Mathematical Methods of Operations Research (2019) 90:61–76
https://doi.org/10.1007/s00186-018-0654-0

ORIGINAL ARTICLE

Simple modeling techniques for base-stock inventory systems with state dependent demand rates

Fredrik Olsson[1]

Received: 19 March 2018 / Accepted: 8 November 2018 / Published online: 14 November 2018

Abstract
In this paper new modeling techniques of $(S-1, S)$ inventory systems (continuous review base-stock inventory systems) with state dependent demand rates are proposed. Examples of single-location $(S-1, S)$ inventory systems where the demand, experienced by the system, varies due to the state of the system are, e.g., inventory models with partial backorders, inventory models with lost sales, inventory models with perishable items, inventory models with emergency replenishments etc. Models of such inventory systems are in general hard to solve due to the fact that the Markov property is often lost, and the prevalent tool used in the literature for providing exact solutions of such models is the theory of partial differential equations. Instead of using partial differential equations with rather complicated analysis of boundary conditions, we suggest considerably simpler techniques which are based on elementary theory of queueing and renewal processes. First, we show that it is possible to use Markov theory in order to prove certain statistical properties of the limiting distribution of the ages of the items in the system. Secondly, we develop a corresponding procedure based on renewal theory, which forms a basis for more complicated models assuming non-Poisson customer demand processes.

Keywords Inventory control · Cox process · State dependent demand · Base-stock policy · Queueing

1 Introduction

In many supply chain systems it is common that the customer demand streams, experienced by the system, vary as a function of the state of the system. For example, in a standard single-level lost sales inventory system, the customer demand rate (from a

✉ Fredrik Olsson
fredrik.olsson@iml.lth.se

[1] Department of Industrial Management and Logistics, Lund University, Ole Römers väg 1, P.O. Box 118, 221 00 Lund, Sweden

modeling point of view) is positive when there is stock on hand, and zero otherwise. Another interesting and related example, which is under the assumption of zero lead-times, mathematically equivalent to lost sales is supply chain systems with the option of emergency orders via alternative supply sources. In such situations, a customer demand may be re-directed to another stock-location if there is no stock on hand at the original location or if, e.g., the residual lead-time of the closest (in time) outstanding order, at the original location, is too long. Yet another important family of systems with similar features is inventory systems with perishable items. In such systems it is common that customer demands vary as a function of the age of the items in the system. Of course, other types of stock-dependent demand models also fall into the class of inventory problems studied in this paper.

The contribution of this paper is primarily to develop novel strategies for modeling $(S-1, S)$ inventory systems (continuous review base-stock inventory systems) with state dependent demand rates which are, from a mathematical point of view, significantly simpler than existing techniques. Instead of using the heavy mathematical machinery of partial differential equations (PDEs) we develop two simple modeling techniques which are based on some important characteristics of the model. The main advantages of our techniques compared to existing ones are two-fold. First, in Section 2.2, we use a classical queueing approach to derive the stationary distribution of the ages of the items in the system. In particular, we show the connection between the stationary distribution of the ages of the items in the system and the corresponding stationary distribution of the inventory level. Secondly, in Sect. 2.3, we use a renewal theory related technique which makes it possible to use our technique in order to extend the analysis to more general renewal demand processes.

In order to avoid confusion, we distinguish between the actual demand rate and the system demand rate which is the demand rate constituting of, by the system, accepted demands. For example, consider a base-stock inventory system with lost sales and Poisson demand with arrival rate λ. This means that the actual demand rate is always λ, disregard of the state of the system. However, the system demand rate is λ when there is stock on hand, and zero otherwise. This means that the system demand rate is itself a stochastic process. From a modeling point of view, the process of interest is the system demand process, and in this paper we consider a Poisson system demand process with state dependent arrival rates, i.e., a Cox process, see e.g. Cox (1955) and Kingman (1964).

As mentioned, there exist numerous inventory/production models that fall into the class of continuous review base-stock inventory systems with state dependent demand rates. A main feature concerning this stream of literature is that these models tend to be rather complex to analyze since the stochastic process describing the number of items in stock is, in general, not a Markov process. We proceed by briefly discussing some related literature concerning similar types of systems from a modeling point of view.

The literature concerning continuous review perishable inventory systems with positive lead-times is very scarce. Schmidt and Nahmias (1985) consider a continuous review base-stock inventory system with perishable items and lost sales. In order to obtain an exact solution Schmidt and Nahmias (1985) derive a system of PDEs and corresponding boundary conditions (a similar approach can be found in Gnedenko and Kovalenko 1968). Clearly, the paper by Schmidt and Nahmias (1985) has been a source of inspiration for many other papers dealing with similar problems. In fact, to the best of our knowledge, most subsequent papers dealing with continuous review base-stock inventory systems with state dependent demand rates use variants of the method developed in Schmidt and Nahmias (1985) as a building block. Olsson and Tydesjö (2010) consider a similar model as Schmidt and Nahmias (1985), but assume that all unsatisfied demands are backordered. Moreover, Olsson and Tydesjö (2010) state that their solution is an approximation since the constant lead-time is replaced by a lead-time which is exponentially distributed with the same mean. We will in this paper show that their solution is indeed exact. In a recent paper, Olsson and Turova (2016) generalize the model studied in Schmidt and Nahmias (1985) by considering a more general demand structure by using the method of PDEs. For more detailed literature reviews concerning perishable items in inventory systems see, e.g., Karaesmen et al. (2011) or Nahmias (2011).

Another stream of research connected to our model concerns inventory systems with emergency replenishments. Moinzadeh and Schmidt (1991) consider a continuous review base-stock inventory problem with two supply modes; one for regular replenishments and the other for emergency replenishments. Customer demands are assumed to follow a Poisson process. Based on information about the net inventory and the timing of all outstanding orders they set up a system of PDEs with corresponding boundary conditions. Hence, the method developed in Moinzadeh and Schmidt is very similar to the method used in Schmidt and Nahmias (1985). Moinzadeh and Aggarwal (1997) extend the model in Moinzadeh and Schmidt (1991) to a multi-echelon setting, by again applying a similar modeling technique as in Schmidt and Nahmias (1985) and Moinzadeh and Schmidt (1991). Song and Zipkin (2009) consider a similar dual-supply model as Moinzadeh and Schmidt (1991), where they keep track on two inventory positions; IP_1 and IP_2. While IP_1 is the standard inventory position (i.e., includes all outstanding orders), IP_2 only includes those outstanding orders that will arrive within a stipulated time window t'. However, the downside of this modeling technique is that the timing of outstanding orders are only partially known. That is, the only information provided is whether the residual lead-time of the closest outstanding order is less than t' or not. In this paper, we provide a solution technique that allows for complete information about the timing of outstanding orders, at all time instances. In a recent paper, Howard et al. (2015) use the modeling technique from Song and Zipkin (2009) in a heuristic two-echelon inventory model. For a more complete overview of inventory problems with emergency replenishments see, e.g., Minner (2003), and Yao and Minner (2017).

Inventory models with lateral transshipments are conceptually very similar to corresponding models with emergency replenishments. The main difference is that a lateral transshipment is requested from a neighbor inventory location, instead of an outside supplier (or a central warehouse). Most models in the literature dealing with lateral transshipments do not consider the timing of outstanding orders. Instead, it is common to assume that a lateral transshipment is requested if there is no stock on hand upon a customer arrival. The lateral transshipment is then realized (with zero lead-time) if a nearby inventory location has stock on hand. Moreover, for model tractability, most literature assume that replenishment lead-times are exponentially distributed. Some examples that fall into this category of papers are Axsäter (1990), Alfredsson and Verrijdt (1999) and Kutanoglu and Mahajan (2009). More recently, Olsson (2015) developed an approximate solution procedure for a model with lateral transshipments and non-zero transshipment lead-times where also the timing of outstanding orders are taken into consideration. The main advantage of our approach compared to the papers above concerning lateral transshipments is that we have complete information about the timing of the items in the system (i.e., the age of all items). This means that it is possible to design transshipment policies which are based on more detailed information than the simple rule described above. Moreover, as mentioned before, by using our approach it is possible to consider more general non-Poisson demand processes.

As already mentioned, another line of research connected to our model concerns inventory models with partial backorders, where some arriving customers are backordered and others are lost depending on the state of the system upon arrival. Moinzadeh (1989) considers a continuous review base-stock inventory system where customers are backordered with a predetermined probability in case there is no stock on hand upon arrival. Hence, the policy in Moinzadeh (1989) is stock level-based, while our policy is age-based. For the case with a fixed ordering cost, Zhang et al. (2003) consider an (R, Q) inventory system with a partial backordering strategy which takes the residual lead-time into account. However, only one outstanding order is allowed. Moreover, the policy is based on the assumption that all arriving customers are lost if the residual lead-time is deemed too long, otherwise backordered. In our model, however, this does not have to be the case. For example, we may assume that, say 10% (i.e., not necessarily 0%), of the customers are willing to wait for an outstanding order, although the residual lead-time is above some threshold value. Olsson (2014) considers an inventory model similar to Moinzadeh (1989), but also includes the case of perishable items. Moreover, Olsson (2014) considers a heuristic approach for the evaluation and optimization of a continuous review (R, Q) inventory model with perishable items.

In the next section, we provide two different techniques for age-based control of an $(S - 1, S)$ inventory system where customers arrive according to a Cox process. In Sect. 3, we revisit the model of Olsson and Tydesjö (2010) and provide a clear relationship between the finite and infinite lifetime model, respectively. Finally, in Sect. 4, some concluding remarks are given.

2 Modeling techniques

2.1 Preliminaries

Consider a single-location inventory model with continuous review, a constant lead-time L and state dependent demands (this demand process will be defined in detail below). The structure of the optimal policy for the considered model would certainly be very complex since it would depend on information concerning the age of all items in the system when deciding whether or not to order. As an alternative to a complex optimal policy, we here consider the commonly used $(S - 1, S)$ policy. Given an

$(S-1, S)$ policy, a new item is ordered whenever an item leaves the inventory. Let $T_1, T_2, \ldots, T_S$ represent the ages of the items in the system in stationarity, where T_1 represents the age of the oldest item in the system, T_2 the age of the second oldest item, etc. Hence, we have the order statistic $0 \leq T_S < T_{S-1} < \cdots < T_1 < \infty$. The age of an item is assumed to begin when the item leaves the supplier. Note that there are always exactly S items in the system (i.e., either in stock or on order) if customers are lost when there is zero stock on hand upon arrival. If customers are assumed to be backordered, $T_1, T_2, \ldots, T_S$ represent the ages of the S youngest items in the system (the other older items are already assigned to waiting customers). All customers are served according to FCFS (first come-first served).

We proceed by deriving the limiting joint density function of the ages of the items in the system. Especially, we will derive the limiting marginal density function of the age of the oldest item in the system (not yet assigned to any customer demand). This marginal density is of particular interest when calculating measures such as average stock on hand, average number of backorders, etc.

In order to develop an intuitive solution methodology, let us start with the simple case of homogeneous Poisson demand. In this case, we formulate the following theorem:

Theorem 1 *The limiting joint density of $T_1, T_2, \ldots, T_S$ assuming a standard single-echelon inventory system with Poisson demand with constant demand intensity $\lambda > 0$, infinite lifetime and full backordering, is*

$$f_{T_1,T_2,\ldots,T_S}(t_1, \ldots, t_S) = \lambda^S e^{-\lambda t_1}, \tag{1}$$

for $0 \leq t_S < t_{S-1} < \ldots < t_1$.

Proof Consider $2 \leq k \leq S$, and let $\tau_{k-1} = T_{k-1} - T_k$ and $\tau_S = T_S$. Then, note that $\tau_k \in \text{Exp}(1/\lambda)$ since τ_k is the time between Poisson demands. Indeed, by the well known PASTA property (see, e.g., Tijms 2003) this holds also for τ_S. Hence, we have

$$f_{\tau_1,\tau_2,\ldots,\tau_S}(v_1, v_2, \ldots, v_S) = \prod_{n=1}^{S} \lambda e^{-\lambda v_n} = \lambda^S e^{-\lambda \sum_{n=1}^{S} v_n}, \quad v_n > 0. \tag{2}$$

Note that v_k in (2), $1 \leq k \leq S$, is the argument of τ_k. This means that $v_1 + v_2 + \ldots + v_S$ in (2) becomes T_1. Hence, by using the so called Transformation Theorem (see, e.g., Theorem I.2.1 in Gut 1995) we immediately get the desired result. □

It is also interesting to note the following statement which follows directly from Theorem 1. Later, in Sect. 2.3 we will comment more about this result.

Corollary 1 *Consider the problem in stationarity, and define $T_k^{(S)}$ as the age of the k:th oldest item with a base stock level S. Then, $T_1^{(S)}$ and $T_2^{(S+1)}$ are equal in distribution, i.e.,*

$$T_1^{(S)} \stackrel{d}{=} T_2^{(S+1)}, \quad S \geq 1.$$

Proof The statement follows directly by deriving the marginal density for T_k (for a given base-stock level S). Given the order statistic, $t_S < \cdots < t_{k+1} < t_k < t_{k-1} < \cdots < t_1$, we obtain

$$\begin{aligned} f_{T_k}(t_k) &= \int_{t_1=t_2}^{\infty} \int_{t_2=t_3}^{\infty} \cdots \int_{t_{k-1}=t_k}^{\infty} \int_{t_{k+1}=0}^{t_k} \cdots \int_{t_{S-1}=0}^{t_{S-2}} \\ &\quad \int_{t_S=0}^{t_{S-1}} \lambda^S e^{-\lambda t_1} dt_S \ldots dt_{k+1} dt_{k-1} \ldots dt_1 \\ &= \int_{t_1=t_2}^{\infty} \int_{t_2=t_3}^{\infty} \cdots \int_{t_{k-1}=t_k}^{\infty} \lambda^S e^{-\lambda t_1} \frac{t_k^{S-k}}{(S-k)!} dt_{k-1} \ldots dt_1 \\ &= \lambda^{S-k+1} e^{-\lambda t_k} \frac{t_k^{S-k}}{(S-k)!}. \end{aligned} \tag{3}$$

Hence, $T_1^{(S)} \stackrel{d}{=} T_2^{(S+1)} \in \text{Erlang}(\lambda, S)$. □

Remark 1 As mentioned, the density f_{T_1} in (3) is of particular interest when deriving a total cost function for the considered model. For example, in a model with linear inventory holding and backorder costs, the total cost function can be readily obtained as

$$TC(S) = \lambda \left(h \int_L^{\infty} (t_1 - L) f_{T_1}(t_1) dt_1 + b \int_0^L (L - t_1) f_{T_1}(t_1) dt_1 \right),$$

where h is the holding cost per unit and time unit, and b the backorder cost per unit and time unit. In a similar fashion, corresponding cost functions can be obtained for models concerning emergency transshipments, lateral transshipments, perishable items, etc.

2.2 A simple modeling technique based on basic queueing theory

Let us now assume a more general case, which arises in many practical applications where the demand rate varies as a function of the age of the oldest item in the system (which is very natural since the issuing rule is FCFS). In more detail, assume that the demand follows a Cox process with state dependent demand rate $\eta(T_1)$ according to

$$\eta(T_1) = \begin{cases} \lambda, & \text{for } 0 \le T_1 < \ell \\ \mu, & \text{for } T_1 \ge \ell, \end{cases} \tag{4}$$

where λ, μ and ℓ are non-negative constants, arbitrarily fixed. That is, here the demand rate is itself a stochastic process. As an example, the stochastic demand rate process in (4) can be used to model lost sales cases by setting $\lambda = 0$ and ℓ equal to the lead-time. Another interesting application of the demand process defined in (4) is inventory models with perishable items. For example, in practice, it is very common that retailers lower the price for a commodity when it is relatively near its due date. Then, a typical scenario is that, when the price is lowered the customer demand rate jumps from λ to a larger value $\mu > \lambda$. Other related examples of the application of (4) may include production maintenance models, inventory models with partial backordering, and inventory models with emergency supply.

Now, if we define $X(T_1)$ as the number of demand occurrences in the interval $(0, T_1]$ we have

$$X(T_1) \in \begin{cases} \text{Po}\,(\lambda T_1) & \text{for } 0 \leq T_1 < \ell \\ \text{Po}\,(\lambda \ell + (T_1 - \ell)\mu) & \text{for } \ell \leq T_1 < \infty, \end{cases} \tag{5}$$

where $\text{Po}(\cdot)$ denotes the Poisson distribution. Then, in view of (1) and (5) an intuitive conjecture is that the limiting joint density of the ages of the items in the system, $T_1, T_2, \ldots, T_S$, has the form presented in Theorem 2 below. We now prove that this is, indeed, the case by using basic queueing theory (a similar version of this theorem is provided in Olsson and Turova (2016), but in that paper the heavy machinery of PDEs is used to prove the result). For a general overview on Markovian queueing systems see, e.g., Kleinrock (1975).

Theorem 2 *Assume the same inventory system as in Theorem 1 but with a state-dependent demand rate according to (4). Then, for $0 \leq t_S < t_{S-1} < \ldots < t_1$ we have*

$$f_{T_1,\ldots,T_S}(t_1, \ldots, t_S) = \begin{cases} Ce^{-\lambda t_1}, & 0 \leq t_1 < \ell \\ Ce^{-(\lambda \ell + (t_1 - \ell)\mu)}, & \ell \leq t_1 < \infty. \end{cases} \tag{6}$$

where C is a positive constant.

Proof First, let us for notational (and readability) purposes and without loss of generality assume that ℓ is the same as the constant lead-time, L. That is, since the joint density in (6) does not depend on L, the lead-time L is just an arbitrary point in time. Then, note that the inventory model described is equivalent to an M/D/∞ queue with state dependent arrival rates, where the number of outstanding orders in the inventory model corresponds to the number of busy servers in the M/D/∞ queue. In more detail, define $IL = n$ as the inventory level so that $n \in \{-\infty, \ldots, S\}$ and then $S - n$ is the number of busy servers in the M/D/∞ queue.

Brumelle (1978) proves, in an elegant manner, that the stationary distribution of the number of busy servers in an M/G/∞ system with state dependent arrival rates is completely determined by the mean service time, and thereby insensitive to the shape of the service time. This means that we can replace the constant lead-time by an exponentially distributed lead-time with the same mean and still obtain the same results. Hence, let us for the moment assume exponential lead-times with a *mean* lead-time equal to ℓ. Now, in our original inventory system (with constant $\ell = L$) we know that the inventory level is positive if and only if $T_1 \geq \ell$, and non-positive otherwise.

This means that (4) corresponds to

$$\eta_n = \begin{cases} \lambda, & \text{for } n \leq 0 \\ \mu, & \text{for } n > 0, \end{cases} \tag{7}$$

in the M/M/∞ model, where η_n is the demand rate given that the inventory level is equal to n.

Remark 2 While the distribution of the number of busy servers is insensitive to the service time distribution of an M/G/∞ queue, this will not be true for all statistics, and in particular it is not true for the longest time in service (which corresponds to T_1). However, recall that we only track the age of the oldest item in the model with a deterministic lead-time. Clearly, it makes no sense to track the ages of the units in the system where lead-times are exponentially distributed, since orders may be received in a different order than they were placed.

By solving a birth-death M/M/∞ queueing system with a state dependent arrival rate according to (7) and a mean lead-time equal to ℓ, we obtain

$$P_n^{(S)} = \begin{cases} \frac{S!}{(S-n)!}(\frac{1}{\mu\ell})^n P_0^{(S)}, & \text{for } n = 1, \ldots, S \\ \frac{S!}{(S-n)!}(\frac{1}{\lambda\ell})^n P_0^{(S)}, & \text{for } n = 0, -1, -2, \ldots, \end{cases} \tag{8}$$

where $P_n^{(S)}$ is the steady state probability that the inventory level is equal to n given the base stock level S. The probability of zero stock on hand, $P_0^{(S)}$, serves as a normalizing constant and is obtained as

$$\frac{1}{P_0^{(S)}} = \sum_{n=1}^{S} \frac{S!}{(S-n)!}\left(\frac{1}{\mu\ell}\right)^n + \sum_{n=-\infty}^{0} \frac{S!}{(S-n)!}\left(\frac{1}{\lambda\ell}\right)^n. \tag{9}$$

The rest of the proof is to show that the joint density in (6) (and the corresponding marginal density of T_1) generates exactly the same inventory level probabilities as the M/M/∞ queue with state dependent demand rates.

Given the joint density in (6), the marginal density of T_1 is obtained as

$$\begin{aligned} f_{T_1}(t_1) &= \int_0^{t_1}\int_0^{t_2}\cdots\int_0^{t_{S-1}} C\exp\{-\int_0^{t_1}\eta(t)dt\}\, dt_S \ldots dt_2 \\ &= C\exp\{-\int_0^{t_1}\eta(t)dt\}\cdot t_1^{S-1}/(S-1)! \\ &= \begin{cases} Ce^{-\lambda t_1} t_1^{S-1}/(S-1)!, & 0 \le t_1 < \ell \\ Ce^{-(\lambda L+(t_1-L)\mu)} t_1^{S-1}/(S-1)!, & \ell \le t_1 < \infty, \end{cases} \end{aligned} \tag{10}$$

where C is the normalizing constant

$$\begin{aligned} \frac{1}{C} &= \int_0^{\ell} e^{-\lambda t_1}\frac{t_1^{S-1}}{(S-1)!}dt_1 + \int_{\ell}^{\infty} e^{-(\lambda\ell+\mu(t_1-\ell))}\frac{t_1^{S-1}}{(S-1)!}dt_1 \\ &= \frac{1}{\lambda^S}\left(1-\sum_{n=0}^{S-1}\frac{(\lambda\ell)^n}{n!}e^{-\lambda\ell}\right) + \frac{e^{(\mu-\lambda)\ell}}{\mu^S}\sum_{n=0}^{S-1}\frac{(\mu\ell)^n}{n!}e^{-\mu\ell}. \end{aligned} \tag{11}$$

Here we have used that the integrals in (11) can be re-written as Gamma distribution functions. Moreover, from (9) and (11) it is easy to show, by simple algebra, the following relation between $P_0^{(S)}$ and C,

$$P_0^{(S)} = \frac{C\ell^S e^{-\lambda\ell}}{S!}. \tag{12}$$

We proceed by deriving $P_n^{(S)}$, for $n = 1, \ldots, S$, by using the marginal density of T_1 in (10). Note that n units in stock is equivalent to $0 \leq T_S \leq T_{S-1} \leq \ldots \leq T_{n+1} < \ell \leq T_n \leq \ldots \leq T_1$. This gives

$$\begin{aligned} P_n^{(S)} &= \int_\ell^\infty \int_\ell^{t_1} \int_\ell^{t_2} \cdots \int_\ell^{t_{n-1}} \int_0^\ell \int_0^{t_{n+1}} \cdots \int_0^{t_{S-1}} Ce^{-(\lambda\ell+(t_1-\ell)\mu)} dt_S dt_{S-1} \ldots dt_2 dt_1 \\ &= C \frac{\ell^{S-n}}{(S-n)!(n-1)!} \int_\ell^\infty Ce^{-(\lambda\ell+(t_1-\ell)\mu)} (t_1-\ell)^{n-1} dt_1 \\ &= C \frac{\ell^{S-n} e^{-\lambda\ell}}{\mu^n (S-n)!}. \end{aligned} \tag{13}$$

By using the relation (12) in (13) we immediately get

$$P_n^{(S)} = C \frac{\ell^{S-n} e^{-\lambda\ell}}{\mu^n (S-n)!} = \frac{S!}{(S-n)!} \left(\frac{1}{\mu\ell}\right)^n P_0^{(S)},$$

which is exactly the expression in (8) for $n = 1, \ldots, S$.

Let us now derive $P_{-n}^{(S)}$, for $n = 0, 1, 2, \ldots$, by using the marginal density of T_1 in (10). Note that, backorders occur when there is no stock on hand when a customer arrives, i.e., when $T_1 < \ell$. Consider the system when all S items are outstanding and when there are no backorders. Denote X as the stochastic residual lead-time for the oldest item among the S items outstanding in the system. This implies that $X = \ell - T_1$, and $f_X(x) = f_{T_1}(\ell - x)$ for $0 < x \leq \ell$. The event of n ($n = 0, 1, 2, \ldots$) backorders in the system and empty stock is equivalent to the event of having n customer orders during the residual lead-time, X, where $0 < X \leq \ell$. Hence, by applying integration by parts we get

$$\begin{aligned} P_{-n}^{(S)} &= \int_0^\ell \frac{(\lambda x)^n}{n!} e^{-\lambda x} \cdot f_X(x) dx = \int_0^\ell \frac{(\lambda x)^n}{n!} e^{-\lambda x} \cdot \frac{C}{(S-1)!} e^{-\lambda(\ell - x)} (\ell - x)^{S-1} dx \\ &= C \frac{(\lambda\ell)^{S+n}}{(S+n)!} \cdot \frac{e^{-\lambda\ell}}{\lambda^S}. \end{aligned}$$

By again using (12), we obtain

$$P_{-n}^{(S)} = \frac{S!}{(S+n)!} (\lambda\ell)^n P_0^{(S)},$$

which after a change of variables gives the desired result in (8). The proof is complete. □

2.3 A simple modeling technique based on basic renewal theory

In this section we derive the limiting density function of the age of the oldest item in the system, T_1, by a completely different method than in the previous section. The method is interesting in its own right by deriving the marginal distribution of T_1 directly without introducing the machinery of PDEs. Notice that since (6) only depends on the age of the oldest unit in the system it is sufficient to derive the limiting marginal density of T_1 for the purpose of calculating the average inventory level, average number of backorders, etc.

Let us start proving (10) for the simple case $S = 1$. Denoting ξ as the customer interarrival times in stationarity, we have $P(\xi > x) = e^{-\int_0^x \eta(y)dy}$. The mean interarrival time, m, becomes

$$m = \int_0^\infty P(\xi > x)dx = \int_0^\ell e^{-\lambda x}dx + \int_\ell^\infty e^{-(\lambda\ell+(x-\ell)\mu)}dx.$$

Using the key renewal theorem (see, e.g., Ross 1992), the limiting distribution of the current age of the item is obtained as

$$\begin{aligned} P(T_1 \le t_1) &= \lim_{t\to\infty} P(T_1(t) \le t_1) = m^{-1}\int_0^{t_1} P(\xi > x)dx \\ &= \begin{cases} m^{-1}\int_0^{t_1} e^{-\lambda x}dx & \text{for } 0 \le t_1 < \ell \\ m^{-1}\int_0^{t_1} e^{-(\lambda\ell+(x-\ell)\mu)}dx & \text{for } \ell \le t_1 < \infty. \end{cases} \end{aligned}$$

Now, differentiation of $P(T_1 \le t_1)$ gives the desired result, i.e.,

$$f_{T_1}(t_1) = \begin{cases} m^{-1} \cdot e^{-\lambda t_1} & \text{for } 0 \le t_1 < \ell \\ m^{-1} \cdot e^{-(\lambda\ell+(t_1-\ell)\mu)}. & \text{for } \ell \le t_1 < \infty. \end{cases} \tag{14}$$

From (14) we note that the normalizing constant C in (10) may be interpreted as the limiting average rate of arrivals per unit time.

Let us now consider the more complicated case $S = 2$. The idea of the proof is quite straightforward and intuitive, and is based on deriving the stationary distribution of the time between customer arrivals. Once the stationary distribution of the time between customer arrivals is known, the limiting distribution of the age of the oldest item in the system is easily obtained. Let us therefore start by deriving the stationary distribution of the time between customer arrivals. Define ξ_k as the k:th interval between successive customer arrivals, see Fig. 1.

In Fig. 1, the ages of the two items can be seen over time. Clearly, just after a demand occurrence the second oldest item becomes the oldest item. As an example, in Fig. 1, consider the time just after the third demand occurrence. At this point in time the age of the oldest item in the system is ξ_3, while the age of the second oldest item is 0. Hence, the next interarrival-time duration, ξ_4, is dependent on if ξ_3 was greater or less than ℓ. In other words, if we know the density of ξ_k, the density of ξ_{k+1}, $f_{\xi_{k+1}}(\cdot)$, is obtained as

$$f_{\xi_{k+1}}(u) = \int_0^\infty g(u|v)f_{\xi_k}(v)dv, \tag{15}$$

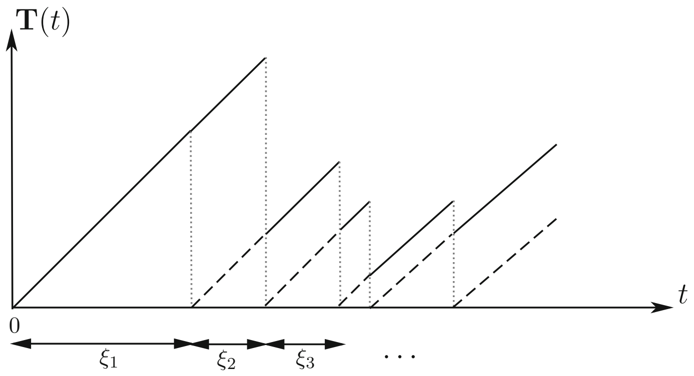

Fig. 1 A representation of $\mathbf{T}(t) = (T_1(t), T_2(t))$ for $S = 2$ with $T_1(0) = T_2(0) = 0$. The solid trajectories describe the age of the oldest item, while the dashed trajectories represent the age of the second oldest item (note that the two items have exactly the same age in the initial interval $[0, \xi_1)$)

where $g(u|v)$ is the kernel (or probability transition function). The kernel, $g(u|v)$, becomes

$$g(u|v) = \begin{cases} \lambda e^{-\lambda u} & \text{for } v + u < \ell \\ e^{-\lambda(\ell - v)} \mu e^{-(u+v-\ell)\mu} & \text{for } v < \ell < v + u \\ \mu e^{-\mu u} & \text{for } v > \ell \end{cases} \tag{16}$$

The intuition behind (16) is, given that the oldest item has the age $\xi_n = v < \ell$ just after the n:th customer occurrence, we know that $f_{\xi_{n+1}}(u)$ becomes $\lambda e^{-\lambda u}$ if $\xi_{n+1} = u$ is less than $\ell - v$, etc.

Let us now consider the stationary case, i.e., when ξ_k and ξ_{k+1} are equal in distribution and have the density $f_\xi(u)$. Then (15) becomes:

$$\begin{aligned} f_\xi(u) &= \int_0^\infty g(u|v) f_\xi(v) dv \\ &= \int_0^{\ell - u} \lambda e^{-\lambda u} f_\xi(v) dv + \int_{\ell - u}^{\ell} e^{-\lambda(\ell - v)} \mu e^{-(u+v-\ell)\mu} f_\xi(v) dv \\ &\quad + \int_\ell^\infty \mu e^{-\mu u} f_\xi(v) dv. \end{aligned} \tag{17}$$

We continue by dividing the functional equation (17) into two cases; $u < \ell$ and $u > \ell$, respectively. Starting with the case $u < \ell$, let us transform (17) into a more dense form. By multiplying both sides of (17) by $e^{\mu u}$ and then differentiating (17) with respect to u yields the following functional equation

$$\mu f_\xi(u) + f'_\xi(u) = (\mu - \lambda) e^{-\lambda u} \left(f_\xi(\ell - u) + \lambda \int_0^{\ell - u} f_\xi(v) dv \right). \tag{18}$$

Now, it is easy to verify that $f_\xi(u) = Ce^{-\lambda u}$, where C is a constant, is the general solution to (18). In complete analogy with the case $u < \ell$ we similarly obtain $f_\xi(u) =$

$De^{-\mu u}$ for $u > \ell$, where D is a constant to be determined. So, to conclude we have

$$f_\xi(u) = \begin{cases} Ce^{-\lambda u} & \text{for } 0 < u < \ell \\ De^{-\mu u} & \text{for } u > \ell. \end{cases} \tag{19}$$

Continuity of $f_\xi(u)$ implies that $f_\xi(\ell) = Ce^{-\lambda\ell} = De^{-\mu\ell}$, which in turn yields that $D = Ce^{(\mu-\lambda)\ell}$. This provides the unique stationary solution, which by the argument of tightness (see Olsson and Turova (2016) and Shiryaev (1996)) also implies convergence of the distribution of the oldest age in the system to the limiting distribution independent of the initial conditions. The limiting density f_ξ is positive for all $u > 0$ (i.e., on the entire considered space). Therefore one can apply the key renewal theorem (Ergodic theorem) to derive the limiting distribution of the age T_1 of the oldest item in the stationary regime

$$f_{T_1}(t_1) = \frac{\mathbf{P}\{\xi_k + \xi_{k+1} > t_1 \cap \xi_k < t_1\}}{\mathbf{E}\{\xi\}} = \frac{\int_{t_1}^{\infty}\int_0^{t_1} g(u-v|v) f_\xi(v)dvdu}{\int_0^\infty u f_\xi(u)du}. \tag{20}$$

By using (16) and (19), the double integral in (20) reduces to (after some easy calculus) the following expression for the case $t_1 < \ell$,

$$\int_{t_1}^{\infty}\int_0^{t_1} g(u-v|v) f_\xi(v)dvdu = Ct_1e^{-\lambda t_1}.$$

Similarly, for the case $t_1 > \ell$, we obtain,

$$\int_{t_1}^{\infty}\int_0^{t_1} g(u-v|v) f_\xi(v)dvdu = Ct_1e^{-(\lambda\ell+(t_1-\ell)\mu)}.$$

Here we have again used that $f_\xi(u)$ is continuous. The average time between successive customer arrivals, $\mathbf{E}\{\xi\}$, follows directly from (19) as

$$\mathbf{E}\{\xi\} = \int_0^\infty u f_\xi(u)du = \int_0^\ell Cue^{-\lambda u}du + \int_\ell^\infty Cue^{-(\lambda\ell+(u-\ell)\mu)}du, \tag{21}$$

by again using the continuity of $f_\xi(u)$.

Hence, we can finally conclude that

$$f_{T_1}(t_1) = \begin{cases} Ct_1e^{-\lambda t_1} & \text{for } t_1 < \ell \\ Ct_1e^{-(\lambda\ell+(t_1-\ell)\mu)} & \text{for } t_1 > \ell, \end{cases} \tag{22}$$

where C is the normalizing constant,

$$C^{-1} = \int_0^\ell t_1e^{-\lambda t_1}dt_1 + \int_\ell^\infty t_1e^{-(\lambda\ell+(t_1-\ell)\mu)}dt_1.$$

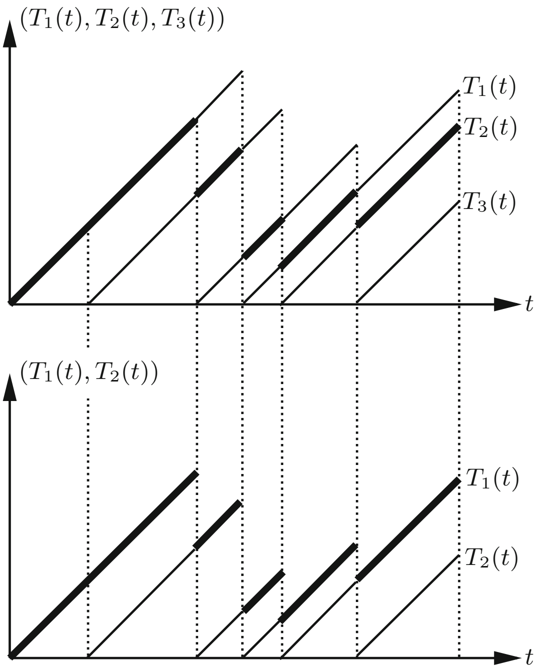

Fig. 2 A representation of the system for $(T_1(0), \dots, T_S(0)) = 0$. The trajectories describe the ages of the units between jumps for $S = 2$ (lower figure) and $S = 3$ (upper figure), respectively

Note that, in this case, $f_\xi(\cdot)$ in (19) is the same as $f_{T_2}(\cdot)$ for $S = 2$. Hence, from (14) and (19) we can conclude that $T_1^{(S=1)} \stackrel{d}{=} T_2^{(S=2)}$, which shows that the statement of Corollary 1 also holds for the Cox process in (4). Using this uniform property it is straightforward to prove the result for a general positive integer S by mathematical induction. □

In order to promote intuition, a particular representation of the inventory system is provided in Fig. 2. In Fig. 2, the trajectories for the ages of items in the system between jumps are represented for two cases. Here, a jump is defined as a customer arrival event. In the upper figure we consider a system with $S = 3$, and in the lower figure we have a corresponding system with $S = 2$. So, given that these two systems face the same sequence of customer demands, we note that $T_2(t)$ for $S = 3$ has exactly the same trajectory as $T_1(t)$ for $S = 2$ (see the bold lines for the two cases). The same equivalence holds between $T_3(t)$ for $S = 3$, and $T_2(t)$ for $S = 2$.

3 An illustrative example

In this section we revisit and reinterpret the model of Olsson and Tydesjö (2010). Consider a $(S-1, S)$ inventory system with perishable items where customer demands follow a Poisson process with rate λ. The lifetime of items, τ, and the lead-time, L, are deterministic. The lifetime of an item begins when the item leaves the supplier, which means, of course, that we assume $\tau > L$. Moreover, when the age of an item reaches its lifetime τ, the item is discarded and a new item is ordered from the supplier. We also assume that all unsatisfied demands are backordered. As noted, this is exactly the same inventory system as considered in Olsson and Tydesjö (2010).

Instead of introducing PDEs as in the related applied literature we will use the following fundamental result:

Lemma 1 *The perishable backorder inventory model considered is equivalent to a corresponding infinite lifetime inventory model with a Cox demand process with state dependent arrival intensity according to:*

$$\eta(T_1) = \begin{cases} \lambda, \ \textit{for}\ 0 \leq T_1 < \tau \\ \mu, \ \textit{for}\ T_1 \geq \tau, \end{cases} \tag{23}$$

where $\mu = \infty$.

Note that Lemma 1 follows directly by modeling perishable events as demand events. First, it is clear that the demand intensity is λ when $T_1 < \tau$, due to the backorder assumption. Secondly, when the lifetime of the oldest item reaches τ, the item should immediately be discarded from inventory and a new item should be ordered. This is equivalent to setting the demand rate to a very large value (∞) when the age of the oldest item, T_1, exceeds τ.

To avoid notational confusion, let $W_1, W_2, \ldots, W_S$ represent the ages of the items in the system when items have a finite fixed lifetime τ (i.e., $T_1, T_2, \ldots, T_S$ represent the ages of the items in the system for the infinite lifetime case described in Sect. 2. Again, we have the order $0 \leq W_S < W_{S-1} < \ldots < W_1 < \tau$. The following theorem provides an important connection between the finite and the infinite lifetime case, respectively:

Theorem 3 *Assume that the lifetime of an item,* τ*, is finite and let* $TEr(\cdot, \cdot)$ *denote the truncated Erlang distribution. Then*

$$W_1 = T_1 | T_1 < \tau \in TEr(\lambda, S) \ \textit{with support}\ [0, \tau).$$

Proof From (10) we have

$$f_{T_1}(t_1) = \begin{cases} De^{-\lambda t} {t_1}^{S-1} & \text{for } 0 \leq t_1 < \tau \\ De^{-(\lambda\tau + (t_1 - \tau)\mu)} {t_1}^{S-1} & \text{for } \tau \leq t_1 < \infty, \end{cases} \tag{24}$$

where $D = C/(S-1)!$ is the normalizing constant. Now, Lemma 1 directly implies that the age of the oldest item (not yet assigned for any waiting customer) in the finite lifetime case has the following density

$$f_{W_1}(t_1) = \begin{cases} De^{-\lambda t_1} {t_1}^{S-1} & \text{for } 0 \leq t_1 < \tau \\ 0 & \text{otherwise}, \end{cases} \tag{25}$$

where D is the normalizing constant,

$$\frac{1}{D} = \int_0^\tau {t_1}^{S-1} e^{-\lambda t_1} dt_1 = \frac{(S-1)!}{\lambda^S} \cdot \left(1 - \sum_{n=0}^{S-1} e^{-\lambda\tau} \frac{(\lambda\tau)^n}{n!} \right). \tag{26}$$

Hence, from (25) and (26) it is clear that $W_1 = T_1|T_1 < \tau$, where $T_1 \in \text{Erlang}(\lambda, S)$. The proof is complete. □

The intuitive interpretation of Theorem 3 is that the distribution of the age of the oldest item in the system in the case with a fixed lifetime, τ, is obtained as the conditional distribution that results from restricting the domain of the corresponding distribution for the infinite lifetime case. Hence, Theorem 3 provides a clear and simple relationship between the age distributions in the finite- and infinite lifetime case, respectively.

Moreover, recall that Olsson and Tydesjö (2010) replaced the constant lead-time by an exponentially distributed lead-time with the same mean for the purpose of deriving various performance measures such as expected inventory on hand, expected number of backorders etc. They stated this as an approximation, but clearly, given the proof of Theorem 2 in Sect. 2.2 it is clear that the result is exact.

4 Conclusions

This paper extends the literature concerning continuous review inventory systems with variable demand rates in two main directions. First, we provide new intuitive queueing related modeling techniques for age-based control of $(S-1, S)$ inventory systems with state dependent demand rates, which are considerably easier than in the existing ones. Instead of introducing heavy mathematical machinery as PDEs with rather complicated analysis of boundary conditions, which tends to cloud simple features of the model, we suggest alternative approaches which are based on probabilistic arguments concerning renewal processes and queueing theory. As discussed in the introduction, the approach in Sect. 2.2 gives a distinct relation between age-based vs. stock-based inventory control. In particular, while the stationary distribution of the inventory level is completely determined by the mean lead-time, the stationary age distribution of the items in the system is certainly not. Secondly, by applying our approach, it may also be possible to further generalize the model to incorporate more complicated demand interarrival patterns compared to the standard Poisson demand arrival process (by changing the interarrival distribution ξ and thereby also the kernel $g(u|v)$ in (16)). The limitations of our approaches are the assumptions of one-for-one ordering policies and unit-sized demands. However, the method considering PDEs suffers from the same restrictions.

As an illustration we revisit and reinterpret a base-stock inventory system with perishable items. Here, we give a clear relation between the stationary age distribution of the items in a finite lifetime and infinite lifetime setting, respectively.

References

Alfredsson P, Verrijdt J (1999) Modeling emergency supply flexibility in a two-echelon inventory system. Manag Sci 45(10):1416–1431

Axsäter S (1990) Modelling emergency lateral transshipments in inventory systems. Manag Sci 36(11):1329–1338

Brumelle SL (1978) A generalization of Erlang's loss system to state dependent arrival and service rates. Math Oper Res 3(1):10–16

Cox DR (1955) The analysis of non-Markovian stochastic processes by the inclusion of supplementary variables. Math Proc Camb Philos Soc 51(3):433–441

Gnedenko BV, Kovalenko IN (1968) Introduction to queueing theory. Israel Program for Scientific Translations, Jerusalem

Gut A (1995) An intermediate course in probability. Springer, New York

Howard C, Marklund J, Tan T, Reijnen I (2015) Inventory control in a spare parts distribution system with emergency stocks and pipeline information. Manuf Serv Oper Manag 17(2):142–156

Karaesmen I, Scheller-Wolf A, Deniz B (2011) Managing perishable and aging inventories: review and future research directions, planning production and inventories in the extended enterprise. In: Kempf K, Keskinocak P, Uzsoy P (eds) Int Ser Oper Res Manag Sci. Springer, New York

Kingman JFC (1964) On doubly stochastic poisson processes. Math Proc Camb Philos Soc 60(4):923–930

Kleinrock L (1975) Queueing systems theory, vol I. Wiley, New York

Kutanoglu E, Mahajan M (2009) An inventory sharing and allocation method for a multi-location service parts logistics network with time-based service levels. Eur J Oper Res 194(3):728–742

Minner S (2003) Multiple-supplier inventory models in supply chain management: a review. Int J Prod Econ 81–82:265–279

Moinzadeh K (1989) Operating characteristics of the $(S-1, S)$ inventory system with partial backorders and constant resupply times. Manag Sci 35(4):472–477

Moinzadeh K, Aggarwal PK (1997) An information based multiechelon inventory system with emergency orders. Oper Res 45(5):694–701

Moinzadeh K, Schmidt CP (1991) An $(S-1, S)$ inventory system with emergency orders. Oper Res 39(2):308–321

Nahmias S (2011) Perishable inventory systems. Springer, New York

Olsson F (2014) Analysis of inventory policies for perishable items with fixed lead-times and lifetimes. Ann Oper Res 217(1):399–423

Olsson F (2015) Emergency lateral transshipments in a two-echelon inventory system with positive transshipment lead-times. Eur J Oper Res 242(2):424–433

Olsson F, Turova TS (2016) Ergodicity of age-dependent inventory control systems. J Appl Probab 53(3):688–699

Olsson F, Tydesjö P (2010) Inventory problems with perishable items: fixed lifetimes and backlogging. Eur J Oper Res 202(1):131–137

Ross SM (1992) Applied probability models with optimization applications. Dover Publications, New York

Schmidt CP, Nahmias S (1985) $(S-1, S)$ policies for perishable inventory. Manag Sci 31(6):719–728

Shiryaev A (1996) Probability. Springer, New York

Song JS, Zipkin P (2009) Inventories with multiple supply sources and networks of queues with overflow bypasses. Manag Sci 55(3):362–372

Tijms H (2003) A first course in stochastic models. Wiley, New York

Yao M, Minner S (2017) Review of multi-supplier inventory models in supply chain management: an update, technical report in SSRN Electronic Journal. https://doi.org/10.2139/ssrn.2995134

Zhang GP, Patuwo BE, Chu C-W (2003) A hybrid inventory system with a time limit on backorders. IIE Trans 35(7):679–687