). Let X be a random variable with values in the probability space $\mathbb{W}$, with a density function PDF_X, given by

$$PDF_X(x) = \sum_{j=1}^{k} \omega_j \cdot PDF_j(x), \quad x \in \mathbb{W}$$

and it holds:

$$\omega_j > 0, \quad \forall j, \quad \sum_{j=1}^{k} \omega_j = 1;$$

$$PDF_j(.) \geq 0, \quad \int_{\mathbb{W}} PDF_j(x)\,dx = 1, \quad \forall j.$$

Then X follows a mixed distribution, with weights ω_j and density components $PDF_j(.)$.

The cumulative density function can be derived from the corresponding expressions of the components:

$$
\begin{aligned}
CDF_X(x) &= \int_{-\infty}^{x} \sum_{j=1}^{k} \omega_j \cdot PDF_j(z)\, dz \\
&= \sum_{j=1}^{k} \omega_j \cdot \int_{-\infty}^{x} PDF_j(z)\, dz \\
&= \sum_{j=1}^{k} \omega_j \cdot CDF_j(x)
\end{aligned}
\tag{4.40}
$$

Chapter 5
Replenishment Processes

This chapter focuses on the replenishment aspect of static stochastic inventory models, namely the modeling of the replenishment lead time generating process. Before starting, let us briefly recapitulate the contents of the previous chapters. To begin with, we described the fundamental terminology related to speaking of inventory systems and telling one system from another in Chap. 2. We then gave an overview of the relevant literature on single-level static stochastic inventory systems in Chap. 3. We will revisit some of the papers that we introduced there in this chapter with respect to the analysis of the replenishment process. Finally, we described some basic methods of stochastic analysis in Chap. 4 that we will partially revisit in the approaches described in this chapter.

Whenever the replenishment lead times in an inventory system are considered to be stochastic, it is indispensable for the analysis to specify the underlying lead time generating process. Even where lead times have the same distribution, the behavior of two systems may be very different if the underlying lead time models vary. This chapter distinguishes three lead time models that are discussed in the literature, namely the cases of:

- *Non-interchangeability* of demand units (Sect. 5.1)
- Replenishment *order crossover* (Sect. 5.2)
- *Sequential arrivals* of replenishment orders (Sect. 5.3)

For a brief literature overview of these three cases see Hayya et al. (2008), for example.One might add a fourth case where lead time *distributions* are *limited* in order to result in the same system behavior for all three basic cases. We consider this a special case of all three general models and discuss it in Sect. 5.4.

5.1 Non-Interchangeability

An analytically attractive idea is to consider non-interchangeable unit demands, i.e., each demand unit can only be satisfied by a specific replenishment unit arriving with a specific replenishment order. The idea was introduced in a technical paper by

T. Wensing, *Periodic Review Inventory Systems*, Lecture Notes in Economics and Mathematical Systems 651, DOI 10.1007/978-3-642-20479-1_5,

Washburn (1973) on a stochastic lead time extension to the classical economic order quantity problem. Similar problems are regarded by Liberatore (1979), Sphicas (1982), Sphicas and Nasri (1984) and more recently He et al. (2005). Common to all four approaches is the assumption of a constant demand rate, allowing for the synchronization of replenishment orders and prospective demand units.

What makes the assumption so attractive is the property that an arbitrary lead time distribution may be considered in the inventory model without causing any stochastic dependencies. Due to the fixed assignment, each demand unit will be fulfilled on its arrival or on the arrival of the corresponding replenishment order, whichever event is the later. The analysis can thus be focused on one replenishment cycle, and lead times may be independent.

For real inventory systems with stochastic demands, however, it is unlikely that a demand unit must inevitably be satisfied by material arriving with a specific replenishment order. It is somewhat contrary to the idea of keeping stock in order to serve an uncertain demand. Nonetheless, it is worth checking wether the assumption is met in a real system, as it significantly simplifies the analysis.

5.2 Order Crossover

Two orders are said to cross over in time if they arrive in a different sequence to that in which they were issued, i.e., the order that was issued later arrives earlier. From a theoretical point of view, this phenomenon may generally be observed in inventory systems if lead times are independent random variables.

A formal definition is given by Riezebos (2006).

Definition 18 (Order crossover). Let A and B be two replenishment orders that were issued at time o_A and o_B, respectively, where $o_A < o_B$ holds, i.e., A was issued earlier than B. Let A and B arrive at times a_A and a_B: these orders then cross if $a_A > a_B$.

Definition 18 implies corollary 1.

Corollary 1. *Let A and B be two orders that observe issue and arrival times as in Definition 18. Furthermore, let l_A and l_B be the lead times of these orders, so that $a_A = o_A + l_A$ and $a_B = o_B + l_B$. Then A and B cross if and only if $o_A - o_B + l_A > l_B$.*

Proof. Corollary 1 follows from inserting equations $a_A = o_A + l_A$ and $a_B = o_B + l_B$ into inequation $a_A > a_B$. □

The corollary directly implies that lead time variability is a prerequisite for order crossover. To be more specific, crossover occurs whenever $o_B - o_A < l_A - l_B$, i.e., it can only be observed in systems with a lead time variability higher than the minimum time between two orders. (Also see Sect. 5.4.)

Riezebos distinguishes stochastic and dynamic lead time variability, both of which are reported to be observable in real-life inventory systems, either solely or in combination. Stochastic fluctuations are related to time-invariant uncertainty in the supply processes, while dynamic fluctuations are due to systematic process changes over time. For the latter, an example of ordering tobacco leaves is given by Riezebos, where individual ordering modalities and production and transport times in different countries of origin introduce high lead time volatility to the overall ordering process. As dynamic lead times – or system dynamics in general – require an inventory system with dynamic parametrization, we will exclusively focus on the (static) stochastic lead time case in the following.

From a process perspective Song and Zipkin (1996) give a clear motivation for real world systems with static stochastic lead times at which order crossover may and may not be observed. Orders may cross over whenever they are processed in parallel, and cannot cross over if they are sequentially processed. The first case is considered in this section; the latter case in general means that lead times are dependent (see Sect. 5.3).

Two further modalities can be distinguished regarding parallel processing. On the one hand, each order may be assigned to a certain parallel process; on the other hand, we have random assignment that may not be controlled by the system that issues orders.

The first case is broadly discussed in the literature in the context of multiple sourcing, which means that a company maintains more than one supplier for the same type of goods. In this context, problems of order crossover are addressed by Ramasesh et al. (1991) and Kelle and Silver (1990), for example. Both articles consider a multiple sourcing strategy for a continuous review system with fixed order sizes, where each order quantity is split among at least two suppliers. Besides the crossover phenomenon, the related problem of determining so-called *effective lead times* may arise in the context of multiple sourcing. See Sculli and Wu (1981), Sculli and Shum (1990), Pam et al. (1991) and Fong et al. (2000), for example. In these papers, the effective lead time is regarded as the lead time of the first, second, etc... arrival of two or more simultaneously issued orders. In Sect. 5.2.3 we consider the more general problem of determining the time that passes between the n-th issue of an order and the n-th order arrival, where the corresponding orders may not be the same if order crossover is possible.

A detailed literature overview on multiple sourcing is given by Minner (2003).

Besides the literature on multiple sourcing, some articles directly focus on the order crossover phenomenon, where the actual cause of crossovers is of secondary interest. The related research goes back to the 1950s. See Galliher et al. (1959), for instance, for an early paper on a continuous review system in which order crossover is not exactly addressed, but considered in the model. In spite of these early works, studies on systems with replenishment processes that may randomly cause order crossovers are limited. The majority of authors consider systems where the possibility that orders may cross is simply ignored or cannot occur by definition.

In recent years, however, more attention has been dedicated to the phenomenon. See for example the works of Robinson et al. (2001), Bradley and Robinson (2005) and Robinson and Bradley (2008) on periodic review systems, and those of He et al. (1998) and Hayya et al. (1995) on continuous review systems, where the latter paper also regards the multi- (here dual-) sourcing aspect. An overview of literature concerning order crossovers when lead times are static stochastic is given by Hayya et al. (2008).

In the remainder of this section we will look more closely at three characteristics of the replenishment process when orders may crossover in time, namely the number of outstanding orders, inventory shortfall and the effective lead time.

5.2.1 Outstanding Orders

Zalkind (1978) analyzes the number of outstanding orders for a periodically distributed inventory system. It is assumed that an order is placed in every review interval, i.e., demand is always sufficiently great between two order events that a replenishment order will be placed. To motivate the calculation, we will firstly assume $r = 1$, and derive the general formulation afterwards. An order is considered outstanding in a period if it arrives no sooner than in the next period.

Let L be a discrete random variable with minimum $l_{min} = 0$ and a finite maximum l_{max}, and let the lead times be a series of independent random variables having the distribution of L. Furthermore, let $e_l \in \{0, 1\}$ indicate wether an order placed l periods ago has arrived (0) or is outstanding (1), and $P\{pt = (e_0, e_1, \ldots, e_{l_{max}-1})\}$ be the probability that the pattern of outstanding orders is given by $pt = (e_0, e_1, \ldots, e_{l_{max}-1})$ during the current period. We will consider the state at the end of each period, so that an order with lead time l counts as arrived and not outstanding after l periods. Therefore $e_{l_{max}} = 0$ always holds. Using the properties that e_l is binary and lead times are independent, the corresponding probability can be calculated according to (5.1):

$$P\{pt = (e_0, e_1, \ldots, e_{l_{max}-1})\} = \prod_{l=0}^{l_{max}-1} (1 - e_l) \cdot P\{L \leq l\} + (e_l) \cdot P\{L > l\} \tag{5.1}$$

As $e_l = 0$ for $l \geq l_{max}$, only patterns of length $l_{max} - 1$ need to be considered. In terms of e_l, the number of outstanding orders is given by (5.2):

$$k = \sum_{l=0}^{l_{max}-1} e_l \tag{5.2}$$

Let PT be the set of all patterns $pt = (e_0, e_1, \ldots, e_{l_{max}-1})$ that satisfy (5.2). The probability that there are k orders outstanding is then given by (5.3):

$$P\{K = k\} = \sum_{pt \in PT} P\{PT = pt\} \quad \forall \quad k = 0, \ldots, l_{max} - 1 \tag{5.3}$$

Obviously, there are $\binom{n}{k} = \frac{n!}{k!\cdot(n-k)!}$ n-element patterns $(e_0, e_1, \ldots, e_{n-1})$ that satisfy (5.2) and need to be considered in (5.3). Thus, the calculation of the probability distribution of the number of orders outstanding would have exponential computational complexity $O(2^n), n = l_{max} - 1$ if $r = 1$, as $\sum_{k=0}^{n} \binom{n}{k} = 2^n$. This circumstance can easily be obtained from the well-studied binomial theorem (5.4), which is included here for the reader's convenience. (Set $x = y = 1$.) It is also obvious though, as all 2^n possible patterns have to be considered to calculate the probabilities of all possible values for k:

$$(x+y)^n = \sum_{k=0}^{n} \binom{n}{k} \cdot x^{n-k} y^k \quad \forall \quad x, y \in \mathbb{R}, n \in \mathbb{Z}^+ \tag{5.4}$$

Zalkind indicates an interesting possibility to reduce computational complexity. He introduces a set of *binary* discrete distributions K_l that are defined according to (5.5) and (5.6):

$$P\{K_l = 0\} = P\{L \leq l\} \tag{5.5}$$

$$P\{K_l = 1\} = P\{L > l\} \tag{5.6}$$

Using this set, the probability distribution of the orders outstanding equals the convolution of $K_0, K_1, \ldots, K_{l_{max}-1}$, where the computational complexity is in $O((l_{max})^2)$ (see Sect. 4.2.2).

This idea is also applicable to a generalized $r \geq 0$, as will be shown in the following. With an arbitrary positive review interval, we have to consider that there may be periods in which no orders are issued, and thus exclude all e_l that do not comply with the ordering pattern. In other words: for $r > 1$, only those order ages are possible for a certain subperiod t that reach back to an order period. While $r = 1$ means that order cycles and periods are the same, it is now necessary to focus on order cycles.

Let us first address the question of the maximum orders that may be outstanding, i.e., the maximum order cycles that have to be considered with respect to a certain maximum lead time. Considering period $t \in \{1, 2, \ldots, r\}$ within an order cycle r, we know that any outstanding order must be at least t periods old. Thus, an order can only be outstanding if it has been issued no later than $l_{max} - t$ periods ago. Adjusting this observation to the order cycle pattern, we derive (5.7), where $c_{max}(r, t)$ is the maximum number of cycles (including the present one) from which an outstanding order might originate:

$$c_{max}(r, t) = \left\lfloor \frac{l_{max} - (t-1)}{r} \right\rfloor + 1 \tag{5.7}$$

K_l also needs adjustment. We will therefore define $K_c(t, r)$ as a function indicating that an order issued c cycles ago will be outstanding in the t-th period of an order cycle of length r (5.8 and 5.9), where $c = 1$ is the present cycle:

$$P\{K_c(t,r) = 0\} = \begin{cases} 1 & \text{if } c_{max}(r,t) \leq 0 \\ P\{L \leq (c-1) \cdot r + t - 1\} & \text{else} \end{cases} \tag{5.8}$$

$$P\{K_c(t,r) = 1\} = \begin{cases} 0 & \text{if } c_{max}(r,t) \leq 0 \\ P\{L > (c-1) \cdot r + t - 1\} & \text{else} \end{cases} \tag{5.9}$$

By convolution of the corresponding cycle-dependent $K_c(t,r)$, we derive the distribution of outstanding orders in period t according to (5.10) if the order cycle is r:

$$K(t,r) \sim \; K_1(t,r) * \ldots * K_{c_{max}}(t,r) \tag{5.10}$$

Thus, the overall probability that the inventory system has k orders outstanding is given by (5.11). K is the mixture of r distributions $K(t,r)$, each having influence $\frac{1}{r}$:

$$P\{K = k\} = \sum_{t=1}^{r} \frac{1}{r} \cdot P\{K(t,r) = k\} \tag{5.11}$$

Numerical Example. Let L be discretely uniformly distributed with values {1,2,3,4} and r = 1. Then we have the following relevant distributions K_l:

$$\begin{aligned} &P\{K_1 = 0\} = 0,\ P\{K_2 = 0\} = 0.25,\ P\{K_3 = 0\} = 0.5,\ P\{K_4 = 0\} = 0.75 \\ &P\{K_1 = 1\} = 1,\ P\{K_2 = 1\} = 0.75,\ P\{K_3 = 1\} = 0.5,\ P\{K_4 = 1\} = 0.25 \end{aligned}$$

By convolution, we obtain the following probabilities that k orders are outstanding:

$$\begin{aligned} &P\{K = 1\} = 0.09375,\ P\{K = 2\} = 0.40625,\ P\{K = 3\} = 0.40625, \\ &P\{K = 4\} = 0.09375 \end{aligned}$$

For $r = 2$, the calculations are as follows:

$$\begin{aligned} P\{K_1(1,2) = 0\} = 0,\ P\{K_2(1,2) = 0\} = 0.5 \\ P\{K_1(1,2) = 1\} = 1,\ P\{K_2(1,2) = 1\} = 0.5 \\ P\{K_1(2,2) = 0\} = 0.25,\ P\{K_2(2,2) = 0\} = 0.25 \\ P\{K_1(2,2) = 1\} = 0.75,\ P\{K_2(2,2) = 1\} = 0.75 \\ P\{K(1,2) = 0\} = 0,\ P\{K(1,2) = 1\} = 0.5,\ P\{K(1,2) = 2\} = 0.5 \\ P\{K(2,2) = 0\} = 0.1875,\ P\{K(2,2) = 1\} = 0.625,\ P\{K(2,2) = 2\} = 0.1875 \\ P\{K = 0\} = 0.09375,\ P\{K = 1\} = 0.5625,\ P\{K = 2\} = 0.34375 \end{aligned}$$

5.2.2 Inventory Shortfall

In the previous section, the distribution of outstanding orders is developed for a periodically distributed inventory system. This figure, however, is of limited help for the proper configuration of an inventory system, as one is more interested in the *amount* of stock missing. This amount may be either part of orders that have already been issued, or may be preconsidered for the next order due date. We will use the term inventory shortfall (SF) for the combined amount, as used by Robinson et al. (2001), for example.

An exact formulation of the inventory shortfall in periodically distributed inventory systems is also developed by Zalkind (1978) using his findings on the number of orders outstanding, as described above.

As indicated, inventory shortfall can be divided into (1) the order amount that has already been issued but which has not yet arrived, and (2) the demand that has occurred since the last order was issued. The first part is described by the mixture of the distributions of demand in $k \cdot r$ periods (5.12). The second part is the demand in $(t-1)$ periods (5.13), where we assume that the demand occurring in an order period is always considered with the corresponding replenishment order.

$$SF_t^{(1)} = \sum_{k=1}^{\infty} P\{K(t,r)=k\} \cdot D^{[k \cdot r]} \tag{5.12}$$

$$SF_t^{(2)} = D^{[t-1]} \tag{5.13}$$

The overall distribution of the inventory shortfall in a specific period t is then obtained by proper convolution:

$$\begin{aligned} SF_t &= SF_t^{(1)} * SF_t^{(2)} \\ &= \left[\sum_{k=0}^{\infty} P\{K(t,r)=k\} \cdot D^{[k \cdot r]} \right] * D^{[t-1]} \\ &= \sum_{k=0}^{\infty} P\{K(t,r)=k\} \cdot D^{[k \cdot r]} * D^{[t-1]} \\ &= \sum_{k=0}^{\infty} P\{K(t,r)=k\} \cdot D^{[k \cdot r+t-1]} \end{aligned} \tag{5.14}$$

With r periods forming one order cycle, distribution of the inventory shortfall for the whole system (5.15) is the mixture of r distributions according to (5.14):

$$\begin{aligned} SF &= \sum_{t=1}^{r} \frac{1}{r} \cdot SF_t \\ &= \sum_{t=1}^{r} \frac{1}{r} \sum_{k=0}^{\infty} P\{K(t,r) = k\} \cdot D^{[k \cdot r + t - 1]} \end{aligned} \tag{5.15}$$

One may not want to use a combination of various different mixed distributions for computational convenience. Regarding (5.15) a little closer, we observe that the underlying mixed distributions may share equal elements. Thus, we can directly formulate (5.15) as mixture of convolved demand distributions (5.16), where SFP is the distribution of the number of demand periods constituting the inventory shortfall:

$$SF = \sum_{sfp=1}^{\infty} P\{SFP = sfp\} \cdot D^{[sfp]} \tag{5.16}$$

$$P\{SFP = sfp\} = \sum_{t=1}^{r} \frac{1}{r} \cdot P\left\{K(t,r) = \frac{sfp - t + 1}{r}\right\} \tag{5.17}$$

Numerical Example. Let L be discretely uniformly distributed with values $\{1,2,3,4\}$, D normally distributed with $(\mu = 100, \sigma = 30)$ and r $=$ 1. Using the results of the previous section's example (we obtain the following shortfall distribution)

$$\begin{aligned} SF &= 0.09375 \cdot Norm(100, 30) + 0.40625 \cdot Norm(200, 42.43) \\ &\quad + 0.40625 \cdot Norm(300, 51.96) + 0.09375 \cdot Norm(400, 60) \\ \mu_{SF} &= 250 \\ \sigma_{SF} &= 92.1954 \end{aligned}$$

Robinson et al. (2001) compare inventory shortfall with lead time demand (LTD), which is commonly used in literature and practice to adjust inventory systems. In contrast to SF, LTD is the demand that occurs from the moment an order is placed until the moment when that particular order arrives. Thus, for our first example ($r = 1$), LTD is calculated as follows. While mean values are equal, we notice a significant disparity of standard deviations.

$$\begin{aligned} LTD(d) &= 0.25 \cdot Norm(100, 30) + 0.25 \cdot Norm(200, 42.43) \\ &\quad + 0.25 \cdot Norm(300, 51.96) + 0.25 \cdot Norm(400, 60) \\ \mu_{LTD} &= 250 \\ \sigma_{LTD} &= 121.4496 \end{aligned}$$

Referring to a technical paper by Zalkind (1976), Robinson et al. (2001) give proof that $E\{SF\} = E\{LTD\}$ and $Var\{SF\} \leq Var\{LTD\}$ hold in periodically distributed inventory systems when $r = 1$.

5.2.3 *Effective Lead Time*

Hayya et al. (2008) introduce the concept of *effective lead times* (ELT) to the crossover context in order to describe the replenishment order arrival time series when order crossover is possible.

Definition 19 (Effective lead time). Let o_i be the issue time of the i-th issued replenishment order in an inventory system and let a_i be the arrival time of the i-th arriving order, where o_i and a_i may correspond to two different orders. Then $elt_i = o_i - a_i$ is the i-th effective lead time observed in the system. ELT is the random distribution of the elt_i.

Note that $ELT = L$ holds if no crossovers may occur in the considered inventory system.

Hayya et al. derive expressions to calculate the effective lead time when the time horizon is limited to two and three periods, mainly to illustrate the computational complexity. The idea is to enumerate all possible sequences of arrivals for orders that were issued in k consecutive periods and calculate the probabilities that each sequence will occur. Applying their approach to a time horizon of t periods, the ELT is the mixture of $t!$ distributions, assuming that each order may observe a lead time of t or more periods.

In the following we will use this idea to derive expressions for a steady-state analysis. Instead of restricting the time horizon, we will restrict the maximum number of orders that may be involved in a crossover. For clarity, we will distinguish purchase orders and deliveries in the remainder of this section. We will refer to an order as a purchase order at the time it is issued, whereas we will speak of a delivery if we consider its arrival. For example, if the first and second order of a sequence cross over, we say that the first purchase order is the second delivery and the second purchase order is the first delivery.

Furthermore, let us introduce the following notation to describe the arrival sequence of purchase orders. Let:

- k denote the number of consecutive purchase orders that are examined for a certain purpose
- m be the maximum number of orders in an arrival sequence that any of the orders may cross
- $as_s^{(k,m)}$ be the arrival sequence s of orders that are identified by their positions in the order issue sequence, in which each order may cross over with a maximum of m other orders
- $as_{s,i}^{(k,m)}$ be the issue position of the i-th order in the arrival sequence $as_s^{(k,m)}$, and let
- $AS^{(k,m)}$ be the set of all possible arrival sequences of k orders, where each order may crossover with a maximum of m other orders

Example. The arrival sequence $as_1^{(3,1)} = (n-1, n+1, n)$ indicates that the n-th and $(n+1)$-th purchase order are involved in a crossover so that each order crosses the other one and the $(n+1)$-th purchase order arrives before the n-th purchase order.

With this notation, we are ready to develop an approach to compute the effective lead times. We will start with the case of $m = 1$, i.e., a maximum of two orders may be involved in a crossover, and then proceed to the general case.

One Order Case. In the event that each order may only cross one other order at most, we observe that only those orders may cross over that have been consecutively issued, i.e. the n-th purchase order can only cross over with either its direct predecessor or its direct successor. From the supplier's perspective, the n-th delivery may only be the $(n-1)$-th, n-th or $(n+1)$-th purchase order.

We thus have three purchase order candidates that may be the n-th delivery, and $AS^{(3,1)} = \{(n-1, n, n+1), (n, n-1, n+1), (n-1, n+1, n)\}$ is the complete set of possible arrival sequences that involve the n-th purchase order. The possible cases that lead to a certain elt of the n-th delivery can be directly derived from these sequences. To do this, we require the middle delivery to exactly meet the designated elt, while the preceding and successive deliveries' lead times need to ensure that the assumed arrival sequence is met, i.e., the preceding delivery may not arrive later and the successive delivery may not arrive earlier than the middle delivery. To save subindexes, we will strictly connect the lead time L to purchase orders and the effective lead time ELT to deliveries. Thus, the n-th delivery observes an effective lead time of elt under the following condition:

$$\begin{aligned} ELT_n = elt \quad \text{if} \quad & L_{n-1} \leq elt + r \wedge L_n = elt \wedge L_{n+1} \geq elt - r \\ \vee \quad & L_{n-1} = elt + r \wedge L_n \leq elt \wedge L_{n+1} \geq elt - r \\ \vee \quad & L_{n-1} \leq elt + r \wedge L_n \geq elt \wedge L_{n+1} = elt - r \end{aligned} \tag{5.18}$$

Note that the cases in (5.18) are not free of conjunctions, e.g., the special case of $L_{n-1} = elt + r \wedge L_n = elt \wedge L_{n+1} = elt - r$ is included in all three of them. Expanding (5.18) to comparisons with the '$<$', '$>$' and '$=$' comparative operators only leads to a conjunction-free representation (5.19), which can be reduced to (5.20) if we again allow for the '$\leq$' and '$\geq$' comparative operators:

$$\begin{aligned} ELT_n = elt \quad \text{if} \quad & L_{n-1} < elt + r \wedge L_n = elt \wedge L_{n+1} > elt - r \\ \vee \quad & L_{n-1} = elt + r \wedge L_n = elt \wedge L_{n+1} > elt - r \\ \vee \quad & L_{n-1} < elt + r \wedge L_n = elt \wedge L_{n+1} = elt - r \\ \vee \quad & L_{n-1} = elt + r \wedge L_n = elt \wedge L_{n+1} = elt - r \\ \vee \quad & L_{n-1} < elt + r \wedge L_n > elt \wedge L_{n+1} = elt - r \\ \vee \quad & L_{n-1} = elt + r \wedge L_n > elt \wedge L_{n+1} = elt - r \\ \vee \quad & L_{n-1} = elt + r \wedge L_n < elt \wedge L_{n+1} > elt - r \\ \vee \quad & L_{n-1} = elt + r \wedge L_n < elt \wedge L_{n+1} = elt - r \end{aligned} \tag{5.19}$$

$$
\begin{aligned}
ELT_n = elt \quad \text{if} \quad & L_{n-1} = elt + r \wedge L_n \leq elt \wedge L_{n+1} \geq elt - r \\
\vee \quad & L_{n-1} < elt + r \wedge L_n = elt \wedge L_{n+1} \geq elt - r \\
\vee \quad & L_{n-1} \leq elt + r \wedge L_n > elt \wedge L_{n+1} = elt - r
\end{aligned} \tag{5.20}
$$

We can derive the effective lead time distribution from the conjunction-free set of cases (5.20) if a maximum of two orders may be involved in a crossover (5.21):

$$
\begin{aligned}
P\{ELT = elt\} = {} & P\{L = elt + r\} \cdot P\{L \leq elt\} \cdot P\{L \geq elt - r\} \\
& + P\{L < elt + r\} \cdot P\{L = elt\} \cdot P\{L \geq elt - r\} \\
& + P\{L \leq elt + r\} \cdot P\{L > elt\} \cdot P\{L = elt - r\}
\end{aligned} \tag{5.21}
$$

Numerical Example. Let L be discretely uniformly distributed with values $\{1, 2, 3, 4\}$ and $r = 2$. Then the corresponding ELT is calculated as follows:

$$
\begin{aligned}
P\{ELT = 1\} &= 0.25 \cdot 0.25 \cdot 1 + 0.5 \cdot 0.25 \cdot 1 + 0.75 \cdot 0.75 \cdot 0 = 0.1875 \\
P\{ELT = 2\} &= 0.25 \cdot 0.5 \cdot 1 + 0.75 \cdot 0.25 \cdot 1 + 1 \cdot 0.5 \cdot 0 = 0.3125 \\
P\{ELT = 3\} &= 0 \cdot 0.75 \cdot 1 + 1 \cdot 0.25 \cdot 1 + 1 \cdot 0.25 \cdot 0.25 = 0.3125 \\
P\{ELT = 4\} &= 0 \cdot 1 \cdot 0.75 + 1 \cdot 0.25 \cdot 0.75 + 1 \cdot 0 \cdot 0.25 = 0.1875
\end{aligned}
$$

General Case. As a first approach to generalize the above findings, let us think of how many purchase orders may possibly arrive as the n-th delivery if we allow each order to cross m other orders at most. Let us first consider an arrival sequence in which all deliveries arrive in the same order as they were issued, except the n-th delivery, which has a relative delay. Let this delayed n-th delivery be issued p order cycles before the n-th purchase order. Then the n-th delivery is obviously overtaken by $p - 1$ other orders, and $m = p - 1$ is the maximum number of orders that are crossed by another one. Following on this situation, we observe that we may only decrease m by somehow moving the n-th delivery to an earlier position in the sequence. Obviously, the number of orders that overtake the n-th delivery can neither be changed by swapping two earlier deliveries, nor by swapping two later deliveries. Even swapping an earlier with a later delivery will not change the situation because it either holds for both deliveries that they have been issued after the n-th delivery and thus one of them still overtakes it, or, if the earlier delivery was also issued before the n-th delivery, this particular order is overtaken by $p - 1 + 1 = p$ orders at least. The latter case would thus induce $m \geq p$ because the particular order is overtaken by all orders that have overtaken the n-th delivery, and furthermore also the n-th delivery itself. The analogous argumentation holds for the case that the n-th delivery was issued p periods after the n-th purchase order. Thus, we have a maximum of $k^* = 2 \cdot m + 1$ purchase orders that may finally arrive in one particular period if no more than m order may be crossed by another one.

Regarding the one order case, we have already noticed that we may not freely permutate these k^* orders to retrieve the possible arrival sequences. For example, $AS^{(3,1)}$ may not contain $as_x = (n+1, n, n-1)$ as $as_{x,1} = n+1$ as well as $as_{x,3} = n-1$ would indicate an order that crosses with two other orders, which would be one more than is allowed. Therefore, $AS^{(k^*,m)}, m \geq 1$ will always be a set of *restricted* permutations to describe all possible arrival sequences. Thus, we call an arrival sequence $as_s^{(k^*,m)}$ *allowed* if

$$as_{s,i}^{(k^*,m)} \leq as_{s,j}^{(k^*,m)} + m \quad \forall \quad i, j \quad i < j \tag{5.22}$$

holds for the corresponding elements.

According to the condition, the i-th delivery must arrive earlier than any delivery j that has been issued more than m periods later. I.e, no issue position $as_{s,j}^{(k^*,m)}$ may be positioned left of $as_{s,i}^{(k^*,m)}$ in the arrival sequence that exceeds the issue position of i by m or more cycles. Let us consider an example here. Let $m = 3$. Then we need to regard $k^* = 7$ consecutively issued orders to cover all possible arrival sequences relative to the middle purchase order. Let us regard an index space $i \in \{-3, -2, \ldots, 3\}$ to emphasize that the sequence regarded is arranged *around* the middle purchase order indexed with 0. The sequence $as_1^{(7,3)} = (-3, -1, -2, 3, 2, 1, 0)$ is allowed as $as_i \leq as_j + 3$ holds for every index pair i, j fulfilling $i < j$, whereas the sequence $as_2^{(7,3)} = (-2, -3, \mathbf{3}, 2, -\mathbf{1}, 0, 1)$ is not allowed, as the pair $as_{2,3}^{(7,3)} = 3$, $as_{2,5}^{(7,3)} = -1$ dissatisfies the condition.

In the literature, several systematic methods are proposed to solve the problem of enumerating restricted permutations. (See Vatter 2008 for an illustrative overview.) For our application, we found the method of generating trees appropriate, both in terms of computational efficiency and confirmability of the approach for constructing the sequences. This method was introduced by Chung et al. (1978); further applications are given by West (1995), West (1996) and Merlini and Verri (2000), for example.

Given the problem of generating all sequences of k consecutive numbers that satisfy a certain set of rules, the basic idea is to iteratively derive the allowed sequences for $x \leq k$ numbers from the allowed insertions into the allowed sequences for $x - 1$ numbers. For the present problem, we start with $x = 1$ and obviously obtain only one trivial sequence $(0 - m)$ that is always allowed. In a following step x, we derive further sequences by inserting x behind the element in position $x - 1, x - 2, \ldots$ as long as the next insertion would result in a forbidden sequence. We can stop here, as all further insertions will obviously be forbidden as well. Figure 5.1 displays the generating tree of allowed sequences for $k = 3$ and $m = 1$.

Enumerating the allowed sequences for combinations of k and m, we observe rapid combinatorial growth. Table 5.1 displays the result using a straightforward Java implementation, without explicit memory management, which limits the generating method to a few hundred thousand permutations. Note that we give the number of allowed sequences for all (computable) pairs of $1 \leq k \leq 20$, $1 \leq m \leq 7$,

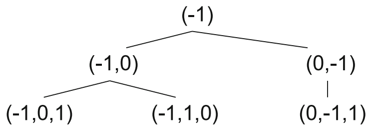

Fig. 5.1 Fully expanded generating tree for (n = 3, m = 1)

Table 5.1 Number of allowed sequences

(k,m)	1	2	3	4	5	6	7
1	1	1	1	1	1	1	1
2	2	2	2	2	2	2	2
3	**3**	6	6	6	6	6	6
4	5	12	24	24	24	24	24
5	8	**25**	60	120	120	120	120
6	13	57	150	360	720	720	720
7	21	124	**399**	1,050	2,520	5,040	5,040
8	34	268	1,145	3,192	8,400	20,160	40,320
9	55	588	3,132	**10,305**	28,728	75,600	181,440
10	89	1,285	8,420	35,505	10,3050	287,280	–
11	144	2,801	22,716	116,620	**390,555**	–	–
12	233	6,118	62,128	374,172	–	–	–
13	377	13,362	169,536	–	–	–	–
14	610	29,168	–	–	–	–	–
15	987	63,685	–	–	–	–	–
16	1,597	139,057	–	–	–	–	–
17	2,584	303,608	–	–	–	–	–
18	4,181	–	–	–	–	–	–
19	6,765	–	–	–	–	–	–
20	10,946	–	–	–	–	–	–

where in terms of the ELT we are only interested in the highlighted pairs for which $k = 2 \cdot m + 1 (= k^*)$ holds.

However, the very sequence is not exactly what we are after. Remember, we are interested in the probability that an order will arrive at a certain position in the arrival sequence. This position is already determined by the set of orders that have arrived before, regardless of the sequence in which they arrived. Calculating the probabilities, we will focus on the middle position only and ask wether an order will arrive either on the exact position, before or after it.

Example. In the event that the range of the independent lead time allows each order to crossover with a maximum of two other orders ($m = 2$), we have to extend the time window observed to five orders ($k^* = 5$). The number of allowed sequences is now 25. Equation (5.23) displays all possible arrival sequences for orders issued at order period n relative to a middle period ($n = 0$):

$$\begin{array}{ccccc}
(-1,0,-2,1,2) & (-1,0,-2,2,1) & (0,-1,-2,1,2) & (0,-1,-2,2,1) & (-2,0,-1,1,2)\\
(-2,0,-1,2,1) & (-2,1,-1,0,2) & (-2,1,-1,2,0) & (0,-2,-1,1,2) & (0,-2,-1,2,1)\\
(-2,-1,0,1,2) & (-2,-1,0,2,1) & (-1,-2,0,2,1) & (-1,-2,0,1,2) & (-2,1,0,-1,2)\\
(-2,-1,1,0,2) & (-2,-1,1,2,0) & (-2,0,1,-1,2) & (-1,-2,1,0,2) & (-1,-2,1,2,0)\\
(0,-2,1,-1,2) & (-2,-1,2,0,1) & (-2,-1,2,1,0) & (-1,-2,2,0,1) & (-1,-2,2,1,0)
\end{array} \tag{5.23}$$

Regarding the first four sequences of the first line, we note that they are equal in terms of orders arriving before and after the middle element. Thus, those four sequences are covered by one case $(\{-1,0\},-2,\{1,2\})$. Similarly, other sequences can be reduced to one case of the described form, leading to (5.24). We will refer to these cases as *partly defined sequences*.

$$\begin{array}{cccc}
(\{-1,0\},-2,\{1,2\}) & (\{-2,0\},-1,\{1,2\}) & (\{-2,1\},-1,\{0,2\}) & (\{-2,-1\},0,\{1,2\})\\
(\{-2,1\},0,\{-1,2\}) & (\{-2,-1\},1,\{0,2\}) & (\{-2,0\},1,\{-1,2\}) & (\{-2,-1\},2,\{0,1\})
\end{array} \tag{5.24}$$

Using the enumeration scheme described above, we can derive the relevant partly defined sequences from the set of allowed sequences. This approach, however, appears to be inefficient both in terms of computational effort and memory requirements. We therefore propose Algorithm 4 to directly enumerate the partly defined sequences for any given (odd) number k^*. The basic idea of the algorithm is as follows. Given $k^* = 2 \cdot m + 1$ numbers, we must consider each number as the middle element. Choosing one of these numbers determines the set of numbers that may be assigned to the left of it. Regarding (5.24), only $\{-2,-1,0\}$ may be assigned to the left of 1, for example. Thus, the algorithm comes up with two nested loops: one outer loop that defines the middle element (me), and one inner loop that chooses one possible left element (ce). To prevent duplicate partly defined sequences, it is ensured in the following that ce is the largest of all left side numbers. Therefore, only some numbers apply for ce, namely those between se and le. (See the structured exposition of the algorithm for details.)

Some numbers are definitely required on the left side of me for certain pairs of me and ce. Regarding (5.24) again, if $me = 0$ and $ce = 1$, then -2 must not be placed on the right of me. Admittedly, the example of (5.24) is too small to observe the full phenomenon, as -2 must always either be on the left, or be me. Thus imagine $m = 3, k^* = 7$, then $ce = 0, me = 1$ would require all numbers smaller than $\max\{ce, me\} - m = -2$ on the left. A base sequence BS is constructed if these required elements are added.

If all slots on the left side have already been taken in a base sequence, the corresponding pair of ce and le only results in one partly defined sequence, which can easily be completed by assigning all numbers to the right that have not been assigned so far.

Algorithm 4: Enumeration of partly defined sequences

```
Input: n
Output: All partly defined sequences PSQ of length n
m = ⌊n/2⌋;
PSQ = new SetOfPartlyDefinedSequences();
for me = 1 to n do
    if me < b then
        se = m + 1;
        le = me + m;
    else
        se = b;
        if me < n − 1 then le = n − 1;
        else if me == n − 1 then le = n − 2;
        else le = m;
    end
    for ce = se to le, ce! = me do
        BS = new PartlyDefinedSequence();
        BS.addAsME(me);              /* add me as middle element */
        BS.addAsLE(ce);            /* add me to the left elements */
        BS.addRequired(ce, me);     /* add required due to me and le */
        FE = BS.getFreeElements();    /* all elements that may but need
        not be included in the set of left elements */
        fs = BS.getFreeSlots();       /* number of undrawn slots on left
        side */
        if fs == 0 then
            BS.assignRE();    /* add disregarded numbers to the right
            side */
            PSQ.add(BS);          /* add basic set to solution set */
        else
            FS = getFreeSubsets(FE, fs);  /* get all fs-subsets of set FE
            */
            foreach Subset fss in FS do
                S = BS.clone();  /* copy elements of basic sequence */
                S.addAsLE(fss); /* add elements of chosen free subset
                */
                S.assignRE();             /* add disregarded numbers */
                PSQ.add(S);                  /* add S to solution set */
            end
        end
    end
end
```

If there are $x > 0$ slots remaining, these may freely be filled with all possible x-subsets of all numbers that may (but need not) stand left of ce and me. Again, the sequences are completed by adding all remaining unassigned numbers to the right.

Experiments suggest that the total number of different partly defined sequences is $(m + 2) \cdot 2^{m-1}$, but providing proof for this assumption remains an open task.

Let PSQ be the set of all partly defined sequences for a given m that we are now able to enumerate, using Algorithm 4. Furthermore, let psq be one member of PSQ, and psq_i the value of element i in a specific psq, i.e., the position in the order issue sequence. We can then describe the probabilities of occurrence of certain effective lead times subject to a specific psq according to (5.25) and (5.26):

$$\begin{aligned} P\{ELT = elt|PSQ = psq\} \cdot \{PSQ = psq\} \\ = \prod_{i=-m+1}^{m-1} P\{W \circ elt + psq_i \cdot r|PSQ = psq\} \end{aligned} \quad (5.25)$$

$$P\{W \circ elt + psq_i \cdot r\} = \begin{cases} P\{W \leq elt + psq_i \cdot r\} & \text{if } i < m\} \\ P\{W = elt + psq_i \cdot r\} & \text{if } i = m \\ P\{W \geq elt + psq_i \cdot r\} & \text{if } i > m\} \end{cases} \quad (5.26)$$

Numerical Example. The partly defined sequence

$$psq = \{-2, 0\}, 1, \{-1, 2\}$$

resolves to

$$\begin{aligned} &P\{ELT = elt|PSQ = psq\} \cdot \{PSQ = psq\} \\ &\quad = P\{W \leq elt - 2 \cdot r\} \cdot P\{W \leq elt\} \cdot P\{W = elt + r\} \\ &\quad \cdot P\{W \geq elt - 1 \cdot r\} \cdot P\{W \geq elt + 2 \cdot r\}. \end{aligned}$$

We are now already able to state (5.27) as upper bound for the efficient lead time probabilities:

$$P\{ELT = elt\} \leq \sum_{psq \in PSQ} P\{ELT = elt|PSQ = psq\} \cdot \{PSQ = psq\} \quad (5.27)$$

Equation (5.27) does not describe the exact probabilities, as the cases we derive from PSQ are not generally free of conjunctions, e.g., the special case of $P\{W = elt + i \cdot r\}\ \forall i$ is included in every psq. We have already observed this phenomenon regarding the special case of $m = 1$. To derive a form that is free of conjunctions, we can use the technique of expansion and reduction as introduced above. Each non-simple comparative operator (i.e., '$\leq$' and '$\geq$') has to be subdivided into two cases of simple comparative operators, leading to 2^{φ} expanded cases, where φ is the number of factors that include a non-simple comparative operator. Expression (5.27) directly implies that we derive $2^{2 \cdot m+1}$ expanded cases from each partly defined sequence.

After the reduction, the resulting cases may again be combined to expressions that allow for the non-simple comparative operators. This may be done by repeatedly combining cases that differ in only one comparison operator. To understand

Table 5.2 Encoding of comparison operators

Operator	$\leq$	$<$	$=$	$<$	$\geq$
Code	-2	-1	0	1	2

Algorithm 5: Algorithm to construct the combined cases

```
Input: Reduced cases (RED), n
Output: Combined cases (COM)
COM = RED.clone();
tempCOM = new SetOfCases();
for i = 1 to n do
    COM.excludingSort(i);          /* Sort COM disregarding position i */
    index = 1;
    while index < |COM| do
        newSequence = COM[index].clone();
        if COM[index].seq(i) · COM[index+1].seq(i) ≥ 0
        && COM[index].excluding(i) == COM[index+1].excluding(i) then
            if COM[index].seq(i) < 0 || COM[index+1].seq(i) < 0 then
                newSequence.seq(i) = -2;
            else
                newSequence.seq(i) = 2;
            end
            index = index +2;
        else
            index = index +1;
        end
        tempCOM.add(newSequence)
    end
    if index < |COM| then
        tempCOM.add(COM.lastElement);
    end
    COM = tempCOM;
end
```

the algorithm, let us introduce the following notation. A case of k^* comparisons is represented by k^* numbers x, $x \in \{-2, -1, 0, 1, 2\}$, where we encode the comparison operators according to Table 5.2.

The basic idea of the approach displayed in Algorithm 5 is to repeatedly combine two cases that meet two conditions. Firstly, both cases are equal except for one element at position i and secondly, x_i^1 and x_i^2, i.e., the elements at position i of both sequences, must either be both smaller than or equal to zero, or they must both be greater than or equal to zero. When these conditions are met, we can replace the two sequences by one, with entry -2 if both x_i^1 and x_i^2 were smaller or equal to 0 and 2 otherwise.

Table 5.3 summarizes the cardinalities of PSQ, the expanded cases (EXP), the reduced cases (RED) and finally the combined cases (COM). Where we have an indication, we denote the theoretical closed expression in the second line. Clearly, the expansion step limits the approach, especially in terms of memory.

Using the algorithm, we derive the following expressions for $m = 2, 3$ and 4. We give the expressions for $m = 2$ in encoded form (5.28) and in plain form (5.29) to

Table 5.3 Effective lead time – relevant cardinalities

m	k^*	|PSQ|	|EXP|	|RED|	|COM|
m	$2 \cdot m + 1$	$(m+2) \cdot 2^{m-1}$ (?)	$(m+2) \cdot 2^{3 \cdot m-1}$ (?)	Unknown	Unknown
1	3	3	12	8	3
2	5	8	128	60	9
3	7	20	1,280	432	26
4	9	48	12,288	3,024	73
5	11	112	114,688	20,736	201

further illustrate the encoding described above. The expressions for $m = 3$ (5.30) and $m = 4$ (5.31) are given in encoded form only.

$$\begin{array}{lllll} (-2,-2,0,1,1) & (-2,-2,2,0,1) & (-2,-2,2,2,0) & (-2,0,-1,1,2) & (-2,0,1,-1,2) \\ (-2,1,0,0,2) & (-2,2,-1,0,2) & (-2,2,0,-1,2) & (0,-1,-1,2,2) & \end{array} \tag{5.28}$$

$$\begin{aligned} &P\{ELT = elt\} \\ &= P\{L \le elt + 2 \cdot r\} \cdot P\{L \le elt + r\} \cdot P\{L = elt\} \cdot P\{L > elt - r\} \cdot P\{L > elt - 2 \cdot r\} \\ &+ P\{L \le elt + 2 \cdot r\} \cdot P\{L \le elt + r\} \cdot P\{L \ge elt\} \cdot P\{L = elt - r\} \cdot P\{L > elt - 2 \cdot r\} \\ &+ P\{L \le elt + 2 \cdot r\} \cdot P\{L \le elt + r\} \cdot P\{L \ge elt\} \cdot P\{L \ge elt - r\} \cdot P\{L = elt - 2 \cdot r\} \\ &+ P\{L \le elt + 2 \cdot r\} \cdot P\{L = elt + r\} \cdot P\{L < elt\} \cdot P\{L > elt - r\} \cdot P\{L \ge elt - 2 \cdot r\} \\ &+ P\{L \le elt + 2 \cdot r\} \cdot P\{L = elt + r\} \cdot P\{L > elt\} \cdot P\{L < elt - r\} \cdot P\{L \ge elt - 2 \cdot r\} \\ &+ P\{L \le elt + 2 \cdot r\} \cdot P\{L > elt + r\} \cdot P\{L = elt\} \cdot P\{L = elt - r\} \cdot P\{L \ge elt - 2 \cdot r\} \\ &+ P\{L \le elt + 2 \cdot r\} \cdot P\{L \ge elt + r\} \cdot P\{L < elt\} \cdot P\{L = elt - r\} \cdot P\{L \ge elt - 2 \cdot r\} \\ &+ P\{L \le elt + 2 \cdot r\} \cdot P\{L \ge elt + r\} \cdot P\{L = elt\} \cdot P\{L < elt - r\} \cdot P\{L \ge elt - 2 \cdot r\} \\ &+ P\{L = elt + 2 \cdot r\} \cdot P\{L < elt + r\} \cdot P\{L < elt\} \cdot P\{L \ge elt - r\} \cdot P\{L \ge elt - 2 \cdot r\} \end{aligned} \tag{5.29}$$

$$\begin{array}{llll} (-2,-2,-2,0,1,1,1) & (-2,-2,-2,2,0,1,1) & (-2,-2,-2,2,2,0,1) & (-2,-2,-2,2,2,2,0) \\ (-2,-2,0,-1,1,1,2) & (-2,-2,0,1,-1,1,2) & (-2,-2,0,1,1,-1,2) & (-2,-2,1,0,0,1,2) \\ (-2,-2,1,0,1,0,2) & (-2,-2,1,2,0,0,2) & (-2,-2,2,-1,0,1,2) & (-2,-2,2,-1,2,0,2) \\ (-2,-2,2,0,-1,1,2) & (-2,-2,2,0,1,-1,2) & (-2,-2,2,2,-1,0,2) & (-2,-2,2,2,0,-1,2) \\ (-2,0,-1,-1,1,2,2) & (-2,0,-1,1,-1,2,2) & (-2,0,1,-1,-1,2,2) & (-2,1,-2,0,0,2,2) \\ (-2,1,0,-1,0,2,2) & (-2,1,0,0,-1,2,2) & (-2,2,-1,-1,0,2,2) & (-2,2,-1,0,-1,2,2) \\ (-2,2,0,-1,-1,2,2) & (0,-1,-1,-1,2,2,2) & & \end{array} \tag{5.30}$$

$$
\begin{array}{lll}
(-2,-2,-2,-2,0,1,1,1,1) & (-2,-2,-2,-2,2,0,1,1,1) & (-2,-2,-2,-2,2,2,0,1,1) \\
(-2,-2,-2,-2,2,2,2,0,1) & (-2,-2,-2,-2,2,2,2,2,0) & (-2,-2,-2,0,-1,1,1,1,2) \\
(-2,-2,-2,0,1,-1,1,1,2) & (-2,-2,-2,0,1,1,-1,1,2) & (-2,-2,-2,0,1,1,1,-1,2) \\
(-2,-2,-2,1,0,0,1,1,2) & (-2,-2,-2,1,0,1,0,1,2) & (-2,-2,-2,1,0,1,1,0,2) \\
(-2,-2,-2,1,2,0,0,1,2) & (-2,-2,-2,1,2,0,1,0,2) & (-2,-2,-2,1,2,2,0,0,2) \\
 & & \\
(-2,-2,-2,2,-1,0,1,1,2) & (-2,-2,-2,2,-1,2,0,1,2) & (-2,-2,-2,2,-1,2,2,0,2) \\
(-2,-2,-2,2,0,-1,1,1,2) & (-2,-2,-2,2,0,1,-1,1,2) & (-2,-2,-2,2,0,1,1,-1,2) \\
(-2,-2,-2,2,2,-1,0,1,2) & (-2,-2,-2,2,2,-1,2,0,2) & (-2,-2,-2,2,2,0,-1,1,2) \\
(-2,-2,-2,2,2,0,1,-1,2) & (-2,-2,-2,2,2,2,-1,0,2) & (-2,-2,-2,2,2,2,0,-1,2) \\
(-2,-2,0,-1,-1,1,1,2,2) & (-2,-2,0,-1,1,-1,1,2,2) & (-2,-2,0,-1,1,1,-1,2,2) \\
(-2,-2,0,1,-1,-1,1,2,2) & (-2,-2,0,1,-1,1,-1,2,2) & (-2,-2,0,1,1,-1,-1,2,2) \\
(-2,-2,1,-2,0,0,1,2,2) & (-2,-2,1,-2,0,1,0,2,2) & (-2,-2,1,-2,2,0,0,2,2) \\
(-2,-2,1,0,-1,0,1,2,2) & (-2,-2,1,0,-1,1,0,2,2) & (-2,-2,1,0,0,-1,1,2,2) \\
(-2,-2,1,0,0,1,-1,2,2) & (-2,-2,1,0,1,-1,0,2,2) & (-2,-2,1,0,1,0,-1,2,2) \\
(-2,-2,1,1,0,0,0,2,2) & (-2,-2,1,2,-1,0,0,2,2) & (-2,-2,1,2,0,-1,0,2,2) \\
(-2,-2,1,2,0,0,-1,2,2) & (-2,-2,2,-1,-1,0,1,2,2) & (-2,-2,2,-1,-1,2,0,2,2) \\
(-2,-2,2,-1,0,-1,1,2,2) & (-2,-2,2,-1,0,1,-1,2,2) & (-2,-2,2,-1,2,-1,0,2,2) \\
(-2,-2,2,-1,2,0,-1,2,2) & (-2,-2,2,0,-1,-1,1,2,2) & (-2,-2,2,0,-1,1,-1,2,2) \\
(-2,-2,2,0,1,-1,-1,2,2) & (-2,-2,2,2,-1,-1,0,2,2) & (-2,-2,2,2,-1,0,-1,2,2) \\
(-2,-2,2,2,0,-1,-1,2,2) & (-2,0,-1,-1,-1,1,2,2,2) & (-2,0,-1,-1,1,-1,2,2,2) \\
(-2,0,-1,1,-1,-1,2,2,2) & (-2,0,1,-1,-1,-1,2,2,2) & (-2,1,-2,-2,0,0,2,2,2) \\
(-2,1,-2,0,-1,0,2,2,2) & (-2,1,-2,0,0,-1,2,2,2) & (-2,1,0,-1,-1,0,2,2,2) \\
(-2,1,0,-1,0,-1,2,2,2) & (-2,1,0,0,-1,-1,2,2,2) & (-2,2,-1,-1,-1,0,2,2,2) \\
(-2,2,-1,-1,0,-1,2,2,2) & (-2,2,-1,0,-1,-1,2,2,2) & (-2,2,0,-1,-1,-1,2,2,2) \\
(0,-1,-1,-1,-1,2,2,2,2) & &
\end{array}
\tag{5.31}
$$

Numerical Example. Let L be a discrete uniform distribution with possible values $l \in \{1, 2, 3, 4, 5\}$, furthermore let $ELT(L, r)$ describe the effective lead time of a replenishment process, where orders are issued every r-th period and replenishment lead times have the distribution of L. Table 5.4 displays the resulting effective lead time probabilities and first two moments for different parameters r.

Regarding these results, we observe a stable mean value and a declining standard deviation when lowering r. This leads us to the following two conjectures.

Conjecture 1. $E(L) = E(ELT(L, r))$ holds for all $r > 0$.

Table 5.4 Effective lead times for different parameters r

	$r \geq 4$	$r = 3$	$r = 2$	$r = 1$
$P\{ELT(L, r) = 1\}$	0.2	0.16	0.12	0.0384
$P\{ELT(L, r) = 2\}$	0.2	0.24	0.2	0.2464
$P\{ELT(L, r) = 3\}$	0.2	0.2	0.36	0.4304
$P\{ELT(L, r) = 4\}$	0.2	0.24	0.2	0.2464
$P\{ELT(L, r) = 5\}$	0.2	0.16	0.12	0.0384
μ_{ELT}	3	3	3	3
σ_{ELT}	1.414214	1.326650	1.166190	0.894427

Conjecture 2. $Var(L) \leq Var(ELT(L, r))$ holds for all $r > 0$.

Incomplete Proof. We may prove the two conjectures for the special case of $m = 1$, which means that two consecutive orders placed at period n and $n + 1$ observe either $elt_n = l_n, elt_{n+1} = l_{n+1}$ (no-crossover) or $elt_n = l_{n+1} + r, elt_{n+1} = l_n - r$ (crossover).

Conjecture 1 holds because of

$$\frac{l_{n+1} + r + l_n - r}{2} = \frac{l_n + l_{n+1}}{2},$$

i.e., the mean effective lead time of the two orders equals their underlying mean lead times, whether they cross over or not.

Considering the variance (Conjecture 2), we have to examine whether

$$l_n^2 + l_{n+1}^2 \geq (l_{n+1} + r)^2 + (l_n - r)^2$$

holds in the event that those two orders cross over. This can be shown by the following conversions, assuming $r > 0$:

$$\begin{aligned}
& l_n^2 + l_{n+1}^2 \geq (l_{n+1} + r)^2 + (l_n - r)^2 \\
\Leftrightarrow\ & l_n^2 + l_{n+1}^2 \geq l_{n+1}^2 + r \cdot l_{n+1} + r^2 + l_n^2 - r \cdot l_n + r^2 \\
\Leftrightarrow\ & 0 \geq r \cdot (l_{n+1} + r) - r \cdot (l_n + r) \\
\Leftrightarrow\ & l_n \geq l_{n+1}
\end{aligned}$$

The last line proves the conjecture for $m = 1$, because we regard the crossover case here, meaning that $l_n \geq l_{n+1}$ holds.

We leave the question open to future research as to wether the two conjectures hold for general assumptions and close with a last unproven conjecture, based on both empirical test results and the author's intuition.

Conjecture 3. $Var(ELT(L, r_1)) \leq Var(ELT(L, r_2))$ holds for all $r_1, r_2 > 0, r_1 \leq r_2$.

5.3 Sequential Arrivals

Based on an idea due to Kaplan (1970), Zipkin (1986b) introduces a general set of conditions to define a replenishment lead time process that rules out order crossover. Let $\{U(t) : t \in \mathbb{R}\}$ be a real-valued, stationary, ergodic stochastic process that satisfies the following conditions. (See Zipkin 1986b, p. 770) $U(t)$ may be interpreted as the age of the oldest order at time t.

1. $U(t) \geq 0$ and $E[U(t)] < \infty$

2. $t - U(t)$ is nondecreasing
3. Sample paths of $\{U(t)\}$ are continuous to the right
4. $\{U(t)\}$ is independent of the placement and size of orders and the demand process

$U(t)$ may be interpreted as the age of the oldest order arriving at time t. According to (1), $U(t)$ returns meaningful (non-negative) lead times with a finite expected value, i.e., the process is stationary. To understand that condition (2) rules out order crossover, consider two orders arriving in t_1 and t_2, $t_1 \leq t_2$. These two orders would cross over, if and only if $U(t_2) > U(t_1) + (t_2 - t_1)$. Reformulating the inequation to $t_1 - U(t_1) > t_2 - U(t_2)$, we clearly see that this is contradictory to (2). Condition (3) is a rather technical condition that ensures continuity of $U(t)$. Typically, we observe jump discontinuities in sample paths of $U(t)$ for those values t at which an (oldest) order arrives. Nonetheless, $U(t)$ is continuous to the right if an order is immediately removed from the stack as soon as it arrives, so that the age of the oldest order at the very point of arrival is then determined by the order that was previously the second oldest. Finally, (4) is self-explanatory.

Note that due to (2), lead times modeled in accordance with $U(t)$ are not independent in general, but form a continuous-time, continuous-state Markov process. See Ehrhardt (1984) or Nahmias (1979) for further insights into formulating lead time processes.

The recursion defined in (5.32) and (5.33) gives an example of a sampling process satisfying (1)–(4), where $S(L_n)$ is a sampling function of an arbitrarily distributed L_n, and $A(n)$ returns the issue date of the n-th order. $A(n) - A(n-1)$ is thus the time between the issue of the $n-1$-st and n-th order. For the case of an (r, S) policy, obviously $A(n) - A(n-1) = r$ holds for every pair of two consecutive orders:

$$l_0 = S(L_0) \tag{5.32}$$

$$l_n = max\{S(L_n), l_{n-1} - (A(n) - A(n-1))\} \tag{5.33}$$

Let $A(n) - A(n-1) = r \; \forall \; n$ and $(L_n^* | L_{n-1}^* = l_{n-1})$ be the lead time generating process defined by the recursion above. Then the probability distribution of the latter is given by (5.34):

$$P\{L_n^* = l_n | L_{n-1}^* = l_{n-1}\} = \begin{cases} 0 & \text{if} \quad l_n < l_{n-1} - r \\ P\{L_n \leq l_{n-1} - r\} & \text{if} \quad l_n = l_{n-1} - r \\ P\{L_n = l_n\} & \text{if} \quad l_n > l_{n-1} - r \end{cases} \tag{5.34}$$

Note that (5.34) forms a Markov chain with state transition probabilities p_{ab} given by $P\{L_n^* = a | L_{n-1}^* = b\}$. On analyzing Markov chains in general, see Meyn and Tweedie (2009), for example.

Let us assume in the following that the L_n are discrete distributions with a common finite set of integer states with probabilities p_l. Then (5.34) forms a discrete-time, discrete-state Markov chain, where steady state probabilities can easily be calculated.

Numerical Example. Let $L_1, L_2, \ldots, L_n$ be identically discretely uniformly distributed with states $\{1, 2, 3, 4\}$ and $r = 1$. Then the matrix of state transitions resolves to (5.35):

$$p_{ab} = \begin{pmatrix} 0.25 & 0.25 & 0.25 & 0.25 \\ 0.25 & 0.25 & 0.25 & 0.25 \\ 0.0 & 0.5 & 0.25 & 0.25 \\ 0.0 & 0.0 & 0.75 & 0.25 \end{pmatrix} \tag{5.35}$$

To determine the steady state probabilities, we derive a system of linear equations (5.36), where we can drop one of the first four lines:

$$\begin{array}{rrrrl} 0.25 \cdot p_1^* & +0.25 \cdot p_2^* & +0.25 \cdot p_3^* & +0.25 \cdot p_4^* & = p_1^* \\ 0.25 \cdot p_1^* & +0.25 \cdot p_2^* & +0.25 \cdot p_3^* & +0.25 \cdot p_4^* & = p_2^* \\ 0.0 \cdot p_1^* & +0.5 \cdot p_2^* & +0.25 \cdot p_3^* & +0.25 \cdot p_4^* & = p_3^* \\ 0.0 \cdot p_1^* & +0.0 \cdot p_2^* & +0.75 \cdot p_3^* & +0.25 \cdot p_4^* & = p_4^* \\ p_1^* & +p_2^* & +p_3^* & +p_4^* & = 1 \end{array}$$

The example lead time process has $p^* = \{0.09375, 0.28125, 0.375, 0.25\}$ steady state probabilities.

Alternatively to the recursion defined in (5.32) and (5.33), it may be reasonable to make use of a truncated distribution, where the probability distribution of $(L_n^* | L_{n-1}^* = l_{n-1})$ is then given by (5.36). See also Sect. 4.4.

$$P\{L_n^* = l_n | L_{n-1}^* = l_{n-1}\} = \begin{cases} 0 & \text{if} \quad l_n < l_{n-1} - r \\ \frac{P\{L_n = l_n\}}{1 - P\{L_n < l_n\}} & \text{if} \quad l_n \geq l_{n-1} - r \end{cases} \tag{5.36}$$

Here, we assume that the probability mass of the *forbidden* values l_n for a given l_{n-1} is transferred proportionally to the remaining *allowed* values. In terms of a probability experiment, this would be equivalent to repeatedly drawing from L_n and discarding forbidden values until we obtain an allowed value. We obtain the following state transition matrix for the above example:

$$p_{ab} = \begin{pmatrix} 0.25 & 0.25 & 0.25 & 0.25 \\ 0.25 & 0.25 & 0.25 & 0.25 \\ 0.0 & 0.333333 & 0.333333 & 0.333333 \\ 0.0 & 0.0 & 0.5 & 0.5 \end{pmatrix} \tag{5.37}$$

This lead time process has $p^* = \{0.0625, 0.1875, 0.375, 0.375\}$ steady state probabilities.

Remark. When lead times are dependent the proceeds for calculating the number of outstanding orders and the amount of stock outstanding as introduced in Sects. 5.2.1

and 5.2.2 are not applicable. We leave it open to future research to create appropriate methods to solve these problems.

5.4 Limited Distributions

While many authors make the assumption that lead times are independent and orders never cross, this is generally self-contradictory. (See e.g. Chen and Zheng 1992.) However, one may specify a system where both assumptions are met, an idea that is due to Hadley and Whitin (1963). Let us have a closer look at corollary 1, p. 58, to understand what we necessarily have to assume.

The corollary implies that in the case of two orders A and B, $o_A < o_B$ do not cross if and only if $o_B - o_A \leq l_A - l_B$, i.e., the difference between the two order lead times must be smaller than the difference between the order issue dates. Let R be a random variable denoting the time between issuing two consecutive orders and let R_n be the random time between two orders A and B that represent the i-th and $(n + i + 1)$-th order in the issue sequence, so that $R_n = O_B - O_A$, where the issuing times of A and B are random. Then A and B do not cross if and only if either $R_n \leq 0$ or $R_n \geq L_A - L_B$.

Excluding the rather theoretical case of a system with $R \equiv 0$, $R_n \leq 0$ is never fulfilled. $R_n \geq L_A - L_B$ is generally fulfilled if $Min\{R_n\} \geq Max\{L_A - L_B\}$. With $R \geq 0$, we have $Min\{R_n\} = Min\{R\}$, while $Max\{L_A - L_B\} = L_{span}$, $L_{span} = l_{max} - l_{min}$. In other words, in the event that the replenishment lead time span is shorter than the minimum time between two consecutive order placements, order crossover can be ruled out even with independent lead times.

For an application using limited distributions in inventory management, see Sphicas and Nasri (1984), for example.

Chapter 6
Analysis and Optimization

The preceding chapters introduced the analysis of inventory systems from a comprehensive perspective. In Chap. 2 we described the general forms of single-level inventory systems along with the basic concepts for describing the state of an inventory system, and the elementary figures for evaluating its performance. We then gave a broad literature overview in Chap. 3, and introduced some basic analytical methods in Chap. 4 with a focus on the application to periodic review models operating on a discrete time axis. In Chap. 5, we – also from a broad perspective – indicated different possibilities for incorporating lead time stochasticity into inventory models.

This chapter inevitably leaves the comprehensive scope and confines our attention to a specific class of inventory systems that we will analyze in detail. Nonetheless, we try to present our results in a manner that may help to adapt them to different systems and further problems. We particularly demonstrate how to consider different assumptions concerning the replenishment and order fulfillment process in the analysis of the same performance metrics.

The organization of this chapter is as follows. In Sect. 6.1, we define the model of the underlying single-level inventory system that we will analyze in the following. Section 6.2 contains a set of example instances that we will use to exemplify the analytical approaches. The analysis then follows in Sect. 6.3, where we distinguish between two model subspecifications. In the first place, we separately consider the two contrary lead time models described in Chap. 5, namely the lead time model that rules out order crossover (as generally described in Sect. 5.3) and the one that allows for the phenomenon (described in Sect. 5.2). Within the corresponding specifications, we will furthermore distinguish whether customer orders may be split in the event of a shortage, or wether they have to be served by one complete delivery.

Finally, in Sect. 6.4, we examine selected aspects of optimizing inventory systems based on the analytical findings.

T. Wensing, *Periodic Review Inventory Systems*, Lecture Notes in Economics and Mathematical Systems 651, DOI 10.1007/978-3-642-20479-1_6,

6.1 Model Formulation and Notation

The single-level inventory model intends to describe an inner node of a multi-level inventory system in which all nodes operate according to periodic review order-up-to (r, S) policies. Following from that, replenishment orders as well as customer orders are placed and arrive on a periodical basis, where the actual frequency depends on the order cycle (r) that induces the orders. To be more specific, customer demands arrive according to a compound renewal process $\{T_n, D_n\}_{n\geq 1}$, where $T_n \in \mathbb{Z}^+$ is a constant order interarrival rate (r_D in the following) induced by the successive stock, and D_n are continuously i.i.d. order sizes. Assuming a reorder policy with variable order sizes, the D_n comply with the successive stock's demand in r_D time units. Lead times are discretely distributed, where we examine both independent lead times with the possibility of order crossover (as described in Sect. 5.2) and the dependent lead time model (as described in Sect. 5.3). Furthermore, we will distinguish two order fulfilment modes, namely the cases of split and full deliveries. In the first case, we allow for split deliveries if stock on hand is sufficient to fulfill only a certain fraction of the volume of a customer order; in the second case, the customer order will only be delivered in full, i.e., the complete volume is delayed if one unit or more is missing.

The discrete time axis comes with the necessity of clarifying the sequence of stock-effecting events. We assume the following sequence. First, the customer order arrives, second, replenishment orders (possibly) arrive, third, the (backordered) customer orders are served if possible, and finally replenishment orders are issued if the present period is an order period.

6.2 Example Configurations

In the following we will repeatedly solve model instances to illustrate the analytical formulae. For a clearer representation, we will define those model instances in this section, and later on only refer to the instance number. Table 6.1 displays the example instances, where $Norm(\mu, \sigma)$ denotes a normal distribution with mean μ and standard deviation σ, and $Unif(\{x, y, \ldots\})$ denotes a discrete uniform distribution with possible values $x, y, \ldots$, i.e., $Unif(\{1, 4\})$ in Table 6.1 means that L is either 1 or 4, both with 50% probability.

6.3 Analysis

The organization of this section is driven by the analytical requirements of the characteristics that we want to examine. In the first place, we distinguish whether the underlying lead time model allows for order crossover or not. For both cases, we will

Table 6.1 Example configurations

No.	r	S	r_D	D	L
1	2	300	1	Norm(100,30)	Unif({1,2})
2	4	500	1	Norm(100,30)	Unif({1,2})
3	4	500	2	Norm(100,30)	Unif({1,2})
4	2	80	1	Unif({10,20,50,100})	Unif({1,2})
5	4	160	1	Unif({10,20,50,100})	Unif({1,2})
6	4	160	2	Unif({10,20,50,100})	Unif({1,2})
7	2	300	1	Norm(100,30)	Unif({1,4})
8	4	650	1	Norm(100,30)	Unif({1,8})
9	4	650	2	Norm(100,30)	Unif({1,8})
10	2	80	1	Unif({10,20,50,100})	Unif({1,4})
11	4	160	1	Unif({10,20,50,100})	Unif({1,8})
12	4	160	2	Unif({10,20,50,100})	Unif({1,8})

Table 6.2 Problems and assumptions

	Dependent lead times	Order crossover
Order view	Sect. 6.3.1	Sect. 6.3.4
Volume view		
Split deliveries	Sect. 6.3.2	Sect. 6.3.5
Full deliveries	Sect. 6.3.3	Sect. 6.3.6

then first be concerned with the characteristics that focus on customer orders, i.e., the distribution of waiting times per order, and then consider the volume-oriented characteristics such as fill rate or average inventory levels. For the latter group we furthermore distinguish whether customer orders may be split or have to be delivered in full. Table 6.2 summarizes the different combinations of problems and assumptions, and shows where to find the corresponding evaluation approaches.

In the following chapters, we will not specially mark the metrics considered within the case being observed, i.e., we will, for example, always plainly identify the fill rate as β without adding any sub- or superscript to indicate that we actually mean the fill rate for a special delivery mode here. We hope that the reader is not confused by the fact that the same symbol is then associated with different equations depending on the case that we assume to be set by the context of the chapter. We think that the alternative use of sub- or superscripts to cleanly distinguish the cases would reduce the readability of the formulae.

6.3.1 Dependent Lead Times, Order View

6.3.1.1 Ready Rate

[*Case:* $\tau = r_D$.] In the event that the lead time model does not allow for order crossover, the ready rate per period may be determined according to the following considerations. Let us initially assume that the lead time l is a constant, and a replenishment order arrives after l periods. We can then obviously state that the stock on hand is sufficient to fulfill the demand in this period if and only if the demand in l periods has not been higher than S. Remember that by definition of the (r, S) policy, each replenishment order immediately raises the inventory position to S. Thus, when the order arrives after l periods, the net inventory levels are at S minus the amount that has been demanded between order issue and arrival. We can furthermore state that the stock on hand is sufficient to fulfill the demand of the next $(r-1)$ periods if the demand in $l+1, l+2, \ldots, l+r-1$ periods was not higher than S. We can keep track of this scheme until the next replenishment order arrives in period $r+l$, where we again observe the same conditions as for period l:

Let the demand rate $r_D = 1$, then (6.1) states the ready rate per customer order arrival, where $T^* = \{0, 1, \ldots, r-1\}$ is the set of subperiods of an order cycle.

$$\alpha^{\tau=1}|L = l = \sum_{t^* \in T^*} \frac{1}{r} \cdot P\{D^{[l+t^*]} \leq S\} \tag{6.1}$$

While (6.1) considers a specific lead time, the overall ready rate per customer order is stated by (6.2), where $P\{L = l\}$ is the (steady state) probability for an order to arrive after l periods. Note that the above argumentation still holds as long as we do not allow for order crossover. The argumentation must then be read in the opposite direction, i.e., there is no stockout in the period just before the replenishment order arrival if the demand in $r+l-1$ periods has not exceeded S.

$$\alpha^{\tau=1} = \sum_{t^* \in T^*} \sum_{l \in L} \frac{P\{L = l\}}{r} \cdot P\{D^{[l+t^*]} \leq S\} \tag{6.2}$$

For an arbitrary customer order rate r_D and an unchanged period demand, we have to consider that the frequency of demand observation changes from $\frac{1}{r}$ to $\frac{r_D}{r}$. To reflect this circumstance, let $T^{**} = \{0, 1, \ldots, \frac{r}{r_D} - 1\}$ be the set of consecutive demand occurrences within an order cycle. Furthermore, we have to alter the amount of demand that has occurred until the corresponding periods after the arrival of a replenishment order. For a clearer representation, let us define an adaptation function here for the latter purpose.

Definition 20 (First adaptation function). Let $nf_1(a, b)$ denote the *first adaptation function*, given as follows:

$$nf_1(a, b) := \left\lfloor \frac{a}{b} \right\rfloor \cdot b$$

Using $nf_1(a, b)$, we can state the ready rate per customer order as follows.

$$\alpha^{\tau=r_D} = \sum_{t^{**} \in T^{**}} \sum_{l \in L} \frac{r_D \cdot P\{L = l\}}{r} \cdot P\{D^{[nf_1(l+r_D \cdot t^{**}, r_D)]} \leq S\} \tag{6.3}$$

Note that (6.4) is equal to (6.3), but results in avoidable multiple calculation of the same values if $r_D > 1$:

$$\alpha^{\tau=r_D} = \sum_{t^* \in T^*} \sum_{l \in L} \frac{P\{L=l\}}{r} \cdot P\{D^{[nf_1(l+t^*, r_D)]} \le S\} \tag{6.4}$$

Numerical Example. Considering Example 1 of Table 6.1 (p. 83), the ready rate per customer order is calculated as follows:

$$\begin{aligned}
\alpha^{\tau=1} &= \frac{P\{L=1\}}{2} \cdot P\{D^{[1]} \le 300\} + \frac{P\{L=2\}}{2} \cdot P\{D^{[2]} \le 300\} \\
&\quad + \frac{P\{L=1\}}{2} \cdot P\{D^{[2]} \le 300\} + \frac{P\{L=2\}}{2} \cdot P\{D^{[3]} \le 300\} \\
&= \frac{1}{4} \cdot CDF_{D^{[1]}}(300) + \frac{1}{2} \cdot CDF_{D^{[2]}}(300) + \frac{1}{4} \cdot CDF_{D^{[3]}}(300) \\
&= 0.25 \cdot 1.0 + 0.5 \cdot 0.990789 + 0.25 \cdot 0.5 \\
&= 0.870394
\end{aligned}$$

[*Case:* $\tau = r$.] The ready rate per replenishment cycle can be calculated similarly. Again, let us first consider an order arrival after a certain lead time l. We will not observe a stockout in the corresponding cycle if S is sufficiently large to cover the demand of $l + r - 1$ periods. In other words, we need to cover the replenishment lead time plus one cycle, where we must not consider the last period due to the chosen event sequence. (The demand of the order period itself will immediately be considered with the new replenishment order.) We thus denote (6.5) for a demand rate $r_D = 1$:

$$(\alpha^{\tau=r} | L = l) = P\{D^{[l+r-1]} \le S\} \tag{6.5}$$

For an arbitrary r_D and discretely distributed replenishment lead times L, we obtain (6.6):

$$\alpha^{\tau=r} = \sum_{L \in l} P\{L=l\} \cdot P\{D^{[nf_1(l+r-1, r_D)]} \le S\} \tag{6.6}$$

Numerical Example. Considering Example 1 of Table 6.1 (p. 83), the ready rate per order cycle is calculated as follows:

$$\begin{aligned}
\alpha^{\tau=r} &= P\{L=1\} \cdot P\{D^{[2]} \le 300\} + P\{L=2\} \cdot P\{D^{[3]} \le 300\} \\
&= 0.5 \cdot 0.990780 + 0.5 \cdot 0.5 \\
&= 0.745394
\end{aligned}$$

In Sect. 2.3.2.1 we have already discussed the problems that we might observe when empirically measuring the ready rate per order cycle under some special assumptions. Although (6.6) exactly indicates the ratio of replenishment orders that will be connected with problems in customer order fulfilment, it is hardly possible to track this effect on a real or simulated inventory system if two or more orders could be outstanding at the same time.

6.3.1.2 Customer Waiting Time per Order

For a start, let us again assume that the lead time is fixed to a value l and customer orders arrive in every period ($r_D = 1$). Let us furthermore regard one period t^* of the order cycle, and assume $S \geq 0$. We can then state that the maximum lead time for any order arriving in this period is $r+l-t^*$, as the corresponding customer order will definitely be taken into account with the next replenishment order, which will be issued in $r-t^*$ periods and will arrive l periods later. The corresponding waiting time is realized if the demand of t^* periods exceeds S, as otherwise the customer order could have been satisfied with the material of some earlier replenishment order. Equation (6.7) thus denotes the probability that a customer order will observe the maximum possible lead time in periods $T^* = \{0, 1, \ldots, r-1\}$:

$$P(W^O = r + l - t^* | L = l, T^* = t^*, r_D = 1, S \geq 0) = P\{D^{[t^*]} > S\} \quad (6.7)$$

Dropping the assumption $S \geq 0$ obviously leads to (6.8), where a customer arriving in period t^* will now observe a waiting time of $r + l - t^*$ *or longer*:

$$P(W^O \geq r + l - t^* | L = l, T^* = t^*, r_D = 1) = P\{D^{[t^*]} > S\} \quad (6.8)$$

Let us now replace t^* by $t^\Delta \in \mathbb{Z}, t^\Delta < r + l$, where t^Δ denotes a period with distance t^Δ versus a specific order period. Formula (6.8) still holds, but we have to consider that t^Δ will match a certain period of the order cycle. Let us define a second adaptation function to identify these matching periods.

Definition 21 (Second adaptation function). Let $nf_2(a, b)$ denote the *second adaptation function*, given as follows:

$$nf_2(a, b) = \left(\frac{a}{b} - \left\lfloor \frac{a}{b} \right\rfloor\right) \cdot b$$

Using $nf_2(a, b)$, we can state the period-specific probabilities as follows:

$$P(W^O \geq r + l - t^\Delta | L = l, T = nf_2(t^\Delta, r), r_D = 1) = P\{D^{[t^\Delta]} > S\} \quad (6.9)$$

Note that even from (6.9) we can still deduce the fundamental argument that leads us to (6.8): In the event that the demand in period t^Δ exceeds S, the replenishment order issued $r + l - t^{\Delta}$ periods ago is not sufficient to cover the demand of period t^{Δ}, which would have resulted in a lead time of $l - t^{\Delta}$ or lower. Instead, we have to wait for the next replenishment order at least, which was released r periods later. This would result in a waiting time of $r + l - t^{\Delta}$.

Now that we allow for negative order-up-to levels S, we need to specify $D^{[t^{\Delta}]}$ for $t^{\Delta} \leq 0$:

$$D^{[0]} \equiv 0 \tag{6.10}$$

$$D^{[t^{\Delta}]} = -D^{[-t^{\Delta}]} \qquad \forall\, t^{\Delta} < 0 \tag{6.11}$$

$$P\{D^{[t^{\Delta}]} > S\} = P\{-D^{[-t^{\Delta}]} \leq -S\} \quad \forall\, t^{\Delta} < 0 \tag{6.12}$$

For the order period $t^{\Delta} = 0$ we have $P\{D^{[0]} > S\} = 0$ if S is positive, or 0 and $P\{D^{[0]} > S\} = 1$ if S is negative. In the first case, the backorder amount including the demand in period $t^{\Delta} = 0$ is definitely considered in full with the replenishment order issued in $t^{\Delta} = 0$, and therefore the waiting time cannot possibly be longer than $r + l - 1$. In the latter case of $S < 0$, the demand in period $t^{\Delta} = 0$ will definitely not be considered in full with the corresponding replenishment order, and thus the demand has a waiting time of $r + l$ or longer. The calculus for $t^{\Delta} < 0$ is somewhat the opposite of the positive case if $S < 0$. Here, a certain backorder level is needed to ensure that a period's demand is considered with the next replenishment order. Namely the demand of period $t^{\Delta} < 0$ will only be considered, if the demand of periods $t^{\Delta} + 1, t^{\Delta} + 2, \ldots, 0$ exceeds $-S$.

By substitution of $w := r + l - t^{\Delta}$ and $t^{\Delta} = r + l - w, w > 0$, we obtain the general probability that an order arriving in a corresponding period of the order cycle has a waiting time of w or longer:

$$\begin{aligned} &P\{W^O \geq w | L = l, T = nf_2(r + l - w, r), w > 0, r_D = 1\} \\ &\quad = P\{D^{[r+l-w]} > S\} \end{aligned} \tag{6.13}$$

Now let us drop the assumption that customer orders arrive in every period. Obviously, we observe the same waiting time probability as for $r_D = 1$ if there is demand in the corresponding period. In the event that there is no demand, we define the probability as zero with the argument that there may be no order that could cause a waiting time.

$$\begin{aligned} &P\{W^O \geq w | L = l, T = nf_2(r + l - w, r), w > 0\} \\ &\quad = \begin{cases} P\{D^{[r+l-w]} > S\} & \text{if} \quad nf_2(r + l - w, r_D) = 0 \\ 0 & \text{if} \quad nf_2(r + l - w, r_D) \neq 0 \end{cases} \end{aligned} \tag{6.14}$$

From (6.14), we can derive the probability that the waiting time *equals* w by subtracting the probability that the waiting time is equal to or greater than the next possible greater lead time. Remember that we always consider a certain period of the order cycle. The next possible waiting time greater than w is therefore $w + r$.

$$\begin{aligned} &P\{W^O = w|L = l, T = nf_2(r + l - w, r), w > 0\} \\ &\quad = P\{W^O \geq w|L = l, T = nf_2(r + l - w, r), w > 0\} \\ &\quad\quad - P\{W^O \geq w + r|L = l, T = nf_2(l - w, r), w > 0\} \end{aligned} \tag{6.15}$$

With an arbitrary arrival rate r_D, we receive (6.16) and (6.17), respectively, for customer orders arriving at random periods of the order cycle. Note that $nf_2(l - w, r_D) = nf_2(r + l - w, r_D)$ always holds because $nf_2(r, r_D) = 0$ always holds due to the model assumptions.

For a clearer representation, let us define a truth function on $nf_2(a, b)$.

Definition 22 (Truth function on $nf_2(a, b)$). Let $NF_2(a, b)$ denote a truth function on $nf_2(a, b)$, given as follows:

$$NF_2(a, b) = \begin{cases} 1 & \text{if} \quad nf_2(a, b) = 0 \\ 0 & \text{if} \quad nf_2(a, b) \neq 0 \end{cases}$$

Using $NF_2(a, b)$, we can denote the lead time conditioned waiting times as follows.

$$P\{W^O = w|L = l, w > 0\}$$

$$= \begin{cases} \frac{r_D}{r} \cdot (P\{D^{[r+l-w]} > S\} - P\{D^{[l-w]} > S\}) & \text{if} \quad nf_2(l - w, r_D) = 0 \\ 0 & \text{if} \quad nf_2(l - w, r_D) \neq 0 \end{cases} \tag{6.16}$$

$$= \frac{r_D}{r} \cdot (P\{D^{[r+l-w]} > S\} - P\{D^{[l-w]} > S\}) \cdot NF_2(l - w, r_D) \tag{6.17}$$

From (6.17), we can easily derive the formula to determine $P\{W^O = 0\}$:

$$\begin{aligned} &P\{W^O = 0|L = l\} = 1 - \sum_{w=1}^{\infty} P\{W^O = w|L = l, w > 0\} \\ &= 1 - \frac{r_D}{r} \cdot \left(\sum_{w=1}^{r} (P\{D^{[r+l-w]} > S\} - P\{D^{[l-w]} > S\}) \cdot NF_2(l - w, r_D) \right. \\ &\quad \left. + \sum_{w=1}^{\infty} (P\{D^{[l-w]} > S\} - P\{D^{[l-w-r]} > S\}) \cdot NF_2(l - w, r_D) \right) \\ &= 1 - \frac{r_D}{r} \cdot \sum_{w=1}^{r} P\{D^{[r+l-w]} > S\} \cdot NF_2(l - w, r_D) \end{aligned} \tag{6.18}$$

At this point, let us bring to mind the connection between the waiting time per order and the ready rate per order. It is easy to show that (6.18) is equal to the lead time conditioned ready rate per order (6.4). The basic idea is to substitute t^* by $r-w$, where $r - w \in \{0, \ldots, r-1\} \Rightarrow -w \in \{0-r, \ldots, r-1-r\} \Rightarrow w \in \{r, \ldots, 1\}$ leads to formula (6.18), as is shown in the following. Note that we can drop function $nf_1(\cdot,\cdot)$ by introducing $NF_2(\cdot,\cdot)$:

$$\begin{aligned}
&\sum_{t^* \in T^*} \frac{1}{r} \cdot P\{D^{[l+t^*]} \leq S\} \cdot NF_2(l+t^*, r_D) \cdot r_D \\
&\quad = \sum_{(r-w) \in T^*} \frac{r_D}{r} \cdot P\{D^{[l+r-w]} \leq S\} \cdot NF_2(l+r-w, r_D) \\
&\quad = \sum_{w=1}^{r} \frac{r_D}{r} \cdot P\{D^{[l+r-w]} \leq S\} \cdot NF_2(l+r-w, r_D) \\
&\quad = \sum_{w=1}^{r} \frac{r_D}{r} \cdot (1 - P\{D^{[l+r-w]} > S\}) \cdot NF_2(l+r-w, r_D) \\
&\quad = \frac{r_D}{r} \cdot \sum_{w=1}^{r} NF_2(l+r-w, r_D) - P\{D^{[l+r-w]} > S\} \cdot NF_2(l+r-w, r_D) \\
&\quad = \frac{r_D}{r} \cdot \frac{r}{r_D} - \frac{r_D}{r} \cdot \sum_{w=1}^{r} P\{D^{[l+r-w]} > S\} \cdot NF_2(l+r-w, r_D) \\
&\quad = 1 - \frac{r_D}{r} \cdot \sum_{w=1}^{r} P\{D^{[l+r-w]} > S\} \cdot NF_2(l+r-w, r_D) \ (q.e.d.)
\end{aligned}$$

Finally, (6.19) and (6.20) denote the overall probabilities for certain customer waiting times:

$$P\{W^O = 0\} = \sum_{L \in l} P^*\{L = l\} \cdot P\{W^O = 0 | L = l\} \tag{6.19}$$

$$P\{W^O = w | w > 0\} = \sum_{l \in L} P^*\{L = l\} \cdot P\{W^O = w | L = l, w > 0\} \tag{6.20}$$

We have already noted that the maximum lead time is $r + l_{max} - 1$ if $S \geq 0$. In the event that $S < 0$, the waiting time is theoretically unbound for continuous demand distributions if demands *near* zero are likely to be observed. For a practical application, we therefore need to define some maximum lead time w_{max} that we want to consider. For this purpose, we propose (6.21), where w_{max} is the minimum value w for which the probability that the according demand does not exceed $-S$ is smaller than a certain small number ϵ. To avoid confusion, note that $-S$ is a positive number in the case of $S < 0$.

$$w_{max} = \min\{w : P\{D^{[-(r+l_{max}-w)]}(x) \leq -S\} \leq \epsilon\} \tag{6.21}$$

Numerical Example. Considering Example 1 of Table 6.1 (p. 83), the customer waiting time per order

$$\begin{aligned}
P\{W^O = 0\} &= P\{L = 1\} \cdot \left(1 - \frac{1}{2} \cdot (P\{D^{[1]} > 300\} + P\{D^{[2]} > 300\})\right) \\
&\quad + P\{L = 2\} \cdot \left(1 - \frac{1}{2} \cdot (P\{D^{[2]} > 300\} + P\{D^{[3]} > 300\})\right) \\
&= 0.5 \cdot (1 - 0.5 \cdot (0.0 + 0.009211)) \\
&\quad + 0.5 \cdot (1 - 0.5 \cdot (0.009211 + 0.5)) \\
&= 0.870394 \\
P\{W^O = 1\} &= P\{L = 1\} \cdot \frac{1}{2} \cdot (P\{D^{[2]} > 300\} - P\{D^{[0]} > 300\}) \\
&\quad + P\{L = 2\} \cdot \frac{1}{2} \cdot (P\{D^{[3]} > 300\} - P\{D^{[1]} > 300\}) \\
&= 0.5 \cdot 0.5 \cdot (0.009211 - 0.0) + 0.5 \cdot 0.5 \cdot (0.5 - 0.0) \\
&= 0.127303 \\
P\{W^O = 2\} &= P\{L = 1\} \cdot \frac{1}{2} \cdot (P\{D^{[1]} > 300\} - P\{D^{[-1]} > 300\}) \\
&\quad + P\{L = 2\} \cdot \frac{1}{2} \cdot (P\{D^{[2]} > 300\} - P\{D^{[0]} > 300\}) \\
&= 0.5 \cdot 0.5 \cdot (0.0 - 0.0) + 0.5 \cdot 0.5 \cdot (0.009211 - 0.0) \\
&= 0.002303
\end{aligned}$$

$$P\{W^O > 2\} \approx 0$$

6.3.1.3 Example Results

Table 6.3 summarizes the solutions to the example instances 1–6 of Table 6.1 for the ready rate per customer order and the customer waiting time per order.

6.3.2 Dependent Lead Times, Volume View, Split Deliveries

6.3.2.1 Mean Backorder

The mean backorder amount is typically calculated for two purposes. In the classical cost-based approaches, it is used to quantify backorder costs, while in

Table 6.3 Example instances 1–6 – order-related metrics

No.	$\alpha^{\tau=r_D}$	$P\{W^O = w\}$			
		$w = 0$	$w = 1$	$w = 2$	$w > 2$
1	0.8704	0.8704	0.1273	0.0023	0.0
2	0.9255	0.9255	0.0685	0.0060	0.0
3	0.9761	0.9761	0.0119	0.0119	0.0
4	0.5039	0.5039	0.2461	0.1875	0.0625
5	0.6708	0.6708	0.1671	0.1094	0.0527
6	0.6797	0.6797	0.1445	0.1445	0.0313

the later service-constraint approaches, it is needed to calculate the two fill-rate characteristics that we discussed in Sects. 2.3.2.2 and 2.3.2.3.

Let us start our considerations with a demand arrival rate $r_D = 1$ and $S \geq 0$, where we will drop the first but not the second assumption in the proceeding. Let us furthermore consider the special case that a replenishment order was issued in period 0, where it has raised the inventory position to S, and arrives in period t. The backorder level in period t then equals the fraction of demand in t periods that has exceeded S. It can be determined according to (6.22), where we assume that the demand per period D is a continuous random variable:

$$E[B_t|L = l, r_D = 1] = \int_S^\infty PDF_{D[t]}(x) \cdot (x - S)\, dx \tag{6.22}$$

Equation (6.22), however, holds not only for the very period in which a replenishment order arrives, but for r consecutive periods until the arrival of the next order. Thus, we may state equation (6.23):

$$E[B_t|L = l, l \leq t < l + r, r_D = 1] = \int_S^\infty PDF_{D[t]}(x) \cdot (x - S) \cdot dx \tag{6.23}$$

Let us furthermore drop the assumption that t is bound to the left. We then have (6.24), which we can no longer interpret as backorder amount in period t, but as an amount that will be on backorder for $l + r - t$ periods at least. We will need this formulation later on to determine the amount that is added to the backorder amount, taking into account the demand of a certain time span.

$$E[B_t|L = l, t < l + r, r_D = 1] = \int_S^\infty PDF_{D[t]}(x) \cdot (x - S)\, dx \tag{6.24}$$

For an arbitrary r_D, we make use of the normalizing function $(nf_1(a,b))$ as defined in Definition 20 on p. 84:

$$E[B_t|L = l, t < l + r] = \int_S^\infty PDF_{D[nf_1(t,r_D)]}(x) \cdot (x - S)\, dx \tag{6.25}$$

Considering the periods of the ordering cycle, we have to determine the mean of r periods from period l after order issue (in which the replenishment order arrives) to $l+r-1$ (which is the period just before the next order arrives). This leads to (6.26):

$$E[B|L=l] = \sum_{t=l}^{l+r-1} \frac{1}{r} \cdot E[B_t|L=l, l \le t < l+r] \\ = \frac{1}{r} \sum_{t^* \in T^*} E[B_{l+t^*}] \tag{6.26}$$

We finally obtain the mean amount on backorder per period for the entire system according to (6.27), where P denotes the steady state probabilities of the lead time model used:

$$E[B] = \sum_{l \in L} P\{L=l\} \cdot E[B|L=l] \tag{6.27}$$

The total mean backorder amount, however, is not exactly what we are after if we want to determine β, for example. Here, we need to know the mean fraction of demand that is *added* to the backorder amount in each period during an order cycle or, generally, in a certain time span. It is tempting to consider this as the change of backorder from one period to the other, but we will see that this conception can be misleading when we examine the crossover case. Instead, we should regard it as the backorder amount minus the backorder amount in the very time span if it were not for the time span's demand. Thus, we introduce $B_{[a,b]}, a \le b < l+r$ as the backorder that originates from time interval $[a+1, b]$. Note that the above B_t then equals $B_{[0,t]}$.

Proceeding from (6.25), we derive (6.28):

$$E[B_{[a,b]}|L=l, a \le b < l+r] \\ = E[B_b|L=l, b < l+r] - E[B_a|L=l, a < l+r] \\ = \int_S^\infty PDF_{D[nf_1(b,r_D)]}(x) \cdot (x-S)\, dx \\ - \int_S^\infty PDF_{D[nf_1(a,r_D)]}(x) \cdot (x-S)\, dx \tag{6.28}$$

We thus derive equation (6.29) for the mean *additional* backorder per period ($E[B^{*,\tau=1}]$), and (6.30) for the mean *additional* backorder per order cycle ($E[B^{*,\tau=r}]$). Note that the selection of a and b in the formulae below always fulfills $a \le b < l+r$, so that we can drop the corresponding conditions of the general formula (6.28).

$$
\begin{aligned}
E[B^{*,\tau=1}] &= \sum_{t^* \in T^*} \sum_{l \in L} \frac{P\{L=l\}}{r} \cdot E[B_{[t^*+l-1,t^*+l]} | L \le l] \\
&= \sum_{l \in L} \frac{P\{L=l\}}{r} \cdot E[B_{[l-1,r+l-1]} | L \le l] \qquad (6.29)\\
E[B^{*,\tau=r}] &= \sum_{l \in L} P\{L=l\} \cdot E[B_{[l-1,r+l-1]} | L \le l] \qquad (6.30)
\end{aligned}
$$

Numerical Example. Considering Example 1 of Table 6.1 (p. 83), the backorder amounts are calculated as follows. See Sect. 4.3 for the determination of the mass integral of normal distributions.

$$
\begin{aligned}
E[B] &= P\{L=1\} \cdot \frac{1}{2} \cdot (E[B_1] + E[B_2]) \\
&\quad P\{L=2\} \cdot \frac{1}{2} \cdot (E[B_2] + E[B_3]) \\
E[B_1] &= \int_S^\infty PDF_{D^{[1]}}(x) \cdot (x-300)\, dx \\
&= 30^2 \cdot PDF_{D^{[1]}}(300) + (100-300) \cdot (1 - CDF_{D^{[1]}}(300)) \\
&= 30^2 \cdot 0 + (-200) \cdot 0 \\
&= 0 \\
E[B_2] &= 2 \cdot 30^2 \cdot PDF_{D^{[2]}}(300) + (200-300) \cdot (1 - CDF_{D^{[2]}}(300)) \\
&= 2 \cdot 30^2 \cdot 0.0005847 + (-100) \cdot (1 - 0.990789) \\
&= 0.131274 \\
E[B_3] &= 20.729649 \\
E[B] &= 0.25 \cdot 0 + 0.5 \cdot 0.131274 + 0.25 \cdot 20.729649 \\
&= 5.248049 \\
E[B^{*,\tau=1}] &= P\{L=1\} \cdot \frac{1}{2} \cdot E[B_{[0,2]}] + P\{L=2\} \cdot \frac{1}{2} \cdot E[B_{[1,3]}] \\
&= P\{L=1\} \cdot \frac{1}{2} \cdot (E[B_2] - E[B_0]) \\
&\quad + P\{L=2\} \cdot \frac{1}{2} \cdot (E[B_3] - E[B_1]) \\
&= 0.25 \cdot (0.131274 - 0) + 0.25 \cdot (20.729649 - 0) \\
&= 5.215231
\end{aligned}
$$

6.3.2.2 Mean Inventory

For a start, let us again assume that we have a fixed lead time l and the customer orders arrive at rate $r_D = 1$. Let $t, l \leq t < l + r$ be the t-th period after the order was issued that arrives after l periods. Then the stock level in t is S minus the demand in t periods if the demand was lower than S, and 0 otherwise. We state:

$$\begin{aligned} &E[I_t|L = l, l \leq t \leq l + r - 1, r_D = 1] \\ &\qquad = \int_{-\infty}^{\infty} PDF_{D[t]}(x) \cdot (S - \text{Min}\{x, S\})\, dx \end{aligned} \tag{6.31}$$

Dropping the assumption of $r_D = 1$ leads to (6.32):

$$E[I_t|L = l, l \leq t < l + r] = \int_{-\infty}^{\infty} PDF_{D[nf_1(t,r_D)]}(x) \cdot (S - \text{Min}\{x, S\})\, dx \tag{6.32}$$

This can be simplified as follows:

$$\begin{aligned} &E[I_t|L = l, l \leq t < l + r] \\ &\quad = \int_{-\infty}^{\infty} PDF_{D^{nf_1(t,r_D)}}(x) \cdot (S - \text{Min}\{x, S\})\, dx \\ &\quad = \int_{-\infty}^{\infty} PDF_{D[nf_1(t,r_D)]}(x) \cdot S - \int_{-\infty}^{\infty} PDF_{D[nf_1(t,r_D)]}(x) \cdot \text{Min}\{x, S\}\, dx \\ &\quad = S - \int_{-\infty}^{S} PDF_{D[nf_1(t,r_D)]}(x) \cdot x\, dx - \int_{S}^{\infty} PDF_{D[nf_1(t,r_D)]}(x) \cdot S\, dx \\ &\quad = S - \int_{-\infty}^{S} PDF_{D[nf_1(t,r_D)]}(x) \cdot x\, dx - (1 - CDF_{D[nf_1(t,r_D)]}(S)) \cdot S \end{aligned} \tag{6.33}$$

For the overall system with $t := l + t^*$, we obtain (6.34). Here, $t^* = 0$ denotes the period in which the replenishment order arrives.

$$E[I] = \sum_{t^* \in T^*} \sum_{l \in L} \frac{P\{L = l\}}{r} \cdot E[I_{t^*+l}|L = l] \tag{6.34}$$

Note that the mean stock levels can also be determined on the basis of the mean demand and mean backorder amount:

$$\begin{aligned} &E[I_t|L = l, l \leq t < l + r] \\ &\quad = S - \int_{-\infty}^{S} PDF_{D[nf_1(t,r_D)]}(x) \cdot x\, dx - \int_{S}^{\infty} PDF_{D[nf_1(t,r_D)]}(x) \cdot S\, dx \end{aligned}$$

$$
\begin{aligned}
&= S - \Bigg[\int_{-\infty}^{S} PDF_{D^{[nf_1(t,r_D)]}}(x) \cdot x \, dx \\
&\quad + \int_{S}^{\infty} PDF_{D^{[nf_1(t,r_D)]}}(x) \cdot (S - x + x) \, dx \Bigg] \\
&= S - \Bigg[\int_{-\infty}^{S} PDF_{D^{[nf_1(t,r_D)]}}(x) \cdot x \, dx + \int_{S}^{\infty} PDF_{D^{[nf_1(t,r_D)]}}(x) \cdot x \, dx \\
&\quad - \int_{S}^{\infty} PDF_{D^{[nf_1(t,r_D)]}}(x) \cdot (x - S) \, dx \Bigg] \\
&= S - \Bigg[\int_{-\infty}^{\infty} PDF_{D^{[nf_1(t,r_D)]}}(x) \cdot x \, dx \\
&\quad - \int_{S}^{\infty} PDF_{D^{[nf_1(t,r_D)]}}(x) \cdot (x - S) \, dx \Bigg] \\
&= S - E[D^{[nf_1(t,r_D)]}] + E[B_t | L = l, l \le t < l + r]
\end{aligned} \tag{6.35}
$$

Numerical Example. Considering Example 1 of Table 6.1 (p. 83), the mean inventory levels are calculated as follows:

$$
\begin{aligned}
E[I] &= P\{L = 1\} \cdot \frac{1}{2} \cdot E[I_1 | L = 1] + P\{L = 2\} \cdot \frac{1}{2} \cdot E[I_2 | L = 2] \\
&\quad + P\{L = 1\} \cdot \frac{1}{2} \cdot E[I_2 | L = 1] + P\{L = 2\} \cdot \frac{1}{2} \cdot E[I_3 | L = 2] \\
&= 0.25 \cdot (300 - E[D_{[1]}] + E[B_1]) + 0.5 \cdot (300 - E[D_{[2]}] + E[B_2]) \\
&\quad + 0.25 \cdot (300 - E[D_{[3]}] + E[B_3]) \\
&= 0.25 \cdot (200 + 0) + 0.5 \cdot (100 + 0.131227) + 0.25 \cdot (0 + 20.729649) \\
&= 105.248049
\end{aligned}
$$

6.3.2.3 Fill Rate

On the basis of the backorder amount determined in Sect. 6.3.2.1, we can state equations (6.36) and (6.37) to compute the classic fill rate and time-weighted fill rate:

$$
\beta = 1 - \frac{E[B^*]}{E[D]} \tag{6.36}
$$

$$
\gamma = 1 - \frac{E[B]}{E[D]} \tag{6.37}
$$

Numerical Example. Considering Example 1 of Table 6.1 (p. 83), we have the following fill-rates:

$$\beta = 1 - 5.215231 \div 100 = 0.947848$$
$$\gamma = 1 - 5.248049 \div 100 = 0.947520$$

6.3.2.4 Customer Waiting Time per Part

In Sect. 6.3.1.2, we considered the customer waiting time per order, i.e., the time that passes until the final part of an order can be fulfilled. In this section, we will consider the waiting time of a single unit, where customer orders may be split and thus two different units of the same order may have different waiting times. The necessary calculations are similar to those for the fill rate. Analogously, we will apply the frequentist interpretation of probabilities, i.e., if x of y parts have to wait for a certain time, we say that the probability to wait is $\frac{x}{y}$ per unit.

Let us again start our considerations with a fixed lead time l and a customer order arrival rate $r_D = 1$. A demand unit will then have to wait for w periods if (1) it belongs to the backorder amount fraction that exceeds S after $r + l - w$ periods, and (2) S has not already been exceeded by demand that arrived r periods before. Thus, we need to determine the amount of *coverable* new backorder (B_w^{**}) that arises $r + l - w$ periods after an order was placed, where we have to keep in mind that a particular waiting time may only occur if the corresponding customer demand arrives in a certain period of the order cycle:

$$\begin{aligned}
&E[B_w^{**}|W^V > 0, L = l, r_D = 1] \\
&\quad = \left[\int_S^\infty PDF_{D[r+l-w]}(x) \cdot (x - S)\, dx \right. \\
&\quad\quad \left. - \int_S^\infty PDF_{D[r+l-w-1]}(x) \cdot (x - S)\, dx \right] \\
&\quad\quad - \left[\int_S^\infty PDF_{D[l-w]}(x) \cdot (x - S)\, dx \right. \\
&\quad\quad \left. - \int_S^\infty PDF_{D[l-w-1]}(x) \cdot (x - S)\, dx \right]
\end{aligned} \tag{6.38}$$

From (6.38) we can directly derive the fraction of the period demand that will have to wait for $w > 0$ periods:

$$P\{W^V = w|w > 0, L = l, r_D = 1\} = \frac{1}{r} \cdot \frac{E[B_w^{**}]}{E[D]} \tag{6.39}$$

We obtain (6.40) and (6.41) for an arbitrary r_D:

$$\begin{aligned}
&E[B_w^{**}|W^V > 0, L = l] \\
&\quad = \left[\int_S^\infty PDF_{D[nf_1(r+l-w,r_D)]}(x) \cdot (x - S)\, dx \right. \\
&\qquad \left. - \int_S^\infty PDF_{D[nf_1(r+l-w-1,r_D)]}(x) \cdot (x - S)\, dx \right] \\
&\qquad - \left[\int_S^\infty PDF_{D[nf_1(l-w,r_D)]}(x) \cdot (x - S)\, dx \right. \\
&\qquad \left. - \int_S^\infty PDF_{D[nf_1(l-w-1,r_D)]}(x) \cdot (x - S)\, dx \right]
\end{aligned} \tag{6.40}$$

$$\begin{aligned}
P\{W^V = w|w > 0, L = l\} &= \frac{r_D}{r} \cdot \frac{E[B_w^{**}|W^V > 0, L = l]}{r_D \cdot E[D]} \\
&= \frac{1}{r} \cdot \frac{E[B_w^{**}|W^V > 0, L = l]}{E[D]}
\end{aligned} \tag{6.41}$$

The probability of $w = 0$ can be derived from (6.41), where we make use of the fact that $\sum_{w=1}^{\infty} P\{W^V = w|w > 0, L = l\}$ contains two telescoping series:

$$\begin{aligned}
&P\{W^V = 0|L = l\} \\
&\quad = 1 - \sum_{w=1}^{\infty} P\{W^V = w|w > 0, L = l\} \\
&\quad = 1 - \frac{1}{r \cdot E[D]} \cdot \sum_{w=1}^{\infty} \left[\left[\int_S^\infty PDF_{D[nf_1(r+l-w,r_D)]}(x) \cdot (x - S)\, dx \right. \right. \\
&\qquad \left. - \int_S^\infty PDF_{D[nf_1(r+l-w-1,r_D)]}(x) \cdot (x - S)\, dx \right] \\
&\qquad - \left[\int_S^\infty PDF_{D[nf_1(l-w,r_D)]}(x) \cdot (x - S)\, dx \right. \\
&\qquad \left. \left. - \int_S^\infty PDF_{D[nf_1(l-w-1,r_D)]}(x) \cdot (x - S)\, dx \right] \right] \\
&\quad = 1 - \frac{1}{r \cdot E[D]} \cdot \left[\sum_{w=1}^{r} \left[\int_S^\infty PDF_{D[nf_1(r+l-w,r_D)]}(x) \cdot (x - S)\, dx \right. \right.
\end{aligned}$$

$$
\begin{aligned}
& - \int_S^\infty PDF_{D[nf_1(r+l-w-1,r_D)]}(x) \cdot (x-S)\,dx \\
& - \int_S^\infty PDF_{D[nf_1(l-w,r_D)]}(x) \cdot (x-S)\,dx \\
& + \int_S^\infty PDF_{D[nf_1(l-w-1,r_D)]}(x) \cdot (x-S)\,dx \Bigg] \\
& + \sum_{w=1}^{\infty} \Bigg[\int_S^\infty PDF_{D[nf_1(l-w,r_D)]}(x) \cdot (x-S)\,dx \\
& - \int_S^\infty PDF_{D[nf_1(l-w-1,r_D)]}(x) \cdot (x-S)\,dx \\
& - \int_S^\infty PDF_{D[nf_1(l-w-r,r_D)]}(x) \cdot (x-S)\,dx \\
& + \int_S^\infty PDF_{D[nf_1(l-w-r-1,r_D)]}(x) \cdot (x-S)\,dx \Bigg] \\
= {} & 1 - \frac{1}{r \cdot E[D]} \cdot \sum_{w=1}^{r} \Bigg[\int_S^\infty PDF_{D[nf_1(r+l-w,r_D)]}(x) \cdot (x-S)\,dx \\
& - \int_S^\infty PDF_{D[nf_1(r+l-w-1,r_D)]}(x) \cdot (x-S)\,dx \Bigg] \\
= {} & 1 - \frac{1}{r \cdot E[D]} \cdot \Bigg[\int_S^\infty PDF_{D[nf_1(r+l-1,r_D)]}(x) \cdot (x-S)\,dx \\
& - \int_S^\infty PDF_{D[nf_1(r+l-2,r_D)]}(x) \cdot (x-S)\,dx \\
& + \int_S^\infty PDF_{D[nf_1(r+l-2,r_D)]}(x) \cdot (x-S)\,dx \\
& - \int_S^\infty PDF_{D[nf_1(r+l-3,r_D)]}(x) \cdot (x-S)\,dx + \quad \ldots \\
& + \int_S^\infty PDF_{D[nf_1(r+l-r,r_D)]}(x) \cdot (x-S)\,dx \\
& - \int_S^\infty PDF_{D[nf_1(r+l-r-1,r_D)]}(x) \cdot (x-S)\,dx \Bigg]
\end{aligned}
$$

$$= 1 - \frac{1}{r \cdot E[D]} \cdot \left[\int_S^\infty PDF_{D^{[nf_1(r+l-1,r_D)]}}(x) \cdot (x - S)\, dx \right.$$
$$\left. - \int_S^\infty PDF_{D^{[nf_1(l-1,r_D)]}}(x) \cdot (x - S)\, dx \right] \qquad (6.42)$$

As previously demonstrated, the overall probability of positive customer waiting time can be calculated using (6.43):

$$P\{W^V = w\} = \sum_{l \in L} P\{L = l\} \cdot P\{W^V = w | L = l\} \qquad (6.43)$$

Fischer (2008) (p.118) gives analogous formulae for $r_D = 1$, where the author classifies them as approximations. As the motivation of these findings remains rather unclear, it is difficult to understand the reasoning behind this conception.

Numerical Example. Considering Example 1 of Table 6.1 (p. 83), the waiting time probabilities per part can be calculated as follows:

$$P\{W^V = 0\}$$
$$= P\{L = 1\} \cdot \left[1 - \frac{1}{2 \cdot 100} \cdot \left(\int_{300}^\infty PDF_{D^{[2]}(x)} \cdot (x - 300)\, dx \right.\right.$$
$$\left.\left. - \int_{300}^\infty PDF_{D^{[0]}(x)} \cdot (x - 300)\, dx \right) \right]$$
$$+ P\{L = 2\} \cdot \left[1 - \frac{1}{2 \cdot 100} \cdot \left(\int_{300}^\infty PDF_{D^{[3]}(x)} \cdot (x - 300)\, dx \right.\right.$$
$$\left.\left. - \int_{300}^\infty PDF_{D^{[1]}(x)} \cdot (x - 300)\, dx \right) \right]$$
$$= 0.5 \cdot (1 - 0.005 \cdot (E[B_2] - E[B_0]))$$
$$+ 0.5 \cdot (1 - 0.005 \cdot (E[B_3] - E[B_1]))$$
$$= 0.5 \cdot (1 - 0.005 \cdot 0.131274) + 0.5 \cdot (1 - 0.005 \cdot 20.729649)$$
$$= 0.947847$$

$$P\{W^V = 1\}$$
$$= P\{L = 1\} \cdot \frac{1}{200} \cdot \Big[(E[B_2] - E[B_1]) - (E[B_0] - E[B_{-1}]) \Big]$$
$$+ P\{L = 2\} \cdot \frac{1}{200} \cdot \Big[(E[B_3] - E[B_2]) - (E[B_1] - E[B_0]) \Big]$$

$$= 0.5 \cdot 0.005 \cdot \Big[(0.131274 - 0) - (0 - 0)\Big]$$
$$+ 0.5 \cdot 0.005 \cdot \Big[(20.729649 - 0.131274) - (0 - 0)\Big]$$
$$= 0.0518241$$
$$P\{W^V = 2\}$$
$$= P\{L = 1\} \cdot \frac{1}{200} \cdot \Big[(E[B_1] - E[B_0]) - (E[B_{-1}] - E[B_{-2}])\Big]$$
$$+ P\{L = 2\} \cdot \frac{1}{200} \cdot \Big[(E[B_2] - E[B_1]) - (E[B_0] - E[B_{-1}])\Big]$$
$$= 0.000328$$
$$P\{W^V > 2\} \approx 0$$

6.3.2.5 Example Results

Table 6.4 summarizes the solutions to the example instances 1–6 of Table 6.1 for the fill rates and customer waiting time per part.

6.3.3 Dependent Lead Times, Volume View, Full Deliveries

6.3.3.1 Mean Backorder

If we do not permit split deliveries, the full period demand will be backordered as soon as one single unit is missing. This complicates the analytical steps, as we will see in the following.

For a start, let us again assume a customer demand rate $r_D = 1$. To get an idea of the necessary analytical steps, let us imagine a stock system with a positive amount of x units of physical stock when a new customer order of d units arrives. After

Table 6.4 Example instances 1–6 – volume-related metrics (split deliveries)

	Service levels		$P\{W^V = w\}$			
No.	β	γ	$w = 0$	$w = 1$	$w = 2$	$w > 2$
1	0.9478	0.9475	0.9478	0.0518	0.0003	0.0
2	0.9651	0.9636	0.9651	0.0335	0.0015	0.0
3	0.997	0.8654	0.997	0.0015	0.0015	0.0
4	0.5495	0.3481	0.5495	0.2769	0.1458	0.0278
5	0.6886	0.4864	0.6886	0.1629	0.1016	0.0469
6	0.783	0.2313	0.783	0.1016	0.1016	0.0139

considering the customer demand, the backorder amount is 0 in the event that d was smaller than or equal to x and d otherwise. If the system is already out of stock, the backorder amount will certainly rise by d. To derive the mean backorder amount in some period of the order cycle, we therefore need to consider three aspects: (1) the probability that the system is out of stock even without considering the current period's demand, (2) distribution of the preceding demand if this demand has not exceeded S, and (3) the demand distribution of the period observed.

We already developed a solution closely related to problem (1) when we examined the ready rate in Sect. 6.3.1.1. At this point, however, we need some more detailed information on the relationship between S and the demand than is given in the ready rate. Let us therefore define a truth function SO_t on the exhaustion of S by the demand of t periods.

Definition 23 (Truth function on the exhaustion of S based on periods t). Let SO_t be a truth function that is defined as follows:

$$SO_t = 0 \text{ if the demand in } t \text{ periods does not exceed } S$$

$$SO_t = 1 \text{ if the demand in } t \text{ periods does exceed } S$$

Equations (6.44) and (6.45) obviously state the probability distribution of SO_t if we observe demand in every period:

$$P\{SO_t = 0 | r_D = 1\} = CDF_{D^{[t]}}(S) \tag{6.44}$$

$$P\{SO_t = 1 | r_D = 1\} = 1 - CDF_{D^{[t]}}(S) \tag{6.45}$$

In the following we will be particularly interested in the distributions of $SO_{l-1}, \ldots,$ SO_{l+r-2} that characterize the *carryover* of stockout probabilities in the r relative periods of the order cycle.

To develop an approach for problems (2) and (3), let us first consider the special case of $L = 1$ and $t = 0$. The previous demand is 0 and we can state the mean new backorder amount according to (6.46):

$$E[B_t^* | L = 1, t^* = 0, r_D = 1]$$
$$= \int_S^\infty PDF_D(x) \cdot x \, dx \tag{6.46}$$

If the period demand exceeds S, we observe the full demand as a new backorder amount. If we have to account for previous demand, we have to distinguish two cases. The previous demand may either have or have not already exceeded S. In the first case, the full demand of the period being observed will be backordered. To handle the second case, we have to reflect the information that the demand has not exceeded S by adjusting the distribution of the previous demand. Namely, this demand equals the truncation (see Sect. 4.4) of the correspondingly convolved period demand within the limits of $-\infty$ (or rather 0 if we assume that demand is always positive) and S:

$$D_t|(SO_t = 0, r_D = 1) = D^{[t]}_{(-\infty,S)} \tag{6.47}$$

Unfortunately $D_t|(SO_t = 0)$ does not provide us with a stable foundation as we had in the above case in the absence of any previous demand. Here, we have to regard all possible combinations of the truncated previous demand and the period's demand that would result in a stockout in the period observed:

$$\begin{aligned} &E[B^*_t|SO_{t-1} = 0, r_D = 1] \\ &= \int_S^{\infty} \Big(\int_{-\infty}^{\infty} PDF_{D^{[t-1]}_{(-\infty,S)}}(v-x) \cdot PDF_D(x) \cdot x \; dv \Big) \, dx \end{aligned} \tag{6.48}$$

Considering that the system may already have been out of stock in the previous period, we derive equation (6.49):

$$\begin{aligned} E[B^*_t|r_D = 1] = P\{(S_{t-1}|r_D = 1) = 1\} \cdot E[D] \\ + P\{(S_{t-1}|r_D = 1) = 0\} \cdot E[B^*_t|SO_{t-1} = 0] \end{aligned} \tag{6.49}$$

Before proceeding, let us introduce a third adaptation function here that helps us to determine the new demand in t.

Definition 24 (Third adaptation function). Let $nf_3(a,b)$ denote the *third adaptation function*, given as follows:

$$\begin{aligned} nf_3(a,b) = \left\lfloor \frac{a}{b} \right\rfloor \cdot b - \left\lfloor \frac{a-b}{b} \right\rfloor \cdot b \\ (=)\; nf_1(a,b) - nf_1(a-1,b) \end{aligned}$$

Using $nf_3(a,b)$, we obtain the following equations for an arbitrary customer demand rate r_D:

$$P\{SO_t = 0\} = CDF_{D^{[nf_1(t,r_D)]}}(S) \tag{6.50}$$

$$P\{SO_t = 1\} = 1 - CDF_{D^{[nf_1(t,r_D)]}}(S) \tag{6.51}$$

$$\begin{aligned} E[B^*_t|SO_{t-1} = 0] = \int_S^{\infty} \Big(\int_{-\infty}^{\infty} PDF_{D^{[nf_1(t-1,r_D)]}_{(-\infty,S)}}(v-x) \\ \cdot PDF_{D^{[nf_3(t,r_D)]}}(x) x \; dv \Big) \, dx \end{aligned} \tag{6.52}$$

$$\begin{aligned} E[B^*_t] = P\{(SO_{t-1}) = 1\} \cdot E[D^{[nf_3(t,r_D)]}] \\ + P\{(SO_{t-1}) = 0\} \cdot E[B^*_t|SO_{t-1} = 0] \end{aligned} \tag{6.53}$$

The overall mean new backorder amount per period can then be determined by (6.54):

$$E[B^*] = \sum_{t^* \in T^*} \sum_{l \in L} \frac{1}{r} \cdot P\{L = l\} \cdot E[B^*_{t^*+l}] \tag{6.54}$$

Let us now turn to the mean backorder amount. Since we already know about the mean new backorder amount, we should be able to determine the mean backorder amount by somehow summing up the new backorders. On the basis of (6.50)–(6.53) we may calculate the mean backorder amount according to (6.55) and (6.56), where we assume $S \geq 0$:

$$E[B|T^* = t^*, L = l] = \sum_{\lambda=1}^{l+t^*} E[B^*_\lambda] \tag{6.55}$$

$$E[B] = \sum_{t^* \in T^*} \sum_{l \in L} \frac{1}{r} \cdot P\{L = l\} \cdot E[B|T^* = t^*, L = l] \tag{6.56}$$

Note that we can obviously determine the mean backorder amount in any interval $[a, b]$ by changing the range of summation in (6.55).

In general, however, there is little hope of resolving (6.52) to a closed form when we assume a certain continuous distribution for the demand per period. In most if not all cases it will thus be inevitable to use some method of numerical integration.

Alternatively, we may discretize the demand distribution D, where we have to transform equation (6.52) according to (6.57):

$$\begin{aligned} &E[B^*_t|SO_{t-1} = 0] \\ &= \sum_{x=S+1}^{\infty} \sum_{v=-\infty}^{\infty} PDF_{D^{[nf_1(t-1,r_D)]}_{(-\infty,S)}}(v - x) \cdot PDF_{D^{[nf_3(t,r_D)]}}(x) \cdot x \end{aligned} \tag{6.57}$$

We may then apply Algorithm 6 to conduct the necessary calculations to evaluate (6.57), where method getNTimeConvolution(X, n) returns convolutions of n random variables having the distribution of X, and method getTruncation(X, a, b) returns the truncation of distribution X within the limits of a and b.

Numerical Example. Considering Example 4 of Table 6.1 (p. 83), the backorder amounts can be calculated as follows:

$$\begin{aligned} E[B^*] &= \frac{1}{2} \cdot P\{L = 1\} \cdot E[B^*_1] + \frac{1}{2} \cdot P\{L = 2\} \cdot E[B^*_2] \\ &= \frac{1}{2} \cdot P\{L = 1\} \cdot E[B^*_2] + \frac{1}{2} \cdot P\{L = 2\} \cdot E[B^*_3] \end{aligned}$$

Algorithm 6: Calculation of mean new backorder t periods after an order was issued (6.57)

```
Input: Discrete distribution D, Integer l, Integer t, Integer S, Integer r_D
Output: Mean new backorder per period t (mnb)
dlt0 = floor((l + t − 1)/r_D) · r_D;
dlt1 = floor((l + t)/r_D) - floor((l + t − 1)/r_D)) · r_D;
if dlt1 == 0 then
    mnb = 0;
else if dlt0 == 0 then
    thisD = getNTimeConvolution(D, dlt1);
    mnb = 0;
    for x = S + 1 to thisD.getMax() do
        mnb += thisD.getPDF(x) ·x ;
    end
else
    prevD = getNTimeConvolution(D, dlt0);
    prevD = getTruncation(prevD, −∞, S);
    thisD = getNTimeConvolution(D, dlt1);
    oLB = thisD.getMin() + prevD.getMin();
    oLB = getMin(oLB, S+1);
    oUB = thisD.getMax() + prevD.getMax();
    mnb = 0;
    for v = oLB to oUB do
        iLB = getMax(thisD.getMin(), v-prevD.getMax());
        iUB = getMin(thisD.getMax(), v-prevD.getMin());
        for x = iLB to iUB do
            mnb += prevD.getPDF(v − x) ·thisD.getPDF(x) ·x ;
        end
    end
end
```

$$
\begin{aligned}
E[B_1^*] &= (1 - CDF_{D^{[0]}}(80)) \cdot E[D^{[1]}] + CDF_{D^{[0]}}(80) \\
&\quad \cdot \int_{80}^{\infty} \left(\int_{-\infty}^{\infty} PDF_{D^{[0]}_{(-\infty,80)}}(v-x) \cdot PDF_{D^{[1]}}(x)\, dv \right) dx \\
&= 0 \cdot 45 + 1 \cdot 100 \cdot 0.25 \\
&= 25 \\
E[B_2^*] &= (1 - CDF_{D^{[1]}}(80)) \cdot E[D^{[1]}] + CDF_{D^{[1]}}(80) \\
&\quad \cdot \int_{80}^{\infty} \left(\int_{-\infty}^{\infty} PDF_{D^{[1]}_{(-\infty,80)}}(v-x) \cdot PDF_{D^{[1]}}(x)\, dv \right) dx \\
&= 0.25 \cdot 45 + 0.75 \cdot 29.166667 \\
&= 33.125 \\
E[B_3^*] &= 0.5 \cdot 45 + 0.5 \cdot 34.0625 \\
&= 39.53125
\end{aligned}
$$

$$
\begin{aligned}
E[B^*] &= 0.25 \cdot 25 + 0.5 \cdot 33.125 + 0.25 \cdot 39.53125 \\
&= 32.695313
\end{aligned}
$$

$$
\begin{aligned}
E[B] = \frac{1}{2} \cdot P\{L = 1\} \cdot \sum_{\lambda=1}^{1} E[B_\lambda^*] + \frac{1}{2} \cdot P\{L = 2\} \cdot \sum_{\lambda=1}^{2} E[B_\lambda^*] \\
\frac{1}{2} \cdot P\{L = 1\} \cdot \sum_{\lambda=1}^{2} E[B_\lambda^*] + \frac{1}{2} \cdot P\{L = 2\} \cdot \sum_{\lambda=1}^{3} E[B_\lambda^*] \\
= 0.25 \cdot 25 + 0.5 \cdot (25 + 33.125) + 0.25 \cdot (25 + 33.125 + 39.53125) \\
= 59.726563
\end{aligned}
$$

6.3.3.2 Mean Inventory

In the event that only full deliveries are allowed, we observe problems that are very similar to those that we had to solve to determine the overall backorder amount in Sect. 6.3.3.1. We have to regard each relevant period rather than being allowed to apply some overall consideration. In this section we will therefore first consider the change in inventory levels after considering the demand of the t-th period ($I_t^{(-)}$), and derive the overall value afterwards.

Let us again use the truth function SO_t as defined in Definition 23. Depending on the value of SO_{t-1} in the previous period, we can state two things. If $SO_{t-1} = 1$, there will be no additional reduction of inventory, as S is already exceeded by previous orders. If $SO_{t-1} = 0$, there will be a reduction to a certain extent if the new demand is lower than what is left of S, and no reduction otherwise. The mean of $I_t^{(-)}$ can thus be calculated according to (6.58):

$$
\begin{aligned}
E[I_t^{(-)}] = P\{SO_{t-1} = 1\} \cdot 0 \\
+ P\{SO_{t-1} = 0\} \cdot \int_{-\infty}^{S} \Big(\int_{-\infty}^{\infty} PDF_{D_{(-\infty,S)}^{[nf_1(t,r_D)]}}(v - x) \\
\cdot PDF_{D^{[nf_3(t,r_D)]}}(x)\, dv \Big)\, dx
\end{aligned}
\tag{6.58}
$$

As for the split deliveries case, the mean reduction of stock levels can also be determined on the basis of the mean demand and mean backorder amount:

$$
\begin{aligned}
E[I_t^{(-)} | SO_{t-1} = 0] \\
= \int_{-\infty}^{S} \Big(\int_{-\infty}^{\infty} PDF_{D_{(-\infty,S)}^{[nf_1(t-1,r_D)]}}(v - x) \cdot PDF_{D^{[nf_3(t,r_D)]}}(x)\, dv \Big)\, dx
\end{aligned}
$$

$$
\begin{aligned}
&= \int_{-\infty}^{\infty} \Big(\int_{-\infty}^{\infty} PDF_{D^{[nf_1(t-1,r_D)]}_{(-\infty,S)}}(v-x) \cdot PDF_{D^{[nf_3(t,r_D)]}}(x)\; dv \Big)\; dx \\
&\quad - \int_{S}^{\infty} \Big(\int_{-\infty}^{\infty} PDF_{D^{[nf_1(t-1,r_D)]}_{(-\infty,S)}}(v-x) \cdot PDF_{D^{[nf_3(t,r_D)]}}(x)\; dv \Big)\; dx \\
&= \int_{-\infty}^{\infty} PDF_{D^{[nf_3(t,r_D)]}}(x)\; dx \\
&\quad - \int_{S}^{\infty} \Big(\int_{-\infty}^{\infty} PDF_{D^{[nf_1(t-1,r_D)]}_{(-\infty,S)}}(v-x) \cdot PDF_{D^{[nf_3(t,r_D)]}}(x)\; dv \Big)\; dx \\
&= E[D^{[nf_3(t,r_D)]}] - E[B_t^*|SO_{t-1} = 0]
\end{aligned}
\tag{6.59}
$$

To understand the third transformation, note that

$$\int_{-\infty}^{\infty} PDF_A(x-v) \cdot PDF_B(x) \cdot x\; dv = PDF_B(x) \cdot x$$

holds for any pair of two random variables A and B because of

$$\int_{-\infty}^{\infty} PDF_A(x-v)\; dv = \int_{-\infty}^{\infty} PDF_A(v)\; dv = 1.$$

Finally, we obtain (6.60) to calculate the mean inventory for the overall system:

$$E[I] = \sum_{t^* \in T^*} \sum_{l \in L} \frac{P\{L = l\}}{r} \cdot \left(S - \sum_{\lambda=1}^{t^*+l} E[I_\lambda^{(-)}] \right) \tag{6.60}$$

Numerical Example. Considering Example 4 of Table 6.1 (p. 83), the mean physical inventory levels can be calculated as follows:

$$
\begin{aligned}
E[I] = &\frac{1}{2} \cdot P\{L = 1\} \cdot \left(S - \sum_{\lambda=1}^{1} E[I_\lambda^{(-)}] \right) \\
&+ \frac{1}{2} \cdot P\{L = 2\} \cdot \left(S - \sum_{\lambda=1}^{2} E[I_\lambda^{(-)}] \right) \\
= &\frac{1}{2} \cdot P\{L = 1\} \cdot \left(S - \sum_{\lambda=1}^{2} E[I_\lambda^{(-)}] \right) \\
&+ \frac{1}{2} \cdot P\{L = 2\} \cdot \left(S - \sum_{\lambda=1}^{3} E[I_\lambda^{(-)}] \right)
\end{aligned}
$$

$$
\begin{aligned}
E[I_1^{(-)}] &= P\{SO_0 = 0\} \cdot \left[E[D^{[1]}] - E[B_1^*|SO_0 = 0]\right] \\
&= 1.0 \cdot (45 - 25) \\
&= 20 \\
E[I_2^{(-)}] &= P\{SO_1 = 0\} \cdot \left[E[D^{[2]}] - E[B_2^*|SO_1 = 0]\right] \\
&= 0.75 \cdot (45 - 29.166667) \\
&= 11.875000 \\
E[I_3^{(-)}] &= P\{SO_2 = 0\} \cdot \left[E[D^{[3]}] - E[B_3^*|SO_2 = 0]\right] \\
&= 0.5 \cdot (45 - 34.0625) \\
&= 5.46875 \\
E[I] &= 0.25 \cdot (80 - 20) + 0.5 \cdot (80 - 20 - 11.875) \\
&\quad + 0.25 \cdot (80 - 20 - 11.875 - 5.46875) \\
&= 49.726563
\end{aligned}
$$

6.3.3.3 Fill Rate

Using (6.55) and (6.56), we can determine the fill rates according to (6.36) and (6.37).

Numerical Example. Example 4 of Table 6.1 (p. 83) results in the following values:

$$
\begin{aligned}
\beta &= 1 - 32.695313 \div 45 = 0.273438 \\
\gamma &= 1 - 59.726563 \div 45 = -0.327257
\end{aligned}
$$

6.3.3.4 Customer Waiting Time per Part

To determine the customer waiting time per part in the case of full deliveries, we can make use of (6.53). In Sect. 6.3.2.4, we introduced the term *coverable* new backorder. Considering full deliveries only, the mean coverable new backorder is equal to the full new backorder in a certain period ($E[B^*_{r+l-w}]$) minus the fraction of this new backorder that has to wait longer than w periods ($E[B^*_{l-w}]$). Thus, we obtain (6.61):

$$
E[B_w^{**}|L = l] = E[B^*_{nf_1(r+l-w,r_D)}] - E[B^*_{nf_1(l-w,r_D)}] \tag{6.61}
$$

Using the adjusted $E[B_w^{**}|L = l]$, equation (6.41) also applies for the case of full deliveries only. The probability $P\{W^V = 0|L = l\}$ can be derived analogously to (6.42):

$$P\{W^V = 0|L = l\} = 1 - \frac{1}{r \cdot E[D]} \cdot \left[\sum_{\lambda=l}^{r+l-1} E[B^*_{nf_1(\lambda,r_D)}] \right] \tag{6.62}$$

Finally, (6.43) applies to determine the non-conditioned lead time probabilities.

Numerical Example. Considering Example 4 of Table 6.1 (p. 83), customer waiting time per part is calculated as follows:

$$\begin{aligned}
P\{W^V = 0\} &= P\{L = 1\} \cdot \Big(1 - \frac{1}{2 \cdot 45} \cdot \sum_{\lambda=1}^{2} E[B^*_\lambda]\Big) \\
&\quad + P\{L = 2\} \cdot \Big(1 - \frac{1}{2 \cdot 45} \cdot \sum_{\lambda=2}^{3} E[B^*_\lambda]\Big) \\
&= 0.5 \cdot (1 - \frac{1}{90} \cdot (25 + 33.125)) + 0.5 \cdot (1 - \frac{1}{90} \cdot (33.125 + 39.53125) \\
&= 0.2734375 \\
P\{W^V = 1\} &= \frac{1}{2} \cdot P\{L = 1\} \cdot \frac{E[B^{**}_1|L = 1]}{E[D]} \\
&\quad + \frac{1}{2} \cdot P\{L = 2\} \cdot \frac{E[B^{**}_2|L = 1]}{E[D]} \\
&= \frac{1}{2} \cdot P\{L = 1\} \cdot \frac{E[B^*_2] - E[B^*_0]}{E[D]} + \frac{1}{2} \cdot P\{L = 2\} \cdot \frac{E[B^*_3] - E[B^*_1]}{E[D]} \\
&= 0.25 \cdot \frac{33.125 - 0}{45} + 0.25 \cdot \frac{39.53125 - 25}{45} \\
&= 0.264757 \\
P\{W^V = 2\} &= \frac{1}{2} \cdot P\{L = 1\} \cdot \frac{E[B^{**}_2|L = 1]}{E[D]} \\
&\quad + \frac{1}{2} \cdot P\{L = 2\} \cdot \frac{E[B^{**}_2|L = 1]}{E[D]} \\
&= \frac{1}{2} \cdot P\{L = 1\} \cdot \frac{E[B^*_1] - E[B^*_{-1}]}{E[D]} + \frac{1}{2} \cdot P\{L = 2\} \cdot \frac{E[B^*_2] - E[B^*_0]}{E[D]} \\
&= 0.25 \cdot \frac{25}{45} + 0.25 \cdot \frac{33.125}{45} \\
&= 0.322917 \\
P\{W^V = 3\} &= 0.25 \cdot 0 + 0.25 \cdot \frac{25}{45} \\
&= 0.138889 \\
P\{W^V > 3\} &= 0
\end{aligned}$$

Table 6.5 Example instances 1–6 – volume-related metrics (full deliveries)

	Service levels		$P\{W^V = w\}$				
No.	β	γ	$w = 0$	$w = 1$	$w = 2$	$w = 3$	$w > 4$
1	0.8516	0.8482	0.8516	0.1449	0.0035	0.0	0.0
2	0.9157	0.908	0.9157	0.0766	0.0077	0.0	0.0
3	0.9692	0.9538	0.9692	0.0154	0.0154	0.0	0.0
4	0.2734	−0.3273	0.2734	0.2648	0.3229	0.1389	0.0
5	0.56	0.2099	0.56	0.1994	0.1487	0.0747	0.0174
6	0.5651	0.2088	0.5651	0.1827	0.1827	0.0347	0.0347

6.3.3.5 Example Results

Table 6.5 summarizes the solutions to the example instances 1–6 of Table 6.1 for the fill rates and customer waiting time per part. To analyze the first three examples, we used discrete transformations of the corresponding normal distributions to compute the convolution of truncated and non-truncated distributions.

6.3.4 Order Crossover, Order View

We have already demonstrated that it is possible to handle the phenomenon of order crossover to a certain extend in Sect. 5.2. Within the computational limits mentioned, we may well determine the effective lead time, for example. However, the effective lead time provides us with insufficient information about the development of stock levels on behalf of the demand side, as we treat replenishment orders as interchangeable. I.e., the case of a large order arriving early in place of a smaller order will be treated equally to the opposite case. We observe the same effect for the distribution of outstanding orders that we will use in the following to derive elementary service levels as well as the mean backorder and mean inventory levels.

The following approaches are clearly exact if orders are in fact interchangeable. This is, however, only the case if customer demand is constant, i.e., if all replenishment orders are for the same amount of material.

If customer demand is stochastic, the treatment of replenishment orders as interchangeable will lead to approximate results. However, we observe a gradual decline in accuracy of the analytical results connected with the comparability of replenishment orders. I.e., the analytical results are fairly accurate if replenishment orders are for more or less the same amount. As obviously the demand process determines the variability of replenishment order amounts in our model, the approaches developed in the following work better the lower the coefficient of demand variation.

In any instance, the following approaches are an improvement on applying the methods for the dependent lead time case to cases where order crossover may occur.

Interestingly, the use of the effective lead time with the methods developed above for the dependent lead time case seems to result in the same values as the approaches we develop in the following. In fact, for the customer waiting time distributions, we cannot offer a better approach than replacing the original lead time distribution by the effective lead time. Nonetheless, we present different approaches for the other metrics that exhibit better scalability than our limited approach to determine the effective lead time.

6.3.4.1 Ready Rate

In Sects. 2.3.2.1 and 6.3.1.1 we have already mentioned the problems of measuring the ready rate on a different time basis than the customer order rate. We particularly observe these problems if the lead time model permits order crossover, where we cannot unambiguously assign orders to a certain order cycle. In this section, we will therefore focus on the ready rate per customer order (*Case:* $\tau = r_D$ in Sect. 6.3.1.1), and disregard the corresponding metric per replenishment order cycle.

As stated in Sect. 5.2, observing the lead time demand is misleading when crossover may occur. We cannot develop the analysis assuming a fixed lead time either, as this would clearly prevent order crossover. Instead, we can make use of the distribution of the amount of outstanding orders and the inventory shortfall, respectively. (See Sects. 5.2.1 and 5.2.2.) We now somewhat invert the perspective and ask how much stock is missing versus our target level S under certain circumstances, while we previously generally asked for the inventory consumption in a certain set of periods. The following approach thus appears to be closer to the actual problem, while the previous approach only works because the general inventory consumption equals the missing stock if orders are not allowed not crossover.

As in Sect. 6.3.1.1, we will first assume that $r_D = 1$. The central idea is to associate replenishment orders with the r-time period demand they are intended to cover. Analogous to (6.1), we have (6.63) for the ready rate per period:

$$\alpha^{\tau=1} = \sum_{t^* \in T^*} \sum_{k \in K(r,t^*)} P\{K(r,t^*) = k\} \cdot \frac{1}{r} \cdot CDF_{D^{[k \cdot r + t^*]}}(S) \tag{6.63}$$

For an arbitrary demand rate r_D, (6.64) applies:

$$\alpha^{\tau=r_D} = \sum_{t^{**} \in T^{**}} \sum_{k \in K(r,t^{**} \cdot r_D)} P\{K(r,t^{**} \cdot r_D) = k\} \cdot \frac{r_D}{r} \cdot CDF_{D^{[k \cdot r + t^{**} \cdot r_D)]}}(S) \tag{6.64}$$

Numerical Example. Considering Example 7 of Table 6.1 (p. 83), the ready rate per period can be calculated as follows:

$$
\begin{aligned}
K(2,0) \sim & Disc\{(0,0.0),(1,0.5),(2,0.5)\} \\
K(2,1) \sim & Disc\{(0,0.25),(1,0.5),(2,0.25)\}
\end{aligned}
$$

$$
\begin{aligned}
\alpha^{\tau=1} &= \frac{1}{2} \cdot P\{K(2,0)=1\} \cdot CDF_{D^{[2]}}(300) \\
&\quad + \frac{1}{2} \cdot P\{K(2,0)=2\} \cdot CDF_{D^{[4]}}(300) \\
&\quad + \frac{1}{2} \cdot P\{K(2,1)=0\} \cdot CDF_{D^{[1]}}(300) \\
&\quad + \frac{1}{2} \cdot P\{K(2,1)=1\} \cdot CDF_{D^{[3]}}(300) \\
&\quad + \frac{1}{2} \cdot P\{K(2,1)=2\} \cdot CDF_{D^{[5]}}(300) \\
&= 0.5 \cdot 0.5 \cdot 0.990789 + 0.5 \cdot 0.5 \cdot 0.047790 \\
&\quad + 0.5 \cdot 0.25 \cdot 1.0 + 0.5 \cdot 0.5 \cdot 0.5 + 0.5 \cdot 0.25 \cdot 0.001435 \\
&= 0.509824
\end{aligned}
$$

Note that (6.63) and (6.64) also apply for the case of dependent lead times, and may be used alternatively to (6.1) and (6.2) if the distribution of the number of outstanding orders is given.

6.3.4.2 Customer Waiting Time per Order

The probabilities of customer waiting times $w > 0$ clearly cannot be derived from the distribution of outstanding orders alone. Besides the information that an order has to wait, we need to know when a replenishment order will finally arrive that is sufficient to cover the waiting customer order's demand.

A promising attempt, we have found here, is to use the effective lead time with the formulae that apply for the dependent lead time case in Sect. 6.3.1.2. As mentioned in the introduction to this section, this approach is not exact as it assumes that orders are interchangeable. We observe a decline in accuracy along with an increase in customer order variability. Table 6.6 illustrates the phenomenon. We examined Example 7 of Table 6.1 (p. 83) for different demand distributions with none, low and high coefficients of variation. The analytical results were compared with the result of simulation experiments, where we conducted 40 replications with 2,000,000 periods for each instance. In the table, the analytical results are printed as single numbers above the 0.98%-t-confidence intervals that resulted from the simulation runs. Note that for a constant demand (here $D = 100$), the simulation results correspond to the analytical results, while $D \sim Unif(\{1, 200\})$ reveals the most significant deviation.

Table 6.6 Waiting times per order – accuracy for different demand distributions

	$D = 100$	$D \sim Norm$ (100, 10)	$D \sim Unif$ ($\{1, 2, \ldots, 200\}$)	$D \sim Unif$ ($\{1, 200\}$)
$P\{W^O = 0\}$	0.7813	0.6875	0.6749	0.5460
	[0.7812, 0.7816]	[0.6875, 0.6879]	[0.6747, 0.6751]	[0.5456, 0.5461]
$P\{W^O = 1\}$	0.125	0.1562	0.1362	0.1445
	[0.1248, 0.1250]	[0.1561, 0.1563]	[0.1361, 0.1362]	[0.1445, 0.1446]
$P\{W^O = 2\}$	0.0625	0.0937	0.0972	0.1232
	[0.0624, 0.0625]	[0.0936, 0.0938]	[0.0952, 0.0954]	[0.1202, 0.1204]
$P\{W^O = 3\}$	0.0313	0.0469	0.0562	0.0955
	[0.0312, 0.0313]	[0.0467, 0.0469]	[0.0551, 0.0552]	[0.0934, 0.0935]
$P\{W^O = 4\}$	0.0	0.0156	0.0257	0.0557
	[0.0, 0.0]	[0.0156, 0.0156]	[0.0274, 0.0275]	[0.0565, 0.0566]
$P\{W^O = 5\}$	0.0	0.0	0.0085	0.0254
	[0.0, 0.0]	[0.0, 0.0]	[0.0095, 0.0096]	[0.0284, 0.0285]
$P\{W^O = 6\}$	0.0	0.0	0.0013	0.0098
	[0.0, 0.0]	[0.0, 0.0]	[0.014, 0.014]	[0.0108, 0.0109]

Table 6.7 Example instances 7–12 – order-related metrics

		$P\{W^O = w\}$					
No.	$\alpha^{\tau = r_D}$	$w = 0$	$w = 1$	$w = 2$	$w = 3$	$w = 4$	$w > 4$
7	0.5098	0.5098	0.2438	0.1815	0.0637	0.0012	0.0
8	0.5679	0.5679	0.1226	0.1085	0.0778	0.0609	0.0623
9	0.5726	0.5726	0.1202	0.125	0.031	0.1202	0.031
10	0.3079	0.3079	0.2058	0.207	0.1543	0.0938	0.0313
11	0.2697	0.2697	0.1054	0.1183	0.1177	0.1089	0.28
12	0.2548	0.2548	0.1229	0.117	0.1111	0.1229	0.2712

6.3.4.3 Example Results

Table 6.7 summarizes the solutions to the example instances 7–12 of Table 6.1 for fill rates and customer waiting time per part.

6.3.5 Order Crossover, Volume View, Split Deliveries

6.3.5.1 Mean Backorder

As for the ready rate, we can exploit the distribution of outstanding orders or rather the shortfall distribution to retrieve the mean backorder amount. In fact, we will see that this leads us to a fairly intuitive representation.

Let us again start our considerations by assuming a customer order arrival rate of $r_D = 1$. Assuming that $K(r, t^*) = k$ replenishment orders are outstanding in period t^* of the ordering cycle, we can determine the total backorder amount in that period according to (6.65):

$$E[B|T^* = t^*, K(r,t^*) = k, r_D = 1] = \int_S^{\infty} PDF_{D[r\cdot k+t^*]}(x)\cdot(x-S)\,dx \tag{6.65}$$

For arbitrary values r_D, we derive equation (6.66). Remember that the volume of each replenishment order corresponds to the customer demand of r periods, where the customer demand in the order period $t^* = 0$ will be considered with the corresponding order:

$$E[B|T^* = t^*, K(r,t^*) = k] = \int_S^{\infty} PDF_{D[r\cdot k+nf_1(t^*,r_D)]}(x)\cdot(x-S)\,dx \tag{6.66}$$

Analogously to (6.27), we derive the mean backorder amount for the whole system according to (6.67):

$$E[B] = \sum_{t^*\in T^*}\sum_{k\in K(r,t^*)} \frac{P\{K(r,t^*) = k\}}{r}\cdot E[B|T^* = t^*, K(r,t^*) = k] \tag{6.67}$$

Now let us turn to the fraction of demand that will be backordered in certain periods t of the order cycle. We have already considered above that this equals the mean backorder in the specific period minus the theoretical backorder if it had not been for the demand in the period observed. In this case, we obtain (6.68):

$$\begin{aligned} E[B^*|T^* = t^*|K(r,t^*) = k] &= \int_S^{\infty} PDF_{D[r\cdot k+nf_1(t^*,r_D)]}(x)\cdot(x-S)\,dx \\ &- \int_S^{\infty} PDF_{D[r\cdot k+nf_1(t^*-1,r_D)]}(x)\cdot(x-S)\,dx \end{aligned} \tag{6.68}$$

For the overall system, we obtain (6.69). The dependence of the number of outstanding orders $K(r,t^*)$ on t^* prevents us from applying similar simplifications to those that lead to (6.29):

$$E[B^*] = \sum_{t^*\in T^*}\sum_{k\in K(r,t^*)} \frac{P\{K(r,t^*) = k\}}{r}\cdot E[B|T^* = t^*|K(r,t^*) = k] \tag{6.69}$$

Numerical Example. Considering Example 7 of Table 6.1 (p. 83), the backorder amounts can be calculated as follows. The distributions of $K(2,0)$ and $K(2,1)$ can be found in the concluding example of Sect. 6.3.4.1.

$$
\begin{aligned}
E[B] &= \frac{1}{2} \cdot P\{K(2,0)=1\} \cdot \int_S^\infty PDF_{D[2]}(x) \cdot (x-S)\,dx \\
&\quad + \frac{1}{2} \cdot P\{K(2,0)=2\} \cdot \int_S^\infty PDF_{D[4]}(x) \cdot (x-S)\,dx \\
&\quad + \frac{1}{2} \cdot P\{K(2,0)=0\} \cdot \int_S^\infty PDF_{D[1]}(x) \cdot (x-S)\,dx \\
&\quad + \frac{1}{2} \cdot P\{K(2,0)=1\} \cdot \int_S^\infty PDF_{D[3]}(x) \cdot (x-S)\,dx \\
&\quad + \frac{1}{2} \cdot P\{K(2,0)=2\} \cdot \int_S^\infty PDF_{D[5]}(x) \cdot (x-S)\,dx \\
&= 0.25 \cdot 0.131247 + 0.25 \cdot 101.189593 + 0.125 \cdot 0.0 \\
&\quad + 0.25 \cdot 20.729649 + 0.125 \cdot 200.027370 \\
&= 55.516050
\end{aligned}
$$

$$
\begin{aligned}
E[B^*] &= \frac{1}{2} \cdot P\{K(2,0)=1\} \cdot \Bigg(\int_S^\infty PDF_{D[2]}(x) \cdot (x-S)\,dx \\
&\quad - \int_S^\infty PDF_{D[1]}(x) \cdot (x-S)\,dx \Bigg) \\
&\quad + \frac{1}{2} \cdot P\{K(2,0)=2\} \cdot \Bigg(\int_S^\infty PDF_{D[4]}(x) \cdot (x-S)\,dx \\
&\quad - \int_S^\infty PDF_{D[3]}(x) \cdot (x-S)\,dx \Bigg) \\
&\quad + (\ldots) \\
&= 0.25 \cdot (0.131274 - 0) + 0.25 \cdot (101.189593 - 20.729649) \\
&\quad + 0.125 \cdot (0 - 0) + 0.25 \cdot (20.729649 - 0.131274) \\
&\quad + 0.125 \cdot (200.027370 - 101.189593) \\
&= 37.652120
\end{aligned}
$$

6.3.5.2 Mean Inventory

Similar to the approach for calculating the mean backorder amount, we can derive mean inventory levels for the crossover case by conceptually replacing the lead time demand by the demand that is bound in outstanding orders. For an arbitrary customer demand rate and a fixed number of outstanding orders we have (6.70), where $t^* = 0$ denotes the period of the order cycle in which replenishment orders are issued:

$$E[I|T^* = t^*, K(r,t^*) = k]$$
$$= \int_{-\infty}^{\infty} PDF_{D^{[r \cdot k + nf_1(t^*, r_D)]}}(x) \cdot (S - \mathrm{Min}\{x, S\})\, dx \tag{6.70}$$

After the analogous restatements as for the dependent lead time case, we obtain (6.71) and finally (6.72):

$$E[I|T^* = t^*, K(r,t^*) = k] = S - \int_{-\infty}^{S} PDF_{D^{[r \cdot k + nf_1(t^*, r_D)]}}(x) \cdot x\, dx$$
$$- (1 - CDF_{D^{[r \cdot k + nf_1(t^*, r_D)]}}(S)) \cdot S \tag{6.71}$$

$$E[I] = \sum_{t^* \in T^*} \sum_{k \in K(r,t^*)} \frac{P\{K(r,t^*) = k\}}{r} \cdot E[I|T^* = t^*, K(r,t^*) = k] \tag{6.72}$$

Analogously to (6.35), we can also determine the mean inventory levels based on the mean demand and mean backorder amount:

$$E[I|T^* = t^*, K(r,t^*) = k]$$
$$= S - E[D^{[k \cdot r + t^*]}] + E[B|T^* = t^*, K(r,t^*) = k] \tag{6.73}$$

Numerical Example. Considering Example 7 of Table 6.1 (p. 83), the mean inventory levels can be calculated as follows. The distributions of $K(2,0)$ and $K(2,1)$ can be found in the concluding example of Sect. 6.3.4.1. Note that for this example the mean backorder amount equals the mean inventory level.

$$\begin{aligned} E[I] = {} & \frac{1}{2} \cdot P\{K(2,0) = 1\} \cdot (S - E[D^{[2]}] + E[B|T^* = 0, K(2,0) = 1]) \\ & + \frac{1}{2} \cdot P\{K(2,0) = 2\} \cdot (S - E[D^{[4]}] + E[B|T^* = 0, K(2,0) = 2]) \\ & + \frac{1}{2} \cdot P\{K(2,1) = 0\} \cdot (S - E[D^{[1]}] + E[B|T^* = 0, K(2,1) = 0]) \\ & + \frac{1}{2} \cdot P\{K(2,1) = 1\} \cdot (S - E[D^{[3]}] + E[B|T^* = 0, K(2,1) = 1]) \\ & + \frac{1}{2} \cdot P\{K(2,1) = 2\} \cdot (S - E[D^{[5]}] + E[B|T^* = 0, K(2,1) = 2]) \end{aligned}$$

$$
\begin{aligned}
&= 0.25 \cdot (300 - 200 + 0.131274) + 0.25 \cdot (300 - 400 + 101.189593) \\
&\quad + 0.125 \cdot (300 - 100 + 0) + 0.25 \cdot (300 - 300 + 20.729649) \\
&\quad + 0.125 \cdot (300 - 500 + 200.027370) \\
&= 55.516050
\end{aligned}
$$

6.3.5.3 Fill Rate

Using equations (6.67) for $E[B]$ and (6.69) for $E[B^*]$, we can apply formulae (6.36) and (6.37) to calculate both fill rate metrics.

Example 7 of Table 6.1 (p. 83) results in the following fill rates:

$$
\begin{aligned}
\beta &= 1 - 37.652120 \div 100 = 0.623479 \\
\gamma &= 1 - 55.516050 \div 100 = 0.444839
\end{aligned}
$$

6.3.5.4 Customer Waiting Time per Part

For the customer waiting time per part, we propose that applying the effective lead time with the formulae described for the dependent lead time case can be used as an approximation. We observe the same problems with this approach as we have already outlined for the customer waiting time per order (Sect. 6.3.4.2). Here, it becomes even clearer that neglecting the non-interchangeability of replenishment orders leads to a systematic error, correlated with the coefficient of demand variation. See Table 6.8 for an illustration of the phenomenon. As above, each simulation run consisted of 40 replications with 2,000,000 periods. Note that both instances with uniformly distributed demand (and high variation) reveal significant deviations between the analytical and simulated values.

6.3.5.5 Example Results

Table 6.9 summarizes the solutions to the example instances 7–12 of Table 6.1 for the performance indicators that we analyzed in this section.

6.3.6 Order Crossover, Volume View, Full Deliveries

6.3.6.1 Mean Backorder

To determine the mean backorder when order crossover may occur, we can apply similar considerations as for the dependent lead time case that we discussed in

Table 6.8 Waiting times per part (split deliveries) – accuracy for different demand distributions

	$D = 100$	$D \sim Norm$ (100, 10)	$D \sim Unif$ ($\{1, 2, \ldots, 200\}$)	$D \sim Unif$ ($\{1, 200\}$)
$P\{W^V = 0\}$	0.7813	0.7751	0.7236	0.6643
	[0.7812, 0.7816]	[0.7750, 0.7753]	[0.7234, 0.7238]	[0.6641, 0.6646]
$P\{W^V = 1\}$	0.125	0.1250	0.1242	0.1240
	[0.1248, 0.1250]	[0.1249, 0.1251]	[0.1267, 0.1268]	[0.1345, 0.1348]
$P\{W^V = 2\}$	0.0625	0.0655	0.0830	0.0963
	[0.0624, 0.0625]	[0.0655, 0.0656]	[0.0802, 0.0804]	[0.0891, 0.0893]
$P\{W^V = 3\}$	0.0313	0.0313	0.0446	0.0641
	[0.0312, 0.0313]	[0.0312, 0.0313]	[0.0428, 0.0429]	[0.0548, 0.0550]
$P\{W^V = 4\}$	0.0	0.0031	0.0187	0.0336
	[0.0, 0.0]	[0.0030, 0.0031]	[0.0201, 0.0201]	[0.0325, 0.0327]
$P\{W^V = 5\}$	0.0	0.0	0.0053	0.0138
	[0.0, 0.0]	[0.0, 0.0]	[0.0061, 0.0062]	[0.0181, 0.0182]
$P\{W^V = 6\}$	0.0	0.0	0.0	0.0040
	[0.0, 0.0]	[0.0, 0.0]	[0.0, 0.0]	[0.0061, 0.0062]

Table 6.9 Example instances 7–12 – volume-related metrics (split deliveries)

	Service levels		$P\{W^V = w\}$					
No.	β	γ	$w = 0$	$w = 1$	$w = 2$	$w = 3$	$w = 4$	$w > 4$
7	0.6235	0.4448	0.6235	0.2241	0.1263	0.0259	0.0002	0.0
8	0.6242	0.0622	0.6242	0.1179	0.0949	0.0683	0.0555	0.0392
9	0.6867	0.2517	0.6867	0.0878	0.1242	0.0068	0.0878	0.0068
10	0.3292	−0.4581	0.3292	0.2233	0.2083	0.1523	0.0729	0.0139
11	0.2765	−1.8467	0.2765	0.1075	0.1187	0.1177	0.1087	0.2709
12	0.3225	−1.4917	0.3225	0.1213	0.1212	0.1026	0.1213	0.2112

Sect. 6.3.3. The fundamental difference is again that we base our calculations on the distribution of outstanding orders instead of the lead time distribution, and consider $t^* = 0$ as first period of the order cycle. In the following, we will right away assume an arbitrary customer demand rate of r_D.

As for the dependent lead time case, we make use of the truth function SO as defined in Definition 23, p. 101. In contrast, we have to drop the concept of regarding the previous period in order to determine the inventory consumption prior to the new demand in the period that we are concerned with. Considering the distribution of outstanding orders, it is no longer sufficient to regard the situation in the preceding period. Instead, we have to determine inventory consumption until the period we are observing, without considering the particular period's own demand. The formulae for the dependent lead time case do in fact allow for the same interpretation. Here, the very fact that we may assume the same number of outstanding orders in the two consecutive periods that we are analyzing allows us to equivalently think of the demand that has been accumulated until the previous period.

In this section we consider the truth function SO_t (as defined in Definition 23, p. 101) in the form $SO_{k \cdot r + t^*}$, which means that we ask for the demand that is bound in k orders plus the demand of t^* subperiods of the order cycle. $SO_{k\cdot r+t^*-1}$ then denotes whether S is exceeded by the demand on order plus subperiod demand minus the demand of the subperiod t^* that we are currently considering. Thus, we obtain the following expression equivalent to (6.52), i.e., the amount of demand in period t^* that will (still) be backordered in the relative period of the order cycle that we are analyzing:

$$\begin{aligned} &E[B^*|T^* = t^*, K(r,t^*) = k, SO_{k\cdot r+t^*-1} = 0] \\ &\quad = \int_S^\infty \left(\int_{-\infty}^\infty PDF_{D^{[nf_1(r\cdot k+t^*-1,r_D)]}_{(-\infty,S)}}(v-x) \right. \\ &\qquad \left. \cdot PDF_{D^{[nf_3(t^*,r_D)]}}(x) \cdot x \, dv \right) dx \end{aligned} \tag{6.74}$$

We may now state the equation that is equivalent to (6.53):

$$\begin{aligned} &E[B^*|T^* = t^*, K(r,t^*) = k] \\ &\quad = P\{SO_{k\cdot r+t^*-1} = 1 | K(r,t^*) = k\} \cdot E[D^{[nf_3(t^*,r_D)]}] \\ &\qquad + P\{SO_{k\cdot r+t^*-1} = 0 | K(r,t^*) = k\} \\ &\qquad \cdot E[B^*|T^* = t^*, K(r,t^*) = k, SO_{k\cdot r+t^*-1} = 0] \end{aligned} \tag{6.75}$$

The overall mean new backorder amount is then given by (6.76):

$$E[B^*] = \sum_{t^*\in T^*} \sum_{k\in K(r,t^*)} \frac{P\{K(r,t^*) = k\}}{r} \cdot E[B^*|T^* = t^*, K(r,t^*) = k] \tag{6.76}$$

Analogously to (6.55) and (6.56), we can determine the mean backorder amount for systems with $S \geq 0$ according to (6.77) and (6.78). (Note that $nf_2(\kappa,r) + \left\lfloor \frac{\kappa}{r} \right\rfloor \cdot r = \left(\frac{\kappa}{r} - \left\lfloor \frac{\kappa}{r} \right\rfloor \right) \cdot r + \left\lfloor \frac{\kappa}{r} \right\rfloor \cdot r = \kappa$.)

$$\begin{aligned} &E[B|T^* = t^*, K(r,t^*) = k] \\ &\quad = \sum_{\kappa=1}^{r\cdot k+t^*} E[B^*|T^* = nf_2(\kappa,r), K(r, nf_2(\kappa,r)) = \left\lfloor \frac{\kappa}{r} \right\rfloor \cdot r] \end{aligned} \tag{6.77}$$

$$E[B] = \sum_{t^*\in T^*} \sum_{k\in K(r,t^*)} \frac{P\{K(r,t^*) = k\}}{r} \cdot E[B|T^* = t^*, K(r,t^*) = k] \tag{6.78}$$

Numerical Example. Considering Example 10 of Table 6.1 (p. 83), the mean backorder levels can be calculated as follows. The distributions of $K(2,0)$ and $K(2,1)$ can be found in the concluding example of Sect. 6.3.4.1:

$$
\begin{aligned}
E[B^*] = \frac{1}{2} \cdot P\{K(2,0) = 1\} \cdot \Big(& P\{SO_{1\cdot 2-1} = 1 | K(2,0) = 1\} \cdot E[D^{[1]}] \\
& + P\{SO_1 = 0 | K(2,0) = 1\} \cdot E[B^* | T^* = 0, K(2,0) = 1, SO_1 = 0] \Big) \\
& + \frac{1}{2} \cdot P\{K(2,0) = 2\} \cdot \Big(P\{SO_{2\cdot 2-1} = 1 | K(2,0) = 2\} \cdot E[D^{[1]}] \\
& + P\{SO_3 = 0 | K(2,0) = 2\} \cdot E[B^* | T^* = 0, K(2,0) = 2, SO_3 = 0] \Big) \\
& + \frac{1}{2} \cdot P\{K(2,1) = 0\} \cdot \Big(P\{SO_{0\cdot 2+1-1} = 1 | K(2,1) = 0\} \cdot E[D^{[1]}] \\
& + P\{SO_0 = 0 | K(2,1) = 0\} \cdot E[B^* | T^* = 1, K(2,1) = 0, SO_0 = 0] \Big) \\
& + \frac{1}{2} \cdot P\{K(2,1) = 1\} \cdot \Big(P\{SO_{1\cdot 2+1-1} = 1 | K(2,1) = 1\} \cdot E[D^{[1]}] \\
& + P\{SO_2 = 0 | K(2,1) = 1\} \cdot E[B^* | T^* = 1, K(2,1) = 1, SO_2 = 0] \Big) \\
& + \frac{1}{2} \cdot P\{K(2,1) = 2\} \cdot \Big(P\{SO_{2\cdot 2+1-1} = 1 | K(2,1) = 2\} \cdot E[D^{[1]}] \\
& + P\{SO_4 = 0 | K(2,1) = 2\} \cdot E[B^* | T^* = 1, K(2,1) = 2, SO_4 = 0] \Big)
\end{aligned}
$$

$$
\begin{aligned}
& P\{SO_1 | K(2,0) = 1\} = CDF_{D^{[1]}}(80) \quad (= 1 - P\{SO_1 | K(2,0) = 0\}) \\
& \quad = 0.75 \\
& P\{SO_3 | K(2,0) = 2\} = CDF_{D^{[3]}}(80) = 0.265625 \\
& P\{SO_0 | K(2,1) = 0\} = CDF_{D^{[0]}}(80) = 0.0 \\
& P\{SO_2 | K(2,1) = 1\} = CDF_{D^{[2]}}(80) = 0.5 \\
& P\{SO_4 | K(2,1) = 2\} = CDF_{D^{[4]}}(80) = 0.078125
\end{aligned}
$$

$$
\begin{aligned}
& E[B^* | T^* = 0, K(2,0) = 1, SO_1 = 0] \\
& \quad = \int_{80}^{\infty} \Big(\int_{-\infty}^{\infty} PDF_{D^{[1]}_{(-\infty,80)}}(v - x) \quad PDF_{D^{[1]}}(x)\, dv \Big)\, dx \\
& \quad = 29.166667 \\
& E[B^* | T^* = 0, K(2,0) = 2, SO_3 = 0] \\
& \qquad = \int_{80}^{\infty} \Big(\int_{-\infty}^{\infty} PDF_{D^{[3]}_{(-\infty,80)}}(v - x) \quad PDF_{D^{[1]}}(x)\, dv \Big)\, dx \\
& \qquad = 40.294118
\end{aligned}
$$

$$
\begin{aligned}
E[B^* | T^* &= 1, K(2,0) = 0, SO_0 = 0] \\
&= \int_{80}^{\infty} \left(\int_{-\infty}^{\infty} PDF_{D^{[0]}_{(-\infty,80)}}(v-x) \quad PDF_{D^{[1]}}(x)\, dv \right) dx \\
&= 25.0 \\
E[B^* | T^* &= 1, K(2,0) = 1, SO_2 = 0] \\
&= \int_{80}^{\infty} \left(\int_{-\infty}^{\infty} PDF_{D^{[2]}_{(-\infty,80)}}(v-x) \quad PDF_{D^{[1]}}(x)\, dv \right) dx \\
&= 34.0625 \\
E[B^* | T^* &= 1, K(2,0) = 2, SO_4 = 0] \\
&= \int_{80}^{\infty} \left(\int_{-\infty}^{\infty} PDF_{D^{[4]}_{-\infty,80}}(v-x) \quad PDF_{D^{[1]}}(x)\, dv \right) dx \\
&= 40.375
\end{aligned}
$$

$$
\begin{aligned}
E[B^*] = {} & 0.5 \cdot 0.5 \cdot (0.25 \cdot 45 + 0.75 \cdot 29.166667) \\
& + 0.5 \cdot 0.5 \cdot (0.734375 \cdot 45 + 0.265625 \cdot 40.294118) \\
& + 0.5 \cdot 0.25 \cdot (0 \cdot 45 + 1.0 \cdot 25) \\
& + 0.5 \cdot 0.5 \cdot (0.5 \cdot 45 + 0.5 \cdot 34.0625) \\
& + 0.5 \cdot 0.25 \cdot (0.921875 \cdot 45 + 0.078125 \cdot 40.375) \\
= {} & 0.25 \cdot 33.125 + 0.25 \cdot 43.75 + 0.125 \cdot 25 + 0.25 \cdot 39.53125 \\
& + 0.125 \cdot 44.638672 \\
= {} & 37.806397
\end{aligned}
$$

$$
\begin{aligned}
E[B] = {} & \frac{1}{2} \cdot P\{K(2,0) = 1\} \\
& \cdot \left(\sum_{\kappa=1}^{2} E[B^* | T^* = nf_2(\kappa,2), K(2, nf_2(\kappa,2)) = \left\lfloor \frac{\kappa}{2} \right\rfloor \cdot 2] \right) \\
& + \frac{1}{2} \cdot P\{K(2,0) = 2\} \\
& \cdot \left(\sum_{\kappa=1}^{4} E[B^* | T^* = nf_2(\kappa,2), K(2, nf_2(\kappa,2)) = \left\lfloor \frac{\kappa}{2} \right\rfloor \cdot 2] \right) \\
& + \frac{1}{2} \cdot P\{K(2,1) = 0\} \\
& \cdot \left(\sum_{\kappa=1}^{1} E[B^* | T^* = nf_2(\kappa,2), K(2, nf_2(\kappa,2)) = \left\lfloor \frac{\kappa}{2} \right\rfloor \cdot 2] \right)
\end{aligned}
$$

$$
\begin{aligned}
&+ \frac{1}{2} \cdot P\{K(2,1) = 1\} \\
&\cdot \Big(\sum_{\kappa=1}^{3} E[B^*|T^* = nf_2(\kappa, 2), K(2, nf_2(\kappa, 2)) = \left\lfloor \frac{\kappa}{2} \right\rfloor \cdot 2] \Big) \\
&+ \frac{1}{2} \cdot P\{K(2,1) = 2\} \\
&\cdot \Big(\sum_{\kappa=1}^{5} E[B^*|T^* = nf_2(\kappa, 2), K(2, nf_2(\kappa, 2)) = \left\lfloor \frac{\kappa}{2} \right\rfloor \cdot 2] \Big) \\
= &\frac{1}{2} \cdot P\{K(2,0) = 1\} \cdot (E[B^*|T^* = 1, K(2,1) = 0] \\
&+ E[B^*|T^* = 0, K(2,1) = 1]) \\
&+ \dots \\
= &\ 0.5 \cdot 0.5 \cdot (25 + 33.125) \\
&+ 0.5 \cdot 0.5 \cdot (25 + 33.125 + 39.53125 + 43.75) \\
&+ 0.5 \cdot 0.25 \cdot (25) \\
&+ 0.5 \cdot 0.5 \cdot (25 + 33.125 + 39.53125) \\
&+ 0.5 \cdot 0.25 \cdot (25 + 33.125 + 39.53125 + 43.75 + 44.638672) \\
= &\ 100.677490
\end{aligned}
$$

6.3.6.2 Mean Inventory

The equations for independent lead times can be developed analogously to the case of dependent lead times described in Sect. 6.3.3.2. The equivalent to (6.58) and (6.59) can then be stated as follows:

$$
\begin{aligned}
E[I^{(-)}|T^* = t^*, K(r, t^*) = k] = P\{SO^{k \cdot r + t^* - 1} = 0\} \cdot (E[D^{[nf_3(t^*, r_D)]}] \\
- E[B^*|T^* = t^*, K(r, t^*) = k, SO_{k \cdot r + t^* - 1} = 0])
\end{aligned} \tag{6.79}
$$

The actual mean inventory is then given according to (6.80):

$$
\begin{aligned}
E[I] = \sum_{t^* \in T^*} \sum_{k \in K_{t^*}} \frac{P\{K_{t^*} = k\}}{r} \\
\cdot \Big(S - \sum_{\kappa=1}^{r \cdot k + t^*} E[I^{(-)}|T^* = nf_2(\kappa, r), K(r, nf_2(\kappa, r)) = \left\lfloor \frac{\kappa}{r} \right\rfloor \cdot r] \Big)
\end{aligned} \tag{6.80}
$$

$$= \sum_{t^* \in T^*} \sum_{k \in K_{t^*}} \frac{P\{K_{t^*} = k\}}{r} \cdot \Big(S - \sum_{\kappa=1}^{r \cdot k + t^*} E[D^{[nf_1(\kappa, r_D)]}]$$

$$+ \sum_{\kappa=1}^{r \cdot k + t^*} E[B^* | T^* = nf_2(\kappa, r), K(r, nf_2(\kappa, r)) = \left\lfloor \frac{\kappa}{r} \right\rfloor \cdot r]\Big)$$

Considering Example 10 of Table 6.1 (p. 83), the mean inventory levels are calculated as follows:

$$E[I] = \frac{1}{2} \cdot P\{K(2,0) = 1\} \cdot \Big(S - \sum_{\kappa=1}^{2} E[D^{[nf_1(\kappa, r_D)]}]$$

$$+ \sum_{\kappa=1}^{2} E[B^* | T^* = nf_2(\kappa, 2), K(2, nf_2(\kappa, 2)) = \left\lfloor \frac{\kappa}{2} \right\rfloor \cdot 2]\Big)$$

$$+ \frac{1}{2} \cdot P\{K(2,0) = 2\} \cdot \Big(S - \sum_{\kappa=1}^{4} E[D^{[nf_1(\kappa, r_D)]}]$$

$$+ \sum_{\kappa=1}^{4} E[B^* | T^* = nf_2(\kappa, 2), K(2, nf_2(\kappa, 2)) = \left\lfloor \frac{\kappa}{2} \right\rfloor \cdot 2]\Big)$$

$$+ \frac{1}{2} \cdot P\{K(2,1) = 0\} \cdot \Big(S - \sum_{\kappa=1}^{1} E[D^{[nf_1(\kappa, r_D)]}]$$

$$+ \sum_{\kappa=1}^{1} E[B^* | T^* = nf_2(\kappa, 2), K(2, nf_2(\kappa, 2)) = \left\lfloor \frac{\kappa}{2} \right\rfloor \cdot 2]\Big)$$

$$+ \frac{1}{2} \cdot P\{K(2,1) = 1\} \cdot \Big(S - \sum_{\kappa=1}^{3} E[D^{[nf_1(\kappa, r_D)]}]$$

$$+ \sum_{\kappa=1}^{3} E[B^* | T^* = nf_2(\kappa, 2), K(2, nf_2(\kappa, 2)) = \left\lfloor \frac{\kappa}{2} \right\rfloor \cdot 2]\Big)$$

$$+ \frac{1}{2} \cdot P\{K(2,1) = 2\} \cdot \Big(S - \sum_{\kappa=1}^{5} E[D^{[nf_1(\kappa, r_D)]}]$$

$$+ \sum_{\kappa=1}^{5} E[B^* | T^* = nf_2(\kappa, 2), K(2, nf_2(\kappa, 2)) = \left\lfloor \frac{\kappa}{2} \right\rfloor \cdot 2]\Big)$$

$$= 0.5 \cdot 0.5 \cdot (80 - 2 \cdot 45 + 25 + 33.125)$$

$$+ 0.5 \cdot 0.5 \cdot (80 - 4 \cdot 45 + 25 + 33.125 + 39.53125 + 43.75)$$

$$
\begin{aligned}
&+ 0.5 \cdot 0.25 \cdot (80 - 1 \cdot 45 + 25) \\
&+ 0.5 \cdot 0.5 \cdot (80 - 3 \cdot 45 + 25 + 33.125 + 39.53125) \\
&+ 0.5 \cdot 0.25 \cdot (80 - 5 \cdot 45 + 25 + 33.125 + 39.53125 + 43.75 \\
&+ 44.638672) \\
= {} & 45.677490
\end{aligned}
$$

6.3.6.3 Fill Rate

With the findings of Sect. 6.3.6.1, we can directly derive both fill rates according to the general formulae given as (6.36) and (6.37). Example 10 of Table 6.1 (p. 83) results in the following fill rates:

$$\beta = 1 - 37.806397 \div 45 = 0.159858$$

$$\gamma = 1 - 100.677490 \div 45 = -1.237278$$

6.3.6.4 Customer Waiting Time per Part

For the customer waiting time with full deliveries only, we observe analogous problems as described in Sects. 6.3.4.2 and 6.3.5.4 and recommend the same remedy, namely using the effective lead time distribution (Table 6.10).

Table 6.10 Waiting times per part (full deliveries) – accuracy for different demand distributions

	$D = 100$	$D \sim Norm$ (100, 10)	$D \sim Unif$ ($\{1, 2, \ldots, 200\}$)	$D \sim Unif$ ($\{1, 200\}$)
$P\{W^V = 0\}$	0.7813	0.6860	0.6261	0.4261
	[0.7810, 0.7815]	[0.6842, 0.6846]	[0.6261, 0.6266]	[0.4259, 0.4263]
$P\{W^V = 1\}$	0.125	0.1568	0.1481	0.1657
	[0.1249, 0.1250]	[0.1578, 0.1579]	[0.1533, 0.1535]	[0.1879, 0.1881]
$P\{W^V = 2\}$	0.0625	0.0943	0.1114	0.1506
	[0.0624, 0.0626]	[0.0942, 0.0944]	[0.1080, 0.1082]	[0.1427, 0.1429]
$P\{W^V = 3\}$	0.0313	0.0471	0.0678	0.1271
	[0.0312, 0.0313]	[0.0473, 0.0474]	[0.0629, 0.0631]	[0.1140, 0.1142]
$P\{W^V = 4\}$	0.0	0.0159	0.0327	0.0779
	[0.0, 0.0]	[0.0161, 0.0162]	[0.0336, 0.0337]	[0.06974, 0.7000]
$P\{W^V = 5\}$	0.0	0.0	0.0118	0.0037
	[0.0, 0.0]	[0.0, 0.0]	[0.0131, 0.0132]	[0.0413, 0.0414]
$P\{W^V = 6\}$	0.0	0.0	0.0021	0.0156
	[0.0, 0.0]	[0.0, 0.0]	[0.0023, 0.0023]	[0.0177, 0.0178]

Table 6.11 Example instances 7–12 – volume-related metrics (full deliveries)

	Service levels		$P\{W^V = w\}$						
No.	β	γ	$w = 0$	$w = 1$	$w = 2$	$w = 3$	$w = 4$	$w = 5$	$w > 5$
7	0.4885	0.1469	0.4885	0.2458	0.1915	0.0724	0.0017	0.0	0.0
8	0.5572	−0.187	0.5572	0.1231	0.1107	0.0802	0.0617	0.0482	0.0188
9	0.5613	−0.1764	0.5613	0.1213	0.125	0.0356	0.1213	0.0	0.0356
10	0.1599	−1.2373	0.1599	0.1761	0.2313	0.2018	0.1615	0.0694	0.0
11	0.211	−2.3651	0.211	0.0959	0.1151	0.1214	0.1161	0.107	0.2334
12	0.1904	−2.3819	0.1904	0.1242	0.1076	0.1181	0.1242	0.0914	0.2442

6.3.6.5 Example Results

Table 6.11 summarizes the solutions to the example instances 7–12 of Table 6.1 for the fill rates and customer waiting time per part. To analyze the first three examples (7–9), we used discrete transformations of the corresponding normal distributions to compute the convolution of truncated and non-truncated distributions.

6.4 Optimization

We are now able to evaluate certain performance aspects of (r, S) inventory systems that meet the given assumptions based of the analytical findings presented in Sect. 6.3. In this section we will ask how we can use these formulae to identify the best system configuration according to a certain optimality criterium.

Clarifying two aspects is preliminary to identifying an optimal (r, S) policy. We have to state which parameters we may change, and we have to define how we prefer one configuration to another one. Regarding the first aspect, we will assume that we may always adjust S, and we will examine the cases where r is prescribed by other needs of the system, and where we may adjust it as well. In terms of the second aspect, we will consider the cost- and service-oriented concepts for controling backorder levels that can be found in the literature, i.e., we will distinguish between cases where backorder costs can be quantified in a satisfying manner, and where the system performance must accomplish some target service level.

We will give separate findings for the cases of split and full deliveries due to the fact that they come with different mathematical properties. In terms of the general solution approach, there is hardly any difference between optimization of the independent and dependent lead time case. We will therefore focus on the dependent lead time case, for which we have exact analytical formulas at hand.

Table 6.12 summarizes the different optimization objectives that we will examine in the following sections.

Table 6.12 Overview of optimization objectives

	Split deliveries	Full deliveries
Set S s.t. Service	Sect. 6.4.1.1	Sect. 6.4.2.1
Set r and S s.t. Service	Sect. 6.4.1.2	Sect. 6.4.2.2
Set S refl. Backorder Costs	Sect. 6.4.1.3	Sect. 6.4.2.3
Set r and S refl. Backorder Costs	Sect. 6.4.1.4	Sect. 6.4.2.4

6.4.1 Split Deliveries

6.4.1.1 Set S Subject to Service

In the first optimization scenario we assume that r is prescribed and a certain target service level should be accomplished. We have to solve the following minimization problem, where $sv(S)$ is the value of the service criterion dependent on S, and sv^* is the service level aspired to:

$$Min!S \tag{6.81}$$

$$s.t.$$

$$sv(S) \geq sv^* \tag{6.82}$$

Conjecture 4 describes the grounds on which we intend to solve the optimization problem stated in (6.81) and (6.82).

Conjecture 4. All service levels that are analyzed in Sects. 6.3.1.1 and 6.3.2.3 monotonically increase in S.

In other words, let A and B be two inventory systems that only differ in their parameters S, where $S_A < S_B$; then system B will perform better or equal in all service levels analyzed in the above sections.

Incomplete Proof. Let us regard the service levels one by one. For the ready rate per customer order we have (6.4), which clearly monotonically increases in S as $P\{X \leq x\} \leq P\{X \leq y\}; x < y$ obviously holds for any random variable X. In view of (6.6), the same argument applies to prove that the ready rate per order cycle also monotonically increases in S.

From (6.36) and (6.37) it follows that the two fill rates monotonically increase in S if the according backorder amount monotonically decreases in S. From (6.25) and (6.27), respectively, we can derive that the backorder amount considered in (6.37) monotonically decreases in S. In (6.25), both the integral area and the multiplier $(x - S)$ decrease when S is increased. For the backorder amount considered in (6.36), it needs to be shown that (6.28) monotonically decreases. Although it is a somewhat clear-cut issue, it is not easily shown that this holds true in general. We will therefore leave the proof open to future research.

Using the presumed property that $sv(S)$ monotonically increases for all service levels, we can apply the same algorithmic idea that we used in Sect. 4.1 to approximate the quantile function of arbitrary distributions. (Algorithm 7).

Algorithm 7: Approximation of the required order-up-to level S

```
Input: Target service level sv*, tolerance t
Output: Required S
S_lb := S_min;
S_ub := S_max;
S := 1/2 · (S_lb + S_ub);
sv_S := sv(S);
while |sv_S − sv*| > t do
    if sv_S < sv* then
    |   S_lb := S;
    else
    |   S_ub := S;
    end
    S := 1/2 · (S_lb + S_ub);
    sv_S := sv(S);
end
```

Finally, let us examine the computational complexity. Obviously we can directly derive it from the analysis given in Sect. 4.1. Thus, we have $CC_{a7} = O((k + l) \cdot CC_{SVS})$, where CC_{SVS} is the computational complexity of calculating the service according to a certain value S.

Alternatively to a service level, one may also consider a service criterion related to the waiting time, i.e., the mean waiting time should be n periods at most. Here, we obtain the following optimization model that we may solve analogously to the above approach using the analytical functions developed in Sects. 6.3.1.2 and 6.3.2.4:

$$Min!S \tag{6.83}$$

$$s.t.$$

$$E[W(S)] \leq n \tag{6.84}$$

6.4.1.2 Set r and S Subject to Service

In Sect. 6.4.1.1, we have seen how to set the parameter S to accomplish a certain service level when r is given. In this section we will take the somewhat next logical step, namely extend the approach to setting both decision parameters of an (r, S) policy.

In the first place, we have to specify cost rates to balance the costs of ordering against those of holding inventory. In Sect. 2.3.1 we have already introduced c_1 as purchase costs and c_2 as inventory holding costs per period. Thus, we can specify the mean costs per period for operating an (r, S) inventory system according to (6.85), where $S_{opt}(r)$ is the solution of the optimization problem stated in (6.81) and (6.82) for a given r:

$$E[C_1(r, S_{opt}(r))] = \frac{c_1}{r} + c_2 \cdot E[I(r, S_{opt}(r))] \tag{6.85}$$

Due to the embedded optimization problem, we cannot simply consider the derivative of (6.85) to solve the problem. However, if we may reasonably assume that (6.85) is convex in r, we may apply some general enumeration method to determine the optimum parameter combination. In that case, Algorithm 8 may serve us to solve the problem, where *getMeanInventory*(r, S) determines the mean inventory levels according to (6.34) and *getRequiredS*(r, sv^*) corresponds to Algorithm 7.

Algorithm 8: Simple algorithm to calculate r_{opt}

```
Input: Target service level sv*, order costs c1, holding costs c2
Output: Optimum cycle r_opt
prevCosts := c1 + c2 · getMeanInventory(1, getRequiredS(1, sv*));
newCosts := c1/2 + c2 · getMeanInventory(2, getRequiredS(2, sv*));
r_opt := 2;
while prevCosts > newCosts do
    prevCosts := newCosts;
    r_opt := r_opt + 1;
    newCosts := c1/r_opt + c2· getMeanInventory(r_opt, getRequiredS(r_opt, sv*));
end
r_opt := r_opt − 1;
```

The computational complexity of Algorithm 8 is given via $CC_{a8} = O(r_{opt} \cdot (CC_{ME} + CC_{a7}))$, where CC_{ME} is the complexity to calculate the mean inventory that mainly depends on the selected demand distribution, and CC_{a7} is the complexity of determining the required S, as described in Sect. 6.4.1.1.

The runtime may be improved by applying the following approach. Let x_{lb} and x_{ub} be two values, where we know that $x_{lb} \leq r_{opt}$ and $x_{ub} \geq r_{opt}$. Furthermore, let x_{lm} and x_{rm} be two values so that $x_{lb} < x_{lm} < x_{rm} < x_{ub}$ holds. Finally, let x_{min} be the one value of the four that results in the least mean purchasing and inventory holding costs. We can then distinguish three cases: (1) x_{min} has a left and right neighbor, where a neighbor is the next smaller (left) or greater (right) value out of the remaining three, (2) x_{min} has a left neighbor only, and (3) x_{min} has a right neighbor only. In case (1), we know that the minimum is between the left and right neighbor as the costs have risen from the current minimum and may not fall again due to the convexity. For the same reason, we know that the minimum is between x_{min} and the left neighbor in case (2), and between x_{min} and the right neighbor in case (3). The algorithmic idea is now to redefine x_{lb} and x_{ub} as those points that we have identified as bounds for the possible location of r_{opt}, then set two new values x_{lm} and x_{rm}, and proceed as above. Algorithm 9 outlines the approach described above.

To set up Algorithm 9, we may set $x_{lb} = 1$ and identify x_{ub} by Algorithm 10, which slightly modifies Algorithm 8.

Let us begin with Algorithm 10 to examine the overall computational complexity. We observe that the inner loop definitely terminates after the second value beyond

Algorithm 9: Improved algorithm to calculate r_{opt}

Input: Lower bound x_{lb}, upper bound x_{ub}
Output: Optimum cycle $P[0].x$
$P := newCoordinate[4];$
$P[0].x := x_{lb}; P[0].y := getCosts(P[0].x);$
$P[1].x := x_{ub}; P[1].y := getCosts(P[1].x);$
$P[2].x := P[0].x + round(0.33 \cdot (P[1].x - P[0].x); P[2].y := getCosts(P[2].x);$
$P[3].x := P[0].x + round(0.66 \cdot (P[1].x - P[0].x); P[3].y := getCosts(P[3].x);$
$stop$:=false;
while !$stop$ **do**
  $sort(P);$ `/* by y, ascending */`
  $i_l = getIndexLeftNeigbor(P[0]);$
  $i_r = getIndexRightNeigbor(P[0]);$
  if $i_l == 0$ **then**
    $P[1].x = P[i_r].x; P[1].y = P[i_r].y;$ `/* set right neighbor, no left neighbor */`
  else if $i_r == 0$ **then**
    $P[1].x = P[i_l].x; P[1].y = P[i_l].y;$ `/* set left neighbor, no right neighbor */`
  else
    $P[0].x = P[i_l].x; P[0].y = P[i_l].y;$ `/* set left neighbor */`
    $P[1].x = P[i_r].x; P[1].y = P[i_r].y;$ `/* set right neighbor */`
  end
  if $P[1] - P[0] == 1$ **then**
    $stop$:=true; `/* P[0] is optimum */`
  else if $P[1] - P[0] == 2$ **then**
    $P[2].x := round(0.5 \cdot (P[1].x + P[0].x)); P[2].y := getCosts(P[2].x);$
    if $P[2].y < P[0].y$ **then**
      $P[0].x := P[2].x; P[0].y := P[2].y;$
    end
    $stop$:= true; `/* P[0] is optimum */`
  else
    $P[2].x := P[0].x + round(0.33 \cdot (P[1].x - P[0].x); P[2].y := getCosts(P[2].x);$
    $P[3].x := P[0].x + round(0.66 \cdot (P[1].x - P[0].x); P[3].y := getCosts(P[3].x);$ `/* continue */`
  end
end

the optimum is examined. Due to the convexity, the second value has higher cost than the first one after the optimum. Thus, the highest value the algorithm may possibly return is $2 \cdot 2 \cdot (r_{opt} - 1)$, leading to a runtime of $CC_{a4} = O(log_2(r_{opt}) \cdot (CC_{ME} + CC_{a7}))$.

Algorithm 9 starts with an interval of length $x_{ub} - x_{lb} = 4 \cdot r_{opt} - 1 \leq 4 \cdot r_{opt}$. In each step of the while-loop, the interval decreases by $\frac{1}{3}$ at least. Thus, the interval has length $(\frac{2}{3})^n \cdot 4 \cdot r_{opt}$ after n iterations. Since the algorithm terminates as soon as the interval is 2 or lower, the upper bound for necessary loops n can be determined as follows:

Algorithm 10: Algorithm to calculate right bound for r_{opt}

Input: Target service level sv^*, order costs c_1, holding costs c_2
Output: Right bound for r_{opt}
$prevCosts := getCosts(1)$;
$newCosts := getCosts(2)$;
$x_{ub} = 2$;
while $prevCosts > newCosts$ **do**
  $prevCosts = newCosts$;
  $x_{ub} = 2 \cdot x_{ub}$;
  $newCosts = getCosts(x_{ub})$;
end

$$\left(\frac{2}{3}\right)^n \cdot 4 \cdot r_{opt} \leq 2$$
$$\Leftrightarrow \left(\frac{3}{2}\right)^n \leq 2 \cdot r_{opt}$$
$$\Leftrightarrow n \leq \log_{\frac{3}{2}}(2 \cdot r_{opt})$$

Thus, the runtime of Algorithm 9 is given by $CC_{a9} = O(\log_{\frac{3}{2}}(r_{opt}) \cdot (CC_{ME} + CC_{a7}))$. At first glance this is an improvement on the simple approach that underlies Algorithm 8. However, the simple algorithm may in fact perform better if the computational complexity of the convolution depends on r, and thus (CC_{ME} and CC_{a7}) depend on r as well. In that case, the advantage of having fewer values to calculate may be overcompensated by the effort of convolving the demand distribution $4 \cdot r_{opt}$-times instead of r_{opt}-times. (See Sect. 4.2.)

With the algorithmic proceeds described above, we are able to determine the optimum order cycle r_{opt} if the relevant cost function is convex in r. Unfortunately, this is not the case for all possible combinations of service levels and demand distributions. For example, if demand is discretely distributed with just a few sampling points and we regard the ready rate per period, we observe jump discontinuities in $sv(S)$, i.e., we may not be able to exactly meet some service level but have to overfulfill the requirements. In the event that we have to overfulfill the service for a value r very significantly, while $r + 1$ allows us to meet the required service level very closely, we may observe an up and down of costs in consecutive values.

However, this problem can be overcome by slightly modifying Algorithm 8 according to Algorithm 11, which also takes into account the difference in service of two consecutive solutions. It stops as soon as the next solution is worse in costs as well as service than the best solution so far.

6.4.1.3 Set S Reflecting Backorder Costs

In Sects. 6.4.1.1 and 6.4.1.2 we applied the strategy of controlling the backorder amount by prescribing a certain minimum service level that the system must

Algorithm 11: Modified algorithm to calculate r_{opt}

Input: Target service level sv^*, order costs c_1, holding costs c_2
Output: Optimum cycle r_{opt}

```
r_opt := 1;
optCosts := c1 + c2 · getMeanInventory(1, getRequiredS(1, sv*));
optService := getService(1, getRequiredS(1, sv*));
r_new := 1;
newCosts := c1/2 + c2 · getMeanInventory(2, getRequiredS(2, sv*));
newService := getService(2, getRequiredS(2, sv*));
while optCosts > newCosts||optService < newService do
    if optCosts > newCosts then
        r_opt := r_new;
        optCosts := newCosts;
        optService := newService;
    end
    r_opt := r_opt + 1;
    newCosts := c1/r_opt + c2 · getMeanInventory(r_opt, getRequiredS(r_opt, sv*));
    newService := getService(r_opt, getRequiredS(r_opt, sv*));
end
```

accomplish. If backorder costs are quantifiable, we should not follow the above approach but instead count costs for inventory holding against costs for backordering. In the proceeding we will separately consider all three backorder cost types that we have introduced in Sect. 2.3.1.3.

Costs per Delayed Customer Order. In the event that a singular penalty will be incurred if an order is delayed we obtain the following cost function:

$$E[C_{31}(S)] = c_2 \cdot E[I] + c_{31} \cdot (1 - \alpha^{\tau = r_D}) \tag{6.86}$$

Resolving $C_{31}(S)$ according to (6.34) and (6.27) leads to expression (6.87):

$$\begin{aligned}
&E[C_{31}(S)] \\
&\quad = \sum_{t^* \in T^*} \sum_{l \in L} \frac{P\{L = l\}}{r} \cdot c_2 \cdot E[I | T^* = t^*, L = l] \\
&\qquad + c_{31} \left(1 - \sum_{t^* \in T^*} \sum_{l \in L} \frac{P\{L = l\}}{r} \cdot (\alpha^{\tau = r_D} | T^* = t^*, L = l) \right) \\
&\quad = \sum_{t^* \in T^*} \sum_{l \in L} \frac{P\{L = l\}}{r} \cdot c_2 \cdot E[I | T^* = t^*, L = l] \\
&\qquad + c_{31} \sum_{t^* \in T^*} \sum_{l \in L} \frac{P\{L = l\}}{r} \cdot (1 - \alpha^{\tau = r_D} | T^* = t^*, L = l)
\end{aligned}$$

$$= \sum_{t^*\in T^*} \sum_{l\in L} \frac{P\{L=l\}}{r} \cdot \big(c_2 \cdot E[I|T^*=t^*, L=l]$$

$$+ c_{31} \cdot (1-\alpha^{\tau=r_D}|T^*=t^*, L=l)\big) \tag{6.87}$$

For simplicity, let us look at the corresponding derivatives of $E[I|T^* = t^*, L = l]$ and $1-\alpha^{\tau=r_D}|T^* = t^*, L = l$ separately. For the mean inventory levels, we apply the fundamental theorem of calculus to get the derivative of an integral expression:

$$E[I|T^*=t^*, L=l] = S - \int_{-\infty}^{S} PDF_{D[nf_1(t^*+l,r_D)]}(x) \cdot x\, dx$$

$$- (1 - CDF_{D[nf_1(t^*+l,r_D)]}(S)) \cdot S \tag{6.88}$$

$$\frac{\delta E[I|T^*=t^*, L=l]}{\delta S} = 1 - PDF_{D[nf_1(t^*+l,r_D)]}(S) \cdot S - 1$$

$$+ PDF_{D[nf_1(t^*+l,r_D)]}(S) \cdot S + CDF_{D[nf_1(t^*+l,r_D)]}(S)$$

$$= CDF_{D[nf_1(t^*+l,r_D)]}(S) \tag{6.89}$$

$$1-\alpha_{\tau=r_D}|T^*=t^*, L=l = 1 - CDF_{D[nf_1(t^*+l,r_D)]}(S)$$

$$\frac{\delta(1-\alpha_{\tau=r_D}|T^*=t^*, L=l)}{\delta S} = -PDF_{D[nf_1(t^*+l,r_D)]}(S) \tag{6.90}$$

Thus, we obtain the following derivative of (6.86):

$$\frac{\delta C(S)}{\delta S} = \sum_{t^*\in T^*} \sum_{l\in L} \frac{P\{L=l\}}{r}$$

$$\cdot \Big(c_2 \cdot CDF_{D[nf_1(t^*+l,r_D)]}(S) - c_{31} \cdot PDF_{D[nf_1(t^*+l,r_D)]}(S)\Big) \tag{6.91}$$

To find the value S that minimizes (6.86), we have to solve the following equation and make sure that the second derivative $\frac{\delta^2 C(S)}{\delta S^2}$ is greater than zero for our solution:

$$0 = \sum_{t^*\in T^*} \sum_{l\in L} \frac{P\{L=l\}}{r}$$

$$\cdot \Big(c_2 \cdot CDF_{D[nf_1(t^*+l,r_D)]}(S) - c_{31} \cdot PDF_{D[nf_1(t^*+l,r_D)]}(S)\Big) \tag{6.92}$$

Unfortunately there are two pieces of bad news about (6.92). First, we cannot solve it to a closed expression, and second, we may have multiple solutions. Hence, we may not (always) apply some general approach like the bisection method or Newton's method. Instead, we propose the following two-phase approach. First, identify an interval around the minimum solution of (6.87) that contains only one solution to

(6.92). Second, approximate the solution of (6.92) within a certain tolerance t using the bisection method. Let *dis* be the minimum distance on the x-axis between two solutions of (6.92). Then we may apply Algorithm 12 to solve the first problem, where the selection of *dis* depends on the distribution of demand per period. The bisection method should then be applied to the interval $[S_{opt} - st, S_{opt} + st]$.

Algorithm 12: Simple algorithm to approximate S_{opt}

Input: Minimum distance *dis*
Output: Approximation of optimum order-up-to level S_{opt}
$S_{lb} := 0;$
$S_{ub} := D^{[l_{max}+r-1]}.getQuantile(0.9999);$
$st = \frac{dis}{2};$
$S_{opt} := S_{lb};$
for $i := st; i \leq S_{ub}; i = i + st$ **do**
  if $getCosts(S_{opt}) > getCosts(i)$ **then**
    $S_{opt} := i;$
  end
end

Obviously, Algorithm 12 exhibits a runtime given via $CC_{a12} = O(\frac{S_{ub}}{dis} \cdot CC_{CO} + \log_2(\frac{dis}{t}) \cdot CC_{SL})$, where CC_{CO} and CC_{SL} are the runtimes for evaluating (6.87) and (6.92), respectively.

Costs per Unit and Time. If we account for backorders per unit and time, we obtain the following cost function:

$$E[C_{33}(S)] = c_2 \cdot E[I] + c_{33} \cdot E[B] \tag{6.93}$$

Resolving $C_{33}(S)$ according to (6.27) and (6.34) leads to the following expression:

$$\begin{aligned} C_{33}(S) = \sum_{t^* \in T^*} \sum_{l \in L} \frac{1}{r} \cdot P\{L = l\} \cdot \Big(c_2 \cdot E[I|T^* = t^*, L = l] \\ + c_{33} \cdot E[B|T^* = t^*, L = l]\Big) \end{aligned} \tag{6.94}$$

For $E[I|T^* = t^*, L = l]$, we already have the derivative with (6.89). To find the derivative of $E[B|T^* = t^*, L = l]$, let us first restate equation (6.25), where we assume that we will always fulfill $t < l + r$ by the order of summation later on:

$$\begin{aligned} &E[B|T^* = t^*, L = l] \\ &\quad = \int_S^\infty PDF_{D^{[nf_1(t^*+l, r_D)]}}(x) \cdot (x - S)\, dx \end{aligned}$$

$$= \int_{-\infty}^{\infty} PDF_{D^{[nf_1(t^*+l,r_D)]}}(x) \cdot (x - S)\, dx$$
$$- \int_{-\infty}^{S} PDF_{D^{[nf_1(t^*+l,r_D)]}}(x) \cdot (x - S)\, dx$$
$$= E[D^{[nf_1(t^*+l,r_D)]}] - S$$
$$- \int_{-\infty}^{S} PDF_{D^{[nf_1(t^*+l,r_D)]}}(x) \cdot x\, dx + CDF_{D^{[nf_1(t^*+l,r_D)]}}(S) \cdot S \tag{6.95}$$

Thus, we have the following derivative:

$$\begin{aligned}\frac{\delta E[B|T^* = t^*, L = l]}{\delta S} &= 0 - 1 - PDF_{D^{[nf_1(t^*+l,r_D)]}}(S) \cdot S \\ &\quad + CDF_{D^{[nf_1(t^*+l,r_D)]}}(S) \\ &\quad + PDF_{D^{[nf_1(t^*+l,r_D)]}}(S) \cdot S \\ &= CDF_{D^{[nf_1(t^*+l,r_D)]}}(S) - 1 \end{aligned} \tag{6.96}$$

In combination, we have the following derivative for the overall cost function:

$$\begin{aligned}\frac{\delta C_{33}(S)}{\delta S} &= \sum_{t^* \in T^*} \sum_{l \in L} \frac{1}{r} \cdot P\{L = l\} \cdot (c_2 \cdot CDF_{D^{[nf_1(t^*+l,r_D)]}}(S) \\ &\quad + c_{33} \cdot (CDF_{D^{[nf_1(t^*+l,r_D)]}}(S) - 1)) \\ &= \sum_{t^* \in T^*} \sum_{l \in L} \frac{1}{r} \cdot P\{L = l\} \cdot ((c_2 + c_{33}) \\ &\quad \cdot CDF_{D^{[nf_1(t^*+l,r_D)]}}(S) - c_{33}) \end{aligned} \tag{6.97}$$

To find the optimum S, we have to solve equation (6.98):

$$\frac{c_{33}}{c_2 + c_{33}} = \sum_{t^* \in T^*} \sum_{l \in L} \frac{1}{r} \cdot P\{L = l\} \cdot CDF_{D^{[nf_1(t^*+l,r_D)]}}(S) \tag{6.98}$$

As the right side of the sum obviously monotonously rises in S, equation (6.98) can be solved by the methods described in Sect. 6.4.1.1 to find the optimum S, subject to a service level constraint.

Costs per Unit. In the event that we account for backorder costs only once they occur, we have to consider the following cost function:

$$E[C_{32}(S)] = c_2 \cdot E[I] + c_{32} \cdot E[B^{*,\tau=1}] \tag{6.99}$$

Since we already know the derivative of $E[I]$ from (6.89), we focus on $E[B^{*,\tau=1}]$, which corresponds to an inventory holding cost rate per period:

$$E[B^{*,\tau=1}] = \sum_{t^* \in T^*} \sum_{l \in L} \frac{1}{r} \cdot P\{L = l\} \cdot \Big(\int_S^\infty PDF_{D[nf_1(t^*+l, r_D)]}(x) \cdot (x - S)\, dx - \int_S^\infty PDF_{D[nf_1(t^*+l-1, r_D)]}(x) \cdot (x - S)\, dx \Big) \tag{6.100}$$

From the derivative of $E[B|T^* = t^*|L = l]$ as stated above, we can directly deduce the derivative of $E[B^{*,\tau=1}|T^* = t^*|L = l]$:

$$\begin{aligned} \frac{\delta E[B^{*,\tau=1}|T^* = t^*|L = l]}{\delta S} &= CDF_{D[nf_1(t^*+l, r_D)]}(S) - 1 \\ &\quad - CDF_{D[nf_1(t^*+l-1, r_D)]}(S) + 1 \\ &= CDF_{D[nf_1(t^*+l, r_D)]}(S) \\ &\quad - CDF_{D[nf_1(t^*+l-1, r_D)]}(S) \end{aligned} \tag{6.101}$$

Thus, to find the optimum S, we have to solve the following equation:

$$0 = \sum_{t^* \in T^*} \sum_{l \in L} \frac{1}{r} \cdot P\{L = l\} \cdot \Big((c_2 + c_{32}) \cdot CDF_{D[nf_1(t^*+l, r_D)]}(S) - c_{32} \cdot CDF_{D[nf_1(t^*+l-1, r_D)]}(S) \Big) \tag{6.102}$$

Again, bisection appears to be a safe method to solve equation (6.102). It is somewhat intuitive, yet hard to prove that there is always just one solution.

6.4.1.4 Set *r* and *S* Reflecting Backorder Costs

When incorporating ordering costs into the calculus, we obtain the following objective function:

$$C(r, S) = \frac{c_1}{r} + C(S) \tag{6.103}$$

Based on the findings of Sect. 6.4.1.3 and assuming that (6.103) is convex in r, we may apply the same algorithmic approaches as described in Sect. 6.4.1.2. The assumption of convexity has the same limitations as mentioned for the service level approach in Sect. 6.4.1.2.

6.4.2 Full Deliveries

6.4.2.1 Set S Subject to Service

If we only want to determine a minimum S that accomplishes a certain target service level, we may apply the same procedure as described in Sect. 6.4.1.1 for the case of full deliveries. For the ready rate it does not make any difference which delivery modality is applied. For the other two service criteria, it analogously holds that rising S may or may not improve the service, but it definitely cannot lower it. Thus, even here the service level functions are monotonic in S, and we may apply the bisection method to approximate the required S.

6.4.2.2 Set r and S Subject to Service

When r and S are to be optimized in terms of purchase costs and holding costs while a specific service level is to be accomplished, we observe the same phenomena as for the case of split deliveries (see Sect. 6.4.1.2). Thus, the same resolution approaches may be applied within the limits mentioned above.

6.4.2.3 Set S Reflecting Backorder Costs

To evaluate the cost functions (6.86), (6.93) and (6.99) that we introduced in Sect. 6.4.1.3, the following formulae are relevant for the full deliveries case.

For the expected average inventory, we have (6.60) and the expected delayed orders per period are given by $\frac{1}{r_D} \cdot (1-\alpha^{\tau=r_D})$ as above. The two types of backorders are calculated according to (6.54) and (6.56), respectively.

However, no matter what type of backorder costs $(c_{31}, c_{32}$ or $c_{33})$ we are considering here, we generally observe that the resulting cost functions have a non-convex shape. Even worse, we are not able to analytically determine their slope. For example, finding the preliminary derivative of the mean inventory in S (6.60) seems a hopeless task.

Therefore, we can only propose a simple linear approach to find the best solution within the limits of a certain step granularity. This may be done by Algorithm 12, where the bisection method should not be applied any further.

6.4.2.4 Set r and S Reflecting Backorder Costs

For setting both parameters r and S, we obtain good results with the same approach as described in Sect. 6.4.1.4. However, we observe significantly longer runtimes than for equivalent split order instances, which is due to the time-consuming basic evaluation of inventory levels and backorder costs for certain parameterizations.

Since the optimum order cycles for the split and full delivery mode tend to be similar or even equal in many instances, a possible way to speed up the procedure could be the following approximation method. Instead of the formulae that consider full deliveries, use those for the split deliveries case to approximate r_{opt}. Then switch to the correct formulae, see if $r_{opt} - 1$ or $r_{opt} + 1$ exhibit lower costs, and finally set S.

Chapter 7
Conclusion

This study is dedicated to the management of inventories with periodic review inventory policies. We have tried to provide the reader with a theoretical and methodological background as well as tangible results for special problems that are ready to implement. In order to achieve this, we started with the basic vocabulary as well as an overview on basic properties of inventory systems and replenishment rules in Chap. 2. The relevant literature was then discussed in Chap. 3, where we pursued two aims. One was to try and provide a comprehensive overview of problems concerning single-level inventory management that have already been solved in the literature. We also attempted to give an introduction to a general mode of thinking and encountering problems in inventory theory, paying special attention to periodic review systems. Before we began to study a specific class of inventory systems, Chap. 4 elaborated on selected elementary methods that we find useful for analyzing inventory systems in general.

Chapter 5 considered four forms of replenishment lead time processes that are commonly found in the literature. We dedicated significant space to the case of independent lead times and the possibility of order crossover. We did so for two reasons. First, the corresponding model – despite its theoretical and practical relevance – has not been very well studied in the literature. Second, the problems and analysis presented provide a practicable introduction to the analysis of inventory systems with related properties. We consider these problems especially suitable for starting with because their focus is merely on one single property, whereas multiple aspects are typically involved even in seemingly simple problems found in inventory theory.

In Chap. 6, we finally attempted to consolidate the background given in the previous sections by addressing and solving a set of problems related to an inventory system that is typically found within a multi-level supply chain. We dedicated most of the chapter to the analytical evaluation of a certain set of properties in order to then be able to construct the corresponding optimization routines on the stable foundations explained previously.

T. Wensing, *Periodic Review Inventory Systems*, Lecture Notes in Economics and Mathematical Systems 651, DOI 10.1007/978-3-642-20479-1_7,

The contribution of this study to theory and practical applications is as follows. In the first place, it contributes to inventory theory by offering novel analytical approaches for several properties of inventory models being considered. To our knowledge, the effective lead time (Sect. 5.2.3) has not been analyzed in this form before. Furthermore, analysis of customer waiting times per order (Sect. 6.3.1.2), consideration of full deliveries (Sect. 6.3.3) for volume-oriented performance indicators and the entire analysis of the independent lead time case (Sects. 6.3.4, 6.3.5 and 6.3.6) have not been conducted for this particular inventory model.

Furthermore, the study contributes to practical inventory management by providing formulae and algorithms that are ready to implement and may help practitioners to adjust related real-world inventory systems. In that context, we put emphasis on designing a modular system, i.e., a system that may easily be enhanced, and which may also be embedded into a larger context of further similar systems that either provide input to it or receive output from the system that is being examined. Not least, the scope of this study was to design an extendable model and give insights into the general methodological possibilities for the task of analyzing periodic review inventory systems of the type examined. This approach follows the general trait of contemporary inventory theory that provides specific – possibly adaptable – methods instead than aiming at a single generic model to meet the requirements of any practical problem.

Future research may be directed towards the further improvement of the model's embeddedness into a multi-level context. Besides this, the customer demand process may be subject to deeper analysis, probably in a similar manner to how we conducted it for the replenishment lead time process. Concerning the first aspect, the model is ready to be used in a linear multi-level context if we consider the customer waiting time of one system as replenishment lead time for the successive system. This approach, however, is only exact for constant customer demands. The case of stochastic demand introduces the problem aspect that customer waiting times are in fact not independent of the customer demand process. Higher customer order volume may clearly induce a higher stockout probability than a lower volume. Therefore the customer waiting time as developed above does not provide sufficient information to exactly analyze the successive inventory system. However, experiments with this approach revealed quite accurate results that decline the more stocks are considered and the more volatile the demand distribution.

Furthermore, the model may be enhanced to cover even more general supply chain or rather supply network structures, i.e., one may allow for multiple successive demand and preceding replenishment processes. Besides that, one may consider that the system being analyzed provides material in combination with other similar systems. For the latter case of converging substructures, it is then necessary to determine the combined delay or service of several systems that serve the same successor. Here, one may distinguish between the obligation of material availability at all preceding systems and the possibility that one system's shortage may be compensated by another that keeps the same goods in stock. The complementary case of diverging structures comes with the necessity to specify a rule to decide which of several successors will be served first if material is insufficient to fulfill the entire demand.

In terms of the demand aspect, it would be also interesting to model the interarrival times as random variables, which would be a generalization of our demand model. Furthermore, the assumption of i.i.d. demand occurrences may not be realistic for many practical problem instances. A simple approach in that context that follows our model's logic would be to replace the classical convolution by a different approach that reflects the corresponding dependencies when determining the combined demand of several periods.

References

Aggarwal, S. (1974). A review of current inventory theory and its application. *International Journal of Production Research*, 12(4):443–482.

Agrawal, V. and Seshadrin, S. (2000). Distribution free bounds for service constrained (q, r) inventory systems. *Naval Research Logistics*, 36(8):635–656.

Akinniyi, F. and Silver, E. (1981). Inventory control using a service constraint on the expected duration of stockouts. *AIIE Transactions*, 13(4):343–348.

Archibald, B. (1981). Continuous review (s,S) policies with lost sales. *Management Science*, 27(10):1171–1177.

Archibald, B. and Silver, E. (1978). (s,S) policies under continuous review and discrete compound poisson demand. *Management Science*, 24(9):899–909.

Arrow, K., Harris, T., and Marshak, J. (1951). Optimal inventory policy. *Econometrica*, 19(3): 250–272.

Atkins, D. and Iyogun, P. (1988). Periodic versus can-order policies for coordinated multi-item inventory systems. *Management Science*, 34(6):791–796.

Axsäter, S. (2006). *Inventory Control*. Springer, New York, 2nd edition.

Bagchi, U., Hayya, J., and Ord, J. (1984). Modeling demand during lead time. *Decision Sciences*, 15(2):157–176.

Bashyam, S. and Fu, M. C. (1998). Optimization of (s, S) inventory systems with random lead times and a service level constraint. *Management Science*, 44(12/2):175–190.

Bather, J. (1966). A continuous time inventory model. *Journal of Applied Probability*, 3(2): 538–549.

Beckmann, M. (1962). An inventory model for arbitrary interval and quantity distributions of demand. *Management Science*, 8(1):35–57.

Bell, C. (1970). Improved algorithms for inventory and replacement-stocking problems. *SIAM Journal on Applied Mathematics*, 18(3):558–566.

Benkherouf, L. (2008). On the optimality of (s,S) inventory policies: a quasivariational approach. *Journal of Applied Mathematics and Stochastic Analysis*, 2008:1–9.

Bensoussan, A., Liu, R., and Sethi, S. (2006). Optimality of an (s,S) policy with compound poisson and diffusion demands: a QVI approach. *SIAM Journal on Control and Optimization*, 44(5):1650–1676.

Berk, E. and Gürler, Ü. (2008). Analysis of the (q, r) inventory model for perishables with positive lead times and lost sales. *Operations Research*, 56(5):1238–1246.

Boyaci, T. and Gallego, G. (2001). Minimizing holding and ordering costs subject to a bound on backorders is as easy as solving a single backorder cost model. *Operations Research Letters*, 29(4):187–192.

T. Wensing, *Periodic Review Inventory Systems*, Lecture Notes in Economics and Mathematical Systems 651, DOI 10.1007/978-3-642-20479-1,

Boyaci, T. and Guillermo, G. (2002). Managing waiting times of backordered demands in single-stage (q,r) inventory systems. *Naval Research Logistics*, 49(6):557–573.

Boylan, J. and Johnston, F. (1994). Relationships between service level measures for inventory systems. *Journal of the Operational Research Society*, 45(7):838–844.

Bradley, J. and Robinson, L. (2005). Improved base-stock approximations for independent stochastic lead times with order crossover. *Manufacturing and Service Operations Management*, 7(4):319–329.

Brooking, S. (1987). Inventory system costs: Source data for analysis. *Engineering Costs and Production Economics*, 13(1):1–12.

Browne, S. and Zipkin, P. H. (1991). Inventory models with contiuous stochastic demands. *The Annals of Applied Probability*, 1(3):419–435.

Burden, R. and Faires, J. (2005). *Numerical Analysis*. Brooks Cole, Pacific Grove, 8th edition.

Cardós, M., Miralles, C., and Ros, L. (2006). An exact calculation of the cycle service level in a generalized periodic review system. *Journal of the Operational Research Society*, 57(10):1252–1255.

Charnes, J. M., Marmorstein, H., and Zinn, W. (1995). Safety stock determination with serially correlated demand in a periodic-review inventory system. *Journal of the Operational Research Society*, 46(8):1006–1013.

Chen, F. (2000). Optimal policies for multi-item echelon inventory policies with batch ordering. *Operations Research*, 48(3):376–389.

Chen, F. and Krass, D. (2001). Inventory models with minimal service level constraints. *European Journal of Operational Research*, 134(1):120–140.

Chen, F. and Zheng, Z.-S. (1992). Waiting time distribution in (t,s) inventory systems. *Operations Research Letters*, 12(3):145–151.

Chiang, C. (2006). Optimal ordering policies for periodic-review systems with replenishment cycles. *European Journal of Operational Research*, 170(1):44–46.

Chiang, C. (2007). Optimal ordering policies for periodic-review systems with a refined intra-cycle time scale. *European Journal of Operational Research*, 177(2):872–881.

Chung, F., Graham, R., Hoggatt, V., and Kleiman, M. (1978). The number of Baxter permutations. *Journal of Combinatorical Theory Series A*, (24):382–394.

Clark, A. and Scarf, H. (1960). Optimal policies for a multi-echelon inventory problem. *Management Science*, 6(4):475–490.

Cohen, M. A., Kleindorfer, P. R., and Lee, H. L. (1989). Near-optimal service constrained stocking policies for spare parts. *Operations Research*, 37(1):104–117.

Cohen, M. A., Kleindorfer, P. R., Lee, H. L., and Pyke, D. F. (1992). Multi-item service constrained (s, S) policies for spare parts logistics systems. *Naval Research Logistics*, 39(4):561–577.

Das, C. (1976). Approximate solution to the (q,r) inventory model for Gamma lead time demand. *Management Science*, 22(9):1043–1047.

Das, C. (1977). The (S-1,S) inventory model under time limit on backorders. *Operations Research*, 25(5):835–850.

Das, C. (1983). Q,r inventory models with time-weighted backorders. *Journal of the Operational Reserch Society*, 34(5):401–412.

Davis, P. J. and Rabinowitz, P. (1984.). *Methods of Numerical Integration*. Academic Press, New York, 2nd edition.

Dominey, M. and Hill, R. (2004). Performance of approximations for compound Poisson distributed demand in the newsboy problem. *International Journal of Production Economics*, 92(2):145–155.

Dunsmuir, W. T. M. and Snyder, R. (1989). Control of inventories with intermittent demand. *European Journal of Operational Research*, 40(1):16–21.

Dvoretzky, A., Kiefer, J., and J.Wolfowitz (1953). On the optimal character of the (S,s) policy in inventory theory. *Econometrica*, 21(4):586–596.

Ehrhardt, R. (1979). The power approximation for computing (s,S) inventory policies. *Management Science*, 25(8):777–786.

Ehrhardt, R. (1984). (s,S) policies for a dynamic inventory model with stochastic lead times. *Operations Research*, 32(1):121–132.

Ehrhardt, R. and Mosier, C. (1984). A revision of the power approximation for computing (s, S) policies. *Management Science*, 30(5):618–622.

Eppen, G. and Martin, R. (1988). Determining safety stock in the presence of stochastic lead time and demand. *Management Science*, 34(11):1380–1390.

Federgruen, A. and Katalan, Z. (1994). Approximating queue size and waiting time distributions in general polling systems. *Queueing Systems*, 18(3-4):353–386.

Federgruen, A. and Zheng, Y. (1992). An efficient algorithm for computing an optimal (r,q) policy in continuous review stochastic inventory systems. *Operations Research*, 40(4):808–813.

Federgruen, A. and Zipkin, P. H. (1984a). Computational issues in an infinite-horizon, multiechelon inventory model. *Operations Research*, 32(4):818–836.

Federgruen, A. and Zipkin, P. H. (1984b). An efficient algorithm for computing optimal (s,S) policies. *Operations Research*, 32(6):1268–1285.

Feeney, G. and Sherbrooke, C. (1966). The (S-1,S) inventory policy under compound poisson demand. *Management Science*, 12(6):391–411.

Feng, Q., Sethi, S., Yan, H., and Zhang, H. (2006a). Are base-stock policies optimal in inventory problems with multiple delivery modes? *Operations Research*, 54(4):801–807.

Feng, Q., Sethi, S., Yan, H., and Zhang, H. (2006b). Optimality and nonoptimality of the base-stock policy in inventory problems with multiple delivery modes. *Journal of Industrial and Management Optimization*, 2(1):19–42.

Feng, Y. and Xiao, B. (2000). A new algorithm for computing optimal (s,S) policies in a stochastic single item/location inventory system. *IIE Transaction*, 32(11):1081–1090.

Fischer, L. (2008). *Bestandsoptimierung für das Supply Chain Management*. Books on Demand GmbH, Norderstedt.

Fong, D., Gempesaw, V., and Ord, J. (2000). Analysis of a dual sourcing inventory model with normal unit demand and Erlang mixture lead times. *European Journal of Operational Research*, 120(1):97–107.

Freeland, J. and Porteus, E. (1980). Evaluating the effectiveness of a new method for computing approximately optimal (s, S) inventory policies. *Operations Research*, 28(2):353–364.

Fung, R., Ma, X., and Lau, H. (2001). (t,s) policy for coordinated inventory replenishment systems under compound Poisson demands. *Production planning and control*, 12(6):575–583.

Gallego, G. (1998). New bounds and heuristics for (q,r) policies. *Management Science*, 44(2): 219–233.

Galliher, H., Morse, P., and Simond, M. (1959). Dynamics of two classes of continuous review inventory systems. *Operations Research*, 7(3):362–384.

Hacking, I. (2001). *An introduction to probability and inductive logicy*. Cambridge University Press, Cambridge.

Hacking, I. (2006). *The Emergence of Probability*. Cambridge University Press, Cambridge, 2nd edition.

Hadley, G. and Whitin, T. M. (1963). *Analysis of Inventory Systems*. Prentice Hall, Englewood Cliffs, N.J.

Hadley, G. and Whitin, T. M. (1962). A family of dynamic inventory models. *Management Science*, 8(4):458–469.

Haehling von Lanzenauer, C. (1988). Service-level constraints measuring shortages during a fiscal period: the normal distribution. *The Journal of the Operational Research Society*, 39(3): 299–304.

Haehling von Lanzenauer, C. and Hamidi-Noori, A. (1986). Fiscal period based service level constraints and safety stock requirements. *International Journal of Production Research*, 24(3):483–492.

Hausman, W., Lee, H., and Zhang, A. (1998). Joint demand fulfillment probability in a multi-item inventory system with independent order-up-to policies. *European Journal of Operational Research*, 109(3):646–659.

Hayya, J., Bagchi, U., Kim, J., and Sun, D. (2008). On static stochastic order crossover. *International Journal of Production Economics*, 114(1):404–413.

Hayya, J., Xu, S., Ramasesh, R., and He, X. (1995). Order crossover in inventory systems. *Stochastic Models*, 11(2):279–309.

He, X., Kim, J., and Hayya, J. (2005). The cost of lead-time variability: The case of the exponential distribution. *International Journal of Production Economics*, 97(2):130–142.

He, X., Xu, S., Ord, K., and Hayya, J. (1998). An inventory model with order crossover. *Operations Research*, 46(3):S112–S119.

Heuts, R. and de Klein, J. (1995). An (s, q) inventory model with stochastic and interrelated lead times. *Naval Research Logistics*, 42(5):839–859.

Higa, I., Feyerherm, A., and Machado, A. (1975). Waiting time in an (S-1,S) inventory system. *Operations Research*, 23(4):674–680.

Hill, R. (1999). On the suboptimality of (S-1, S) lost sales inventory policies. *International Journal of Production Economics*, 59:387–393.

Hollier, R. and Vrat, P. (1978). A proposal for classification of inventory systems. *Omega*, 6(3):277–279.

Hordijk, A. and van der Duyn Schouten, F. (1986). On the optimality of (s,S)-policies in continuous review inventory models. *SIAM Journal on Applied Mathematics*, 46(5):912–929.

Iglehart, D. (1963). Optimality of (s,S) policies in the infinite horizon dynamic inventory problem. *Management Science*, 9(2):259–267.

Iyer, A. and Schrage, L. (1992). Analysis of the deterministic (s,S) inventory problem. *Management Science*, 38(9):1299–1313.

Janssen, F., Heuts, R., and de Kok, T. (1998). On the (R, s, Q) inventory model when demand is modelled as compound Bernoulli process. *European Journal of Operational Research*, 104(3):423–436.

Johansen, S. and Hill, R. (2000). The (r,Q) control of a periodic-review inventory system with continuous demand and lost sales. *International Journal of Production Economics*, 68(3): 279–286.

Johansen, S. and Thorstenson, A. (1993). Optimal and approximate (Q, r) inventory policies with lost sales and Gamma-distributed lead time. *International Journal of Production Economics*, 30–31(1):179–194.

Johansen, S. and Thorstenson, A. (1996). Optimal (r, Q) inventory policies with Poisson demands and lost sales: discounted and undiscounted cases. *International Journal of Production Economics*, 46–47:359–371.

Johnson, E. (1967). Optimality and computation of (σ, S) policies in the multi item-infinite horizon inventory problem. *Management Science*, 13(7):475–491.

Johnson, E. (1968). On (s,S) policies. *Management Science*, 15(1):80–101.

Johnson, M. E., Lee, H. L., Davis, T., and Hall, R. (1995). Expressions for item fill rates in periodic inventory systems. *Naval Research Logistics*, 42(1):57–80.

Johnson, N., Kotz, S., and Balakrishnan, N. (1994). *Continuous Univariate Distributions Volume 1*. Wiley, New York, 2nd edition.

Kalin, D. (1980). On the optimality of (σ, S) policies. *Mathematics of Operations Research*, 5(2):475–491.

Kalpakam, S. and Sapna, K. (1994). Continuous review (s, S) inventory system with random lifetimes and positive leadtimes. *Operations Research Letters*, 16(2):115–119.

Kaplan, R. (1970). A dynamic inventory model with stochastic lead times. *Management Science*, 16(7):491–507.

Karlin, S. (1958). One stage inventory models with uncertainty. In Arrow, K., Karlin, S., and Scarf, H., editors, *Studies in the Mathematical Theory of Inventory and Production*. Stanford University Press, Stanford.

Karlin, S. (1960). Dynamic inventory policy with varying stochastic demands. *Management Science*, 6(3):231–258.

Kelle, P. and Silver, E. (1990). Safety stock reduction by order splitting. *Naval Research Logistics*, 37(5):725–743.

Kiesmüller, G. and de Kok, A. (2006). The customer waiting time in an (R, s, Q) inventory system. *International Journal of Production Economics*, 104:354–364.

Kim, D. and Park, K. (1985). (Q, r) inventory model with a mixture of lost sales and time-weighted backorders. *Journal of the Operational Research Society*, 36(3):231–238.

Knuth, D. E. (1997). *The Art of Programming, Volume 3: Sorting and Searching*. Adisson Wesley, Reading.

Krishnamoorthy, A. and Islam, M. (2004). (s,S) inventory system with postponed demands. *Stochastic Analysis and Applications*, 22(3):827–842.

Kruse, W. (1980). Waiting time in an S-1,S inventory system with arbitrarily distributed lead times. *Operations Research*, 28(2):348–352.

Kruse, W. (1981). Waiting time in a continuous-review (S-1,S) inventory system with constant lead time. *Operations Research*, 29(1):202–207.

Küenle, C. and Küenle, H. (1977). Durchschnittsoptimale Strategien in Markovschen Entscheidungsmodellen bei unbeschränkten Kosten. *Optimization*, 8(4):549–564.

Larsen, C. and Thorstenson, A. (2008). A comparison between the order and the volume fill rate for a base-stock inventory control system under a compound renewal demand process. *Journal of the Operational Research Society*, 59(6):798–804.

Larson, C. and Kiesmüller, G. (2007). Developing a closed-form cost expression for an (r, s, nq) policy where the demand process is compound generalized Erlang. *Operations Research Letters*, 35:567–572.

Lee, H. and Nahmias, S. (1993). Single-product, single-location models. In Graves, S., Kan, A. R., and Zipkin, P. H., editors, *Handbooks in Operations Research and Management Science, Volume 4: Logistics of Production and Inventory*. North-Holland, Amsterdam.

Lee, H.-S. (1995). On continuous review stochastic (s,S) inventory systems with ordering delays. *Computers and Industrial Engineering*, 28(4):763–771.

Liberatore, M. (1979). The EOQ model under stochastic lead time. *Operations Research*, 27(2):391–396.

Liu, L. (1990). (s,S) continuous review models for inventory with random lifetimes. *Operations Research Letters*, 9(3):161–167.

Merlini, D. and Verri, M. C. (2000). Generating trees and proper Riordan arrays. *Discrete Mathematics*, 218(1–3):167–183.

Meyn, S. and Tweedie, R. (2009). *Markov Chains and Stochastic Stability*. Cambridge University Press, Cambridge, 2nd edition.

Minner, S. (2003). Multiple-supplier inventory models in supply chain management: A review. *International Journal of Production Economics*, 81–82:265–279.

Moors, J. J. A. and Strijbosch, L. W. G. (2002). Exact fill rates for (R, s, S) inventory control with Gamma distributed demand. *Journal of the Operational Research Society*, 53(11):1268–1274.

Morton, T. (1978). The non-stationary infinite horizon inventory problem. *Management Science*, 24(14):1474–1482.

Naddor, E. (1975). Optimal and heuristic decisions in single- and multi-item inventory systems. *Management Science*, 21(11):1234–1249.

Nahmias, S. (1979). Simple approximations for a variety of dynamic leadtime lost-sales inventory models. *Operations Research*, 27(5):904–924.

Norman, J. M. and White, D. J. (1968). A method for approximate solutions to stochastic dynamic programming problems using expectations. *Operations Research*, 16(2):296–306.

Ord, J. and Bagchi, U. (1983). The truncated normal-Gamma mixture as a distribution for lead time demand. *Naval Research Logistics*, 30(2):359–365.

Pam, A., Ramasesh, R., Hayya, J., and Ord, J. (1991). Multiple sourcing: The determination of lead times. *Operations Research Letters*, 10(1):1–7.

Porteus, E. (1971). On the optimality of generalized (s,S) policies. *Management Science*, 17(7):411–426.

Porteus, E. (1972). The optimality of generalized (s,S) policies under uniform demand densities. *Management Science*, 18(11):644–646.

Porteus, E. (1985). Numerical comparisons of inventory policies for periodic review systems. *Operations Research*, 33(1):134–152.

Prasad, S. (1994). Classification of inventory models and systems. *International Journal of Production Economics*, 34(2):209–222.

Ramasesh, R., Ord, J., Hayya, J., and Pan, A. (1991). Sole versus dual sourcing in stochastic lead-time (s,q) inventory models. *Management Science*, 37(4):428–443.

Rao, U. (2003). Properties of the periodic review (r,t) inventory control policy for stationary stochastic demand. *Manufacturing and Service Operations Management*, 5(1):37–53.

Riezebos, J. (2006). Inventory order crossovers. *International Journal of Production Economics*, 104(2):666–675.

Robinson, L., Bradley, J., and Thomas, L. (2001). Consequences of order crossover under order-up-to inventory policies. *Manufacturing & Service Operations Management*, 3(3):175–188.

Robinson, L. and Bradley, R. (2008). Further improvements on base-stock approximations for independent stochastic lead times with order crossover. *Manufacturing & Service Operations Management*, 10(2):325–327.

Ronen, D. (1982). Measures of product availability. *Journal of Business Logistics*, 3(1):45–58.

Ross, M. (2006). *Introduction to Probability Models*. 9 edition.

Sahin, I. (1982). On the objective function behavior in (s,S) inventory models. *Operations Research*, 30(4):709–724.

Sahin, I. (1988). Optimality conditions for regenerative inventory systems under batch demands. *Applied Stochastic Models and Data Analysis*, 4(3):173–183.

Sahin, I. and Kilari, P. (1984). Performance of an approximation to continuous review (S,s) policies under compound renewal demand. *International Journal of Production Research*, 22(6):1027–1032.

Sahin, I. and Sinha, D. (1987). Renewal approximation to optimal order quantity for a class of continuous-review inventory systems. *Naval Research Logistics*, 34(5):655–667.

Sani, B. and Kingsman, B. (1997). Selecting the best periodic inventory control and demand forcasting methods for low demand items. *Journal of the Operational Research Society*, 48(7):700–713.

Scarf, H. (1960). The optimaltity of (s,S) policies in the dynamic inventory problem. In Arrow, K., Karlin, S., and Suppes, P., editors, *Mathematical Methods in the Social Sciences*. Stanford University Press, Stanford.

Schäl, M. (1976). On the optimality of (s,S)-policies in dynamic inventory models with finite horizon. *SIAM Journal on Applied Mathematics*, 30(3):528–537.

Schneider, H. (1978). Methods for determining the re-order point of an (s,S) ordering policy when a service level is specified. *Journal of the Operational Research Society*, 29(12):1181–1193.

Schneider, H. (1981). Effect of service-levels on order-points or order-levels in inventory models. *International Journal of Production Research*, 19(6):615–631.

Schneider, H. and Ringuest, J. L. (1990). Power approximation for computing (s, S) policies using service level. *Management Science*, 36(7):822–834.

Sculli, D. and Shum, Y. (1990). Analysis of a continuous review stock-control model with multiple suppliers. *Journal of the Operational Research Society*, 41(9):873–877.

Sculli, D. and Wu, S. (1981). Stock control with two suppliers and normal lead times. *Journal of the Operational Research Society*, 32:1003–1009.

Sethi, S. and Cheng, F. (1997). Optimality of (s,S) policies in inventory models with markovian demand. *Operations Research*, 45(6):931–939.

Sherbrooke, C. (1975). Waiting time in an (S-1,S) inventory system - constant service time case. *Operations Research*, 23(4):819–820.

Silver, E. (1981). Operations research in inventory management: A review and critique. *Operations Research*, 29(4):628–645.

Silver, E. (2008). Inventory management: An overview, canadian publications, practical applications and suggestions for future research. *INFOR*, 46(1):15–28.

Silver, E., Pyke, D., and Peterson, R. (1998). *Inventory Management and Production Planing and Scheduling*. Wiley, New York, 3rd edition.

Silver, E. A., Naseraldin, H., and Bischak, D. P. (2009). Determining the reorder point and order-up-to-level in a periodic review system so as to achieve a desired fill rate and a desired average time between replenishments. *Journal of the Operational Research Society*, 60(9):1244–1253.

Smith, A. (1977). Optimal inventories for an (S-1,S) system with no backorders. *Management Science*, 23(5):522–528.

Song, J.-S., Zhang, H., Hou, Y., and Wang, M. (2010). The effect of lead time and demand uncertainties in (r, q) inventory systems. *Operations Research*, 58(1):68–80.

Song, J.-S. and Zipkin, P. H. (1996). The joint effect of leadtime variance and lot size in a parallel processing environment. *Management Science*, 42(9):1352–1363.

Sphicas, G. (1982). On the solution of an inventory model with variable lead times. *Operations Research*, 30(2):404–410.

Sphicas, G. and Nasri, F. (1984). An inventory model with finite-range stochastic lead times. *Naval Research Logistics Quarterly*, 31(4):609–616.

Strichartz, R. S. (2003). *A Guide to Distribution Theory and Fourier Transforms*. 2 edition.

Strijbosch, L. W. G. and Moors, J. J. A. (2005). The impact of unknown demand parameters on (R,S)-inventory control performance. *European Journal of Operational Research*, 162(3): 805–815.

Strijbosch, L. W. G. and Moors, J. J. A. (2006). Modified normal demand distributions in (R,S)-inventory control. *European Journal of Operational Research*, 172(1):201–212.

Suchanek, B. (1996). *Sicherheitsbestände zur Einhaltung von Servicegraden*. Europäischer Verlag der Wissenschaften, Frankfurt am Main.

Svoronos, A. and Zipkin, P. H. (1988). Estimating the performance of multi-level inventory systems. *Operations Research*, 36(1):57–72.

Tempelmeier, H. (1985). Inventory control using a service constraint on the expected customer order waiting time. *European Journal of Operational Research*, 19(3):313–323.

Tempelmeier, H. (2000). Inventory service-levels in the customer supply chain. *OR Spektrum*, 22(3):361–380.

Tempelmeier, H. (2006). *Inventory Management in Supply Networks*. Books on Demand GmbH, Norderstedt.

Tempelmeier, H. and Fischer, L. (2010). Approximation of the probability distribution of the customer waiting time under an (r, s, q) inventory policy in discrete time. *International Journal of Production Research*, 48(21):6275–6291.

Teunter, R. H. (2009). Note on the fill rate of single-stage general periodic review inventory systems. *Operations Research Letters*, 37(1):67–68.

Tijms, H. and Groenevelt, H. (1984). Simple approximations for the reorder point in periodic and continuous review (s,S) inventory systems with service level constraints. *European Journal of Operational Research*, 17(2):175–190.

Tijms, H. C. (1994). *Stochastic Models*. Wiley, Chichester.

Titterington, D., Smith, A., and Makov, U. (1985). *Statistical Analysis of Finite Mixture Distributions*. Wiley, New York.

van Beek, P. (1981). Modelling and analysis of a multi-echelon inventory system. *European Journal of Operational Research*, 6(4):380–385.

van der Heijden, M. C. and de Kok, A. G. (1992). Customer waiting times in an (R,S) inventory system with compound poisson demand. *ZOR - Methods and Models of Operation Research*, 36(4):315–332.

van der Heijden, M. C. and de Kok, A. G. (1998). Estimating stock levels in periodic review inventory systems. *Operations Research Letters*, 22:179–182.

van der Veen, B. (1981). Safety stocks - an example of theory and practice in O.R. *European Journal of Operational Research*, 6(4):367–371.

van der Veen, B. (1984). Safety stocks and the paradox of the empty period. *European Journal of Operational Research*, 16(1):19–33.

van der Veen, B. (1986). Safety stocks and the order quantity that leads to the minimal stock. *European Journal of Operational Research*, 27(1):34–49.

Vatter, V. (2008). Enumeration schemes for restricted permutations. *Combinatorics, Probability and Computing*, 17(1):137–159.

Veinott, A. (1965). The optimal inventory policy for batch ordering. *Operations Research*, 13(3):424–432.

Veinott, A. (1966). On the optimality of (s,S) inventory policies: new conditions and a new proof. *SIAM Journal on Applied Mathematics*, 14(5):1067–1083.

Veinott, A. and Wagner, H. (1965). Computing optimal (s,S) inventory policies. *Management Science*, 11(5):525–552.

Wagner, H., O'Hagen, M., and Lundh, B. (1965). An empirical study of exactly and approximately optimal inventory policies. *Management Science*, 11(7):690–723.

Washburn, A. (1973). A bi-modal inventory study with random lead times. Technical Report AD769404, Naval Postgraduate School, Monterey, California.

Weiss, H. (1980). Optimal ordering policies for continuous review perishable inventory models. *Operations Research*, 28(2):365–374.

Weiss, H. (1988). Sensitivity of continuous review stochastic (s, S) inventory systems to ordering delays. *European Journal of Operational Research*, 36(2):174–179.

West, J. (1995). Generating trees and the Catalan and Schröder numbers. *Discrete Mathematics*, 146(1–3):247–262.

West, J. (1996). Generating trees and forbidden subsequences. *Discrete Mathematics*, 157(1–3):363–374.

Wijngaard, I. and van Winkel, E. (1979). Average costs in a continuous review (s,S) inventory system with exponentially distributed lead time. *Operations Research*, 27(2):396–401.

Yano, C. (1985). New algorithms for (Q,r) systems with complete backordering using a fill-rate criterion. *Naval Research Logistics Quarterly*, 32(4):675–688.

Zalkind, D. (1976). Further results for order-level inventory systems with independent stochastic leadtimes. Technical Report 76-6, Department of Health Administration and Curriculum in Operations Research and Systems Analysis, University of North Carolina at Chapel Hill, Chapel Hill, NC.

Zalkind, D. (1978). Order-level inventory systems with independent stochastic lead times. *Management Science*, 24(13):1385–1392.

Zhang, J. and Zhang, J. (2007). Fill rate of single-stage general periodic review inventory systems. *Operations Research Letters*, 35(4):503–509.

Zhang, R., Hopp, W., and Supatgiat, C. (2001). Spreadsheet implementable inventory control for a distribution center. *Journal of Heuristics*, 7(2):185–203.

Zheng, Y. (1991). A simple proof for optimality of (s,S) policies in infinite-horizon inventory systems. *Journal of Applied Probability*, 28(4):802–810.

Zheng, Y. and Federgruen, A. (1991). Finding optimal (s,S) policies is about as simple as evaluating a single policy. *Operations Research*, 39(4):654–665.

Zheng, Y.-S. and Chen, F. (1992). Inventory policies with quantized ordering. *Naval Research Logistics*, 39(3):285–305.

Zinn, W., Mentzer, J., and Croxton, K. (2002). Customer-based measures of inventory availability. *Journal of Business Logistics*, 23(2):19–43.

Zipkin, P. H. (1986a). Inventory service-level measures: convexity and approximation. *Managment Science*, 32(8):975–981.

Zipkin, P. H. (1986b). Stochastic leadtimes in continuous-time inventory models. *Naval Research Logistics Quarterly*, 33(4):763–774.

Zipkin, P. H. (2000). *Foundations of Inventory Management*. McGraw-Hill, Boston.

Glossary of Symbols

Fundamental concepts as defined in Chaps. 2 and 4

α^τ	Ready rate for a basis interval of length τ (def. 7, p. 14)
β	Fill rate (def. 8, p. 15)
γ	Time-weighed fill rate (def. 9, p. 17)
AI	Arriving inventory (as defined on p. 6)
B	Number of backorders (as defined on p. 5)
c_1	Fixed costs per replenishment order (def. 2, p. 11)
c_2	Inventory holding costs (def. 3, p. 12)
c_{31}	Costs for being in stockout state (def. 4, p. 13)
c_{32}	Costs per missing unit (def. 5, p. 13)
c_{33}	Costs per missing unit and time (def. 6, p. 13)
CC_{ax}	Computational complexity of algorithm x
$CDF_X(.)$	Cumulative density of the distribution of random variable X
$CDF^{-1}(.)$	Inverse cumulative probability density function or quantile function (def. 13, p. 44)
D	Demand per underlying base period (as defined on p. 6)
h	Inventory holding cost rate without costs for capital binding (def. 3, p. 12)
i	Interest rate (def. 3, p. 12)
I	Physical inventory (as defined on p. 5)
II	Issued inventory (as defined on p. 6)
IO	Inventory on order (as defined on p. 6)
IP	Inventory position (as defined on p. 6)
IS	Inventory system
K	Number of outstanding orders (as defined on p. 6)
L	(Replenishment) lead time (as defined on p. 6)
LTD	Lead time demand (as shortly discussed in Sect. 5.2.2)
NI	Net inventory (as defined on p. 6)

p	Acquisition price per unit (def. 1, p. 11)
$PDF_X(.)$	Probability density function of the distribution of random variable X
q	Deterministic order quantity (policy parameter, see Sect. 2.2.3)
Q	Stochastic order quantity (as defined in Sect. 2.2.3)
r	Order cycle (policy parameter, see Sect. 2.2.3)
s	Reorder point (policy parameter, see Sect. 2.2.3)
S	Order-up-to level (policy parameter, see Sect. 2.2.3)
SF	Inventory shortfall or inventory on order (as defined on p. 6)
t	Review interval (policy parameter, see Sect. 2.2.3)
SS	Safety stock (as described on p. 6)
W^O	Customer waiting time per order (def. 10, p. 18)
W^V	Customer waiting time per part (def. 11, p. 18)
x_{min}	Minimum value of the support of random variable X
x_{max}	Maximum value of the support of random variable X
X_{span}	Value span of the distribution of X, $X_{span} := x_{max} - x_{min}$

Additional concepts as introduced in Chap. 6

$\alpha^{\tau=1}$	Ready rate per period
$\alpha^{\tau=r_D}$	Ready rate per customer order arrival
$\alpha^{\tau=r}$	Ready rate per order cycle
$B_{[a,b]}$	Backorder amount that originates from time interval $[a+1, b]$
B_w^{**}	Additional coverable backorder amount (see Sect. 6.3.2.4, p. 96 et seq.)
$B^{*,\tau=1}$	Additional backorder amount per period
$B^{*,\tau=r}$	Additional backorder amount per order cycle
D^x	Demand in x periods
$Disc(\{(x, p_x), \ldots\})$	Discrete distribution defined by pairs of value and probability $(x, p_x), \ldots$
ELT	Effective lead time (def. 19, p. 65)
$I^{(-)}$	Change of inventory
$nf_1(a, b)$	*First adaption function* (def. 20, p. 84)
$nf_2(a, b)$	*Second adaption function* (def. 21, p. 86)
$nf_3(a, b)$	*Third adaption function* (def. 24, p. 102)
$NF_2(a, b)$	Truth function on $nf_2(a, b)$ (def. 22, p. 88)
SO_t	Truth function on the exhaustion of S (def. 23, p. 101)
sv^*	Aspired (target) service level
T^*	Set of subperiods of an order cycle, $T^* = \{0, 1, \ldots, r-1\}$

T^{**}	Consecutively numbered demand periods of an order cycle, $T^{**} = \{0, 1, \ldots, \frac{r}{r_D} - 1\}$
$Norm(\mu, \sigma)$	Normal distribution with mean μ and standard deviation σ
r_D	Customer demand cycle, i.e. the frequency of customer demand occurrences
$Unif(\{x, y, \ldots\})$	Discrete uniform distribution with possible values $\{x, y, \ldots\}$