AN \((s, S)\) INVENTORY MODEL WITH LEVEL-DEPENDENT $G/M/1$-TYPE STRUCTURE

Sung-Seok Ko
Department of Industrial Engineering, Konkuk University
120 Neungdong-ro, Gwangjin-gu, Seoul, 143-701, Korea

Jangha Kang and E-Yeon Kwon
Department of Industrial Engineering, Chosun University
309 Pilmun-daero, Dong-gu, Gwangju, 501-759, Korea

(Communicated by Shoji Kasahara)

Abstract. Inventory models are widely used in a variety of real-world applications. In particular, inventory systems with perishable items have received a significant amount of attention. We consider an \((s, S)\) continuous inventory model with perishable items, impatient customers, and random lead times. Two characteristic behaviors of impatient customers are balking and reneging. Balking is when a customer departs the system if the item they desire is unavailable. Reneging occurs when a waiting customer leaves the system if their demand is not met within a set period of time. The proposed system is modeled as a two-dimensional Markov process with level-dependent $G/M/1$-type structure. We also consider independent and identically distributed replenishment lead times that follow a phase-type distribution. We find an efficient approximation method for the joint stationary distribution of the number of items in the system, and provide formulas for several performance measures. Moreover, to minimize system costs, we find the optimal values of \(s\) and \(S\) numerically and perform a sensitivity analysis on key parameters.

1. Introduction. Inventory models have been widely studied, and great theoretical advances have been made over the past several decades. Extensions and variants of traditional inventory systems, including customer and product characteristics and replenishment time uncertainty, continue to be studied for their application to various real-world problems (see [8] for a more complete list of applications).

Many products do not last forever; food, photographic film, and medications, for example, all have expiration dates. Even blood banks are limited to storing units of blood for a maximum of 21 days. The analysis of inventory systems with perishable items has received a significant amount of attention from many researchers [9, 10, 11, 12, 14, 16, 17, 18, 22, 23, 24, 25] for the purpose of continuous and periodic review policies. However, most studies were conducted under the assumption of infinite customer patience. A relatively small number of studies consider both these phenomena, i.e., perishable products and finite customer patience.

Patients in need of blood transfusions are typically impatient customers. Other practical examples of impatient customers can be found in retail stores and service
systems such as restaurants and hair salons. Typical behaviors of impatient customers are “balking” and “reneging.” When a customer arrives at a store and the items they desire are unavailable, they may choose to depart immediately without taking the items (balking) or wait in a queue to receive the items at a later time. A customer waiting in a queue can also exhibit impatience. They may choose leave the system before receiving their items if they feel they have waited too long (reneging). Customer reneging and product expiration are similar phenomena. A customer whose wait time expires withdraws their order from the backlog queue; likewise, a product whose lifetime has expired is removed from inventory. Queuing systems with impatient customers have been studied by many researchers [1, 3, 4, 13, 26, 27].

Perry and Stadje [21] studied inventory systems having both perishable items with exponential or deterministic lifetimes and impatience customer having due times (reneging times). Their model was based on finite shelf size and finite waiting room for demands. Chakravarthy and Daniel [6] studied an \((s,S)\) inventory system where demands occur according to a Markovian arrival process (MAP), replenishing times follow a phase-type distribution, and products have exponential lifetimes. They also permitted back orders up to a certain level. Ioannidis et al. [9] studied inventory-production systems with both perishable items and impatient customers under the base stock and base backlog policies. A base stock policy controls inventory, and a base backlog policy controls pending orders.

In this study, we consider an \((s,S)\) perishable inventory model with impatient customers and random lead times. A Markov process with a level-dependent \(G/M/1\)-type structure is used to represent this inventory model. In particular, we consider an \((s,S)\) continuous review inventory model where items are perishable and have exponentially distributed lifetimes. We assume that the replenishment lead times are independent and identically distributed with a phase-type distribution. Because of the diversity of phase-type distributions, we are able to find a phase-type distribution that is sufficiently close to an arbitrary distribution of nonnegative random variables [15, 19, 20]. We also consider a customer’s impatience. The impatience phenomena of customers waiting in a queue is natural in real-world situations. Our model is a generalization of previously considered models; for example, the models in [10, 14, 17] are all special cases of our model. The impatience characteristics of customers in a queue and a replenishment time with a phase-type distribution lead us to represent the inventory model as a continuous-time Markov process with a level-dependent \(G/M/1\)-type structure. As seen in Theorem A.1, the stationary distribution of the Markov process with a level-dependent \(G/M/1\)-type structure is very complicated; therefore, it is not useful for computational purposes. The main contribution of our research is that we find an approximation method for the stationary distribution of our inventory system.

The remainder of this paper is organized as follows. In Section 2, we describe our main inventory model. We find an approximation method for the stationary distribution of our inventory system and discuss several steady state performance measures in Section 3. Numerical examples are shown in Section 4, and concluding remarks are presented in Section 5. In appendix, we give the matrix analytic properties of a continuous-time Markov process with level-dependent \(G/M/1\)-type structure, which is the basic structure for our analysis.

2. Model description. We consider an \((s,S)\) continuous review inventory model with perishable items, impatient customers, and random lead times (Figure 1).
Customers arrive according to a Poisson process with rate $\lambda$. Each arriving customer requests one unit of product. When stock of finished products is available in the system, an arriving customer is immediately satisfied. During stock-out periods, arriving customers may balk, i.e., decide not to place orders, and leave the system with probability $1 - p$, where $0 \leq p \leq 1$. In addition to balking, customer impatience may lead to reneging. We assume that when a customer finds out that the product they desire is out of stock, they may not balk with probability $p$ and are willing to wait a set amount of time, called the due time, before their order is shipped. The due time is the difference between a customer’s deadline to receive a product and the customer’s arrival time. We assume that due times are independent and exponentially distributed with mean $1/\gamma$. Products cannot be indefinitely stored. Each product has a random duration of time before it perishes. If it is still in stock after this time elapses, it becomes unusable and is discarded. The time between the production and expiration of a product is referred to as its lifetime. Product lifetimes are assumed to be exponentially distributed with mean $1/\mu$.

Orders for items proceed according to the rule of $(s,S)$ policy. The order and supply procedures are described as follows. If the number of items in inventory is less than or equal to $s$, $S - s$ items are ordered. The ordered items are delivered after a random lead time with the phase-type distribution $(\alpha, T)$, where $\alpha$ is the $m$-dimensional row vector representing the initial distribution of the phase, and $T$ is the $m \times m$ matrix representing the generator of the transient Markov process for the phase-type distribution. For the definition and properties of the phase-type distribution, see Chapter 2 in [19] or Chapter 2 in [15]. If the number of items in inventory is still less than $s$ after the delivery of the ordered items, another order of $S - s$ items occurs immediately. If the number of items in the inventory is larger than $s$ after the delivery of ordered items, the next order of $S - s$ items occurs when the number of items in inventory decreases to $s$. We also assume that lead times are independent and identically distributed.

The state of the system at time $t$ is described by its inventory/backlog status and phase of lead time, denoted by $\{I(t), X(t)\}$. When the system is out of stock, the value of $I(t)$ is negative and $-I(t)$ customer orders are backlogged and pending. $I(t) = 0$ means there are no pending orders and no stock, and $1 \leq I(t) \leq S$ when

![Figure 1. Inventory Model](image)
there are $I(t)$ items in stock. By $(s,S)$ policy properties, when $I(t) > s$, $X(t) = 0$ and $I(t) \leq s$, $X(t) \in I$, where $l = \{1, 2, \ldots, m\}$. Therefore, let $\{(I(t), X(t)) : t \geq 0\}$ be a continuous-time Markov process with state space $E = \{(n, i) : n \leq S, i \in l(n)\}$, where $l(n)$ is denoted by

$$l(n) = \begin{cases} \{0\}, & \text{if } n > s, \\ \{1, 2, \ldots, m\}, & \text{if } n \leq s. \end{cases}$$

The generator $Q$ for the Markov process, in block form, is given by,

$$Q = \begin{bmatrix} S & S-1 & \cdots & S-s+1 & S-s & \cdots & s-1 & s-2 & \cdots & -1 & \cdots \\ S-1 & B_S & A_S & A_{S-1} & & & & & & & \\ \vdots & & & & & & & & & & \vdots \\ S-s & & & & & & & & & & \\ s & D_s & D_{s-1} & & & & & & & & \\ s-1 & & & & & & & & & & \\ \vdots & & & & & & & & & & \vdots \\ -2 & & & & & & & & & & \\ -1 & & & & & & & & & & \\ \vdots & & & & & & & & & & \vdots \\ & & & & & & & & & & \end{bmatrix}, \quad (1)$$

where

$$A_n = \begin{cases} \lambda n \mu, & s + 2 \leq n \leq S, \\ \{\lambda + (s + 1)\mu\} \alpha, & n = s + 1, \\ (\lambda + n \mu)I, & 0 < n \leq s, \\ \lambda p I, & n \leq 0, \end{cases}$$

$$B_n = \begin{cases} -\lambda n \mu, & s < n \leq S, \\ T - (\lambda + n \mu)I, & 0 < n \leq s, \\ T - (\lambda p - n \gamma I), & n \leq 0, \end{cases}$$

$$C_n = \begin{cases} 0, & n \geq 0, \\ -n \gamma I, & n < 0, \end{cases}$$

$$D_n = \begin{cases} T_0, & 2s - S < n \leq s, \\ T_0 \alpha, & n \leq 2s - S. \end{cases}$$

Here, $T_0 \equiv -T e$ where $e$ is a column vector of all ones with appropriate size. Though the elements (submatrices) of generator $Q$ are not square, the basic property of $Q$, $Qe = 0$, is satisfied. When $p = 0$, generator $Q$ is determined by taking the submatrix of level $S$ by level 0 in the right-hand side of (1).

The following observations can be made: i) when $p = 0$, $\{(I(t), X(t))\}$ is a finite Markov process, ii) when $0 < p \leq 1$ and $\gamma = 0$, $\{(I(t), X(t))\}$ is a Markov process with a level-independent $G/M/1$-type structure, except for a finite number of levels, and iii) when $0 < p \leq 1$ and $\gamma > 0$, $\{(I(t), X(t))\}$ is a Markov process with a level-dependent $G/M/1$-type structure. In this study, we will focus on the last case and verify the stability condition for the second case.

3. **Analysis of an $(s,S)$ inventory policy model.** In this section, we show the ergodicity of the Markov process $\{(I(t), X(t))\}$, which guarantees the existence of the steady-state distribution $\pi$. We also propose an approximation method to calculate the stationary distribution $\pi$. 
Theorem 3.1. When $\gamma > 0$, the Markov process $\{(I(t), X(t))\}$ is ergodic.

Proof. It is straightforward to verify that the Markov process $\{(I(t), X(t))\}$ is irreducible. It remains to be shown that it is positive recurrent. Let $\{(I_n, X_n) : n = 0, 1, 2, \cdots\}$ be the embedded Markov chain (EMC) of $\{(I(t), X(t))\}$ immediately after the transition epochs. It suffices to show that the EMC $\{(I_n, X_n)\}$ is positive recurrent, since

$$\inf_{(k,i) \in E} |Q_{(k,i)(k,i)}| > 0.$$  

Now, we show that the EMC $\{(I_n, X_n)\}$ is positive recurrent using Mustafa’s criterion [7]. Define $f : E \to [0, \infty)$ by $f(n,i) = S - n$. Then,

$$x_{k,i} = E[f(I_{n+1}, X_{n+1}) - f(I_n, X_n) | (I_n, X_n) = (k,i)]$$

has finite value for all $(k,i) \in E$, and

$$\lim_{k \to -\infty} x_{k,i} = \lim_{k \to -\infty} \frac{\lambda p - (-k)\gamma - (S - s)(T_0)_i}{\lambda p + (-T_i) + (-k)\gamma} = -1,$$

where $(T_0)_i$ denotes the $i$th component of the column vector $T_0$. Hence, $x_{k,i} < -\epsilon$ except for finitely many $(k,i) \in E$ for some $\epsilon > 0$ (for example, $\epsilon = \frac{1}{2}$). Therefore, the EMC $\{(I_n, X_n)\}$ is positive recurrent by Mustafa’s criterion.

This theorem shows that our model works similar to infinite server system if $\gamma > 0$ since backlogged orders are satisfied with receiving ordered items and are disappeared by their reneging properties, which each backlogged items have i.i.d. reneging times.

Theorem 3.2. When $\gamma = 0$, the Markov process $\{(I(t), X(t))\}$ is ergodic if and only if

$$\lambda p d < S - s,$$

where $d = -\alpha T^{-1} e$ is the mean lead time.

Proof. Note that $C_k = 0$ for $k \leq 0$, $A_k$ and $B_k$ are independent of $k$ for $k \leq 0$, and $D_k$ is independent of $k$ for $k \leq 2s - S$. Therefore, we use $A$, $B$, and $D$ instead of $A_k$, $B_k$ ($k \leq 0$), and $D_k$ ($k \leq 2s - S$), respectively. Let $Q^* = A + B + D$. Then, $Q^* = T + T_0 \alpha$ is a generator for an irreducible Markov process since the representation $(\alpha, T)$ is irreducible. Let $\pi^*$ be the unique probability vector satisfying $\pi^* Q^* = 0$ and $\pi^* e = 1$. By Theorem 1.71 in [19], $\{(I(t), X(t))\}$ is ergodic if and only if

$$\pi^* A e < (S - s) \pi^* D e.$$  \hfill (3)

Since $A = \lambda p I$, the left-hand side of (3) becomes $\lambda p$. Since $\pi^* = \frac{1}{\alpha T^{-1} e} \alpha T^{-1}$ and $D = T_0 \alpha$, the right-hand side of (3) becomes

$$\frac{S - s}{\alpha T^{-1} e} \alpha T^{-1} T_0 \alpha e = \frac{S - s}{\alpha T^{-1} e} \alpha T^{-1} (T_0) e = \frac{S - s}{\alpha T^{-1} e}.$$ 

Since $d = -\alpha T^{-1} e$, (2) is obtained from (3).

Theorem 3.3 says that the stability condition of our model is same as batch service system when $\gamma = 0$. In this case, there is no reneging orders and our model is operated by a $(s, S)$ inventory policy. Hence this system works as a batch service system, whose stability condition is that service capacity is greater arrival rate. And, the effective arrival rate of this system is $\lambda p$ and service capacity is $(S - s)/d$. 

\textbf{AN $(s, S)$ INVENTORY MODEL WITH G/M/1-TYPE STRUCTURE 613}
The Markov process \{\{(I(t), X(t))\}\} has the level-dependent, G/M/1-type structure described in Appendix. In order to find the stationary distribution \(\pi = (\pi_S, \pi_{S-1}, \cdots)\), we define the rate matrices \(R_n\) and the first passage probability matrix \(G_n\) in a way similar to that given in Appendix.

In this model, \((R_n)_{ij}\) is \(-\{(B_{n+1})_{ij}\}\) times the expected time spent in state \((n, j)\) before the first return to level \(n+1\), given that the Markov process starts at \((n+1, i)\). Note that \(R_n\) is a \(1 \times 1\) matrix if \(s < n < S\), a \(1 \times m\) matrix if \(n = s\), and an \(m \times m\) matrix if \(n < s\). \((G_n)_{ij}\) represents the probability that the first hitting state is \((n, j)\) in the class of states whose levels are larger than or equal to \(n\), given that the initial state is \((n-1, i)\), i.e., for \(n = S, S-1, \cdots\).

The stationary distribution \(\pi = (\pi_S, \pi_{S-1}, \cdots)\) is expressed as follows (see (16) and (17) in Section 2):

\[
\pi_n = \pi_S \prod_{i=S-1}^{n} R_i \\
\pi_S = \left( 1 + \sum_{n=s+1}^{S-1} \prod_{i=S-1}^{n} R_i + \sum_{n=-\infty}^{s} \prod_{i=S-1}^{n} R_i e \right)^{-1},
\]

where \(\prod_{i=1}^{m} A_i = A_n A_{n-1} \cdots A_m\) when \(n \geq m\). Since this process has an infinite state space and \(R_n\) is difficult to calculate directly, we approximate \(R_n\) using the following theorem and corollary. First, we introduce matrices \(G_n, n = S, S-1, \cdots\), whose \((i, j)\) entry \((G_n)_{ij}\) represents the probability that the first hitting state is \((n, j)\) in the class of states whose levels are larger than or equal to \(n\), given that the initial state is \((n-1, i)\); that is, for \(n = S, S-1, \cdots\),

\[
(G_n)_{ij} = P\{(I(\tau), X(\tau)) = (n, j)| (I(0), X(0)) = (n-1, i)\},
\]

where \(\tau = \inf\{t \geq 0 : I(t) \geq n\}\). The following theorem states properties of \(G_n\) and \(R_n\).

**Theorem 3.3.** The following equalities hold for for \(\gamma > 0\):

i) \(\lim_{n \to -\infty} (G_n)_{ij} = \delta_{ij}, 1 \leq i, j \leq m,\) where \(\delta_{ij}\) equals 1 if \(i = j\) and 0 if \(i \neq j\).

ii) \(R_n = \lambda p (\gamma I - T)^{-1} + o\left(\frac{1}{n^2}\right),\) as \(n \to -\infty,\) and for all \(i, j \in \{1, \cdots, m\},\)

\[
\lim_{n \to -\infty} n^2 \left\{ (R_n)_{ij} - \lambda p (\gamma I - T)^{-1} \right\} = 0.
\]

iii) \(R_n = -A_{n+1}(B_n + R_n-1 C_{n-1} + \prod_{i=n-1}^{n-(S-s)} R_i D_{n-(S-s)})^{-1}.\) \hspace{1cm} (6)

**Proof.** i) Given \((I(0), X(0)) = (n-1, i)\), with probability \(\frac{(1-n)^{\gamma}}{\lambda p + (1-n)^{\gamma} - T_{ii}}\), the Markov process jumps to \((n, i)\). Hence,

\[
(G_n)_{ii} \geq \frac{(1-n)^{\gamma}}{\lambda p + (1-n)^{\gamma} - T_{ii}}.
\]

Since \((G_n)_{ii} \leq 1\), we have

\[
\lim_{n \to -\infty} (G_n)_{ii} = 1, \text{ for } i = 1, \cdots, m.
\]

Since \(\sum_{j=1}^{m} (G_n)_{ij} \leq 1\), we have

\[
0 \leq (G_n)_{ij} \leq 1 - (G_n)_{ii} \text{ for } j \neq i.
\]
Hence,
\[ \lim_{n \to -\infty} (G_n)_{ij} = 0 \text{ for } j \neq i. \]

ii) It follows from (22) in Section 2 that
\[ R_n = -A_{n+1}(B_n + A_n G_n)^{-1} = \lambda p \left( -n \gamma I - T + \lambda p(I - G_n) \right)^{-1}, \quad n < 0. \]
Hence, for \( n < 0 \),
\[ R_n - \lambda p \left( (-n) \gamma I - T \right)^{-1} = \lambda p \left( (-n) \gamma I - T + \lambda p(I - G_n) \right)^{-1} - \lambda p \left( (-n) \gamma I - T \right)^{-1} \]
\[ = \lambda p \left( (-n) \gamma I - T \right)^{-1} \left( \left\{ I + \lambda p(I - G_n) \left( (-n) \gamma I - T \right)^{-1} \right\}^{-1} - I \right) \]
\[ = \lambda p \left( (-n) \gamma I - T \right)^{-1} \left\{ I + \lambda p(I - G_n) \left( (-n) \gamma I - T \right)^{-1} \right\}^{-1} - \lambda p(G_n - I) \left( (-n) \gamma I - T \right)^{-1}. \]

Since
\[ \lim_{n \to -\infty} (-n) \left( (-n) \gamma I - T \right)^{-1} = \frac{1}{\gamma} I \]
and
\[ \lim_{n \to -\infty} (G_n - I) = 0, \]
property ii) holds. Property iii) is proved by (23) in Section 3.

From Theorem 3.3 (ii), we have the following corollary:

**Corollary 3.1.** \( R_n \) has the following limiting behavior as \( k \) tends to negative infinity:
\[ \lim_{n \to -\infty} (-n) R_n = \frac{\lambda p}{\gamma} I. \]

Theorem 3.3 (ii) enables us to reasonably approximate \( R_n \) as \( \tilde{R}_n \):
\[ \tilde{R}_n \equiv \lambda p \left( (-n) \gamma I - T \right)^{-1}, \quad \text{for } n \leq -N. \]

We now determine a large value of \( N \) to make the relative error sufficiently small. For a predetermined small \( \epsilon > 0 \), we determine \( N \) such that
\[ \frac{||\tilde{R}_n - R_n||}{||R_n||} < \epsilon \quad \text{if } n \leq -N, \]
where the matrix norm \( ||\cdot|| \) is defined as
\[ ||K|| = \max_{ij} |K_{ij}|, \quad \text{for any matrix } K. \]

Let \( |K| \) denote the matrix whose \((i, j)\)th entry is \( |K_{ij}| \) for any matrix \( K \). Then, by (8), we have
\[ |\tilde{R}_n - R_n| \leq |\tilde{R}_n| \left\{ I + |(I - G_n)\tilde{R}_n| + |(I - G_n)\tilde{R}_n|^2 + \cdots \right\} ||I - G_n|| \tilde{R}_n||, \quad (10) \]
for \( n \leq -N \) where the inequality between matrices is interpreted entrywise. For \( 1 \leq i \leq m, \) since
\[ \sum_{j \neq i} |(I - G_n)_{ij}| \leq ||I - G_n||_i \leq \frac{\lambda p - T_{ii}}{\lambda p + (1 - n)\gamma - T_{ii}}, \]
from (7), we have

\[(I - G_n) \hat{R}_n)_{ij} \leq (I - G_n)_{ii} (\hat{R}_n)_{ii} \leq a_n \| \hat{R}_n \|\]

and

\[(I - G_n) \hat{R}_n)_{ij} \geq -\sum_{i \neq i} (I - G)_{di} \| \hat{R}_n \| \geq -a_n \| \hat{R} \|,\]

where \(a_n = \max_i \frac{\lambda p - T_i}{\sigma p + (1 - n) \gamma - T_i}\).

Thus,

\[| (I - G_n) \hat{R}_n \| \leq a_n \| \hat{R}_n \| \| e \| t, \]

where \(e\) denotes the \(m\)-dimensional row vector of all ones. By (10), if \(ma_n \| \hat{R}_n \| < 1\) and \(n \leq -N\),

\[| \hat{R}_n - R_n \| \leq \| \hat{R}_n \| e \| t \left( \sum_{i=0}^{\infty} (a_n \| \hat{R}_n \| e \| t )^i \right) a_n \| \hat{R}_n \| e \| t = \frac{ma_n \| \hat{R}_n \|^2 \| e \| t}{1 - ma_n \| R_n \|} \]

and

\[\| \hat{R}_n - R_n \| \leq \frac{ma_n \| \hat{R}_n \|^2}{1 - ma_n \| R_n \|}.\]

Therefore, if \(2ma_n \| \hat{R}_n \| < 1\) and \(n \leq -N\),

\[\frac{\| \hat{R}_n - R_n \|}{\| R_n \|} \leq \frac{\| \hat{R}_n - R_n \|}{\| R_n \| - \| R_n - R_n \|} \leq \frac{ma_n \| \hat{R}_n \|}{1 - 2ma_n \| R_n \|}.\]

Hence, we choose \(N\) such that, if \(n \leq -N\),

\[0 \leq \frac{ma_n \| (\gamma a I - T)^{-1} \|}{1 - 2ma_n \| (\gamma a I - T)^{-1} \|} < \epsilon. \tag{11}\]

We calculate \(\hat{R}_n\) \((n \leq -N)\), using (9) and calculate \(\hat{R}_n\) \((n = 1 - N, 2 - N, \cdots, S - 1)\), recursively by (6) substituting \(\hat{R}_n\) in place of \(R_n\) \((n = k - 1, k - 2, \cdots, k - (S - s))\). The approximate stationary distribution \(\pi = (\pi_1, \pi_2, \cdots)\) for \(\pi = (\pi_1, \pi_2, \cdots)\) is obtained by (4) and (5) with \(\hat{R}_n\) instead of \(R_n\) \((n = S - 1, S - 2, \cdots)\).

In summary, the approximate distribution \(\hat{\pi} = (\hat{\pi}_1, \hat{\pi}_2, \cdots)\) is given by the following algorithm:

- **STEP 1**: For a given sufficiently small \(\epsilon > 0\), choose \(N\) such that (11) is satisfied.
- **STEP 2**: Calculate \(\hat{R}_n, n = -N, -N - 1, \cdots, -N - (S - s)\) by (9).
- **STEP 3**: Calculate \(\hat{R}_n, n = 1 - N, 2 - N, \cdots, S - 1\), recursively by

\[\hat{R}_n = -A_{n+1}(B_n + \hat{R}_{n-1}C_{n-1} + \prod_{i=n-1}^{n-(S-s)} \hat{R}_i D_{n-(S-s)})^{-1}.\]

- **STEP 4**: Calculate \(\hat{\pi}_S\) by

\[\hat{\pi}_S = (1 + \sum_{n=s+1}^{S-1} \prod_{i=S-1}^{n} \hat{R}_i + \sum_{n=-N-(S-s)}^{s} \prod_{i=S-1}^{n} \hat{R}_i e_i)^{-1}.\]

- **STEP 5**: Calculate \(\hat{\pi}_n, n = S - 1, S - 2, \cdots, \) by

\[\hat{\pi}_n = \hat{\pi}_{n+1} \hat{R}_n.\]

From the stationary distribution \(\pi\) of \(\{(I(t), X(t))\}\), we obtain the performance measures of the system as follows:
• Long run supply rate $r_s$ of items:
  \[ r_s = (S - s) \sum_{n=-\infty}^{s} \pi_n T_0. \]

• Mean number of items in inventory $E_i$ at steady state:
  \[ E_i = \sum_{n=1}^{s} n\pi_n e + \sum_{n=s+1}^{S} n\pi_n. \]

• Perishing probability $P_p$ of an item:
  \[
  P_p = \frac{\text{long run perishing rate of items}}{\text{long run supply rate of items}} = \frac{\mu E_i}{r_s} = \frac{\mu \left\{ \sum_{n=1}^{s} n\pi_n e + \sum_{n=s+1}^{S} n\pi_n \right\}}{(S - s) \sum_{n=-\infty}^{s} \pi_n T_0}.
  \]

• Blocking probability $P_b$ of a customer’s demand:
  \[ P_b = \lim_{t \to \infty} P\{I(t) \leq 0\} \text{ (by PASTA)} = \sum_{n=-\infty}^{0} \pi_n e. \]

• Mean number of waiting customers $E_d$ at steady state:
  \[ E_d = \sum_{n=-\infty}^{0} (-n)\pi_n e. \]

• Loss probability $P_l$ of a demand:
  \[ P_l = P\{\text{Immediate loss}\} + P\{\text{loss after waiting time}\} = (1 - p)P_b + \frac{\text{long run loss rate of waiting customers}}{\text{long run occurrence rate of demands}} \]
  \[ = (1 - p) \sum_{n=-\infty}^{0} \pi_n e + \frac{\gamma E_d}{\lambda} \]
  \[ = \sum_{n=-\infty}^{0} \left\{ (1 - p) + \frac{\gamma}{\lambda} (-n) \right\} \pi_n e. \]

• Mean inventory holding time $E_h$ of an item:
  \[ E_h = \frac{E_i}{r_s} \text{ (by Little’s formula)} = \frac{\sum_{n=1}^{s} n\pi_n e + \sum_{n=s+1}^{S} n\pi_n}{(S - s) \sum_{n=-\infty}^{s} \pi_n T_0}. \]

• Mean waiting time $E_w$ of a delayed demand:
  \[ E_w = \frac{E_d}{\lambda p} \text{ (by Little’s formula)} = \frac{\sum_{n=-\infty}^{-1} (-n)\pi_n e}{\lambda p P_b} \]
  \[ = \frac{\sum_{n=-\infty}^{-1} (-n)\pi_n e}{\lambda p \sum_{n=-\infty}^{0} \pi_n e}. \]
4. Numerical examples. In this section, we discuss the problem of finding the optimal \((s, S)\) values to minimize the cost rate. Moreover, we perform a sensitivity analysis under the cost structure with the following cost factors:

- \(C_h\): The inventory holding cost per unit item per time unit.
- \(C_r\): The replacement cost per unit perished item.
- \(C_b\): The blocking cost per unit blocked demand.
- \(C_w\): The delay cost per unit demand per unit waiting time.
- \(C_l\): The loss cost per unit loss of demand.
- \(C_o\): The ordering cost per order.
- \(C_p\): The purchase cost per unit item.

With the above cost factors, the total cost rate (TCR), which is the expected total cost per unit time, is expressed in terms of the system parameters and the performance measures derived in Section 3 as follows:

\[
TCR = C_h E_i + C_r r_S P_p + C_b b P_b + C_w w E_w + C_l \lambda P_l + C_o \frac{r_S}{S - s} + C_p r_S.
\]

Although we have not established analytically, our experience with considerable numerical examples indicates the function \(TCR\), for fixed \(S\), to be convex in \(s\). Hence we adopted simple numerical search procedure to determine the optimal values of \(s^*\) and \(S^*\).

Table 1 shows the total cost rate with the following system parameters and cost factors.

- \(p = 0.5\), \(C_h = 3\), \(\lambda = 1\), \(C_r = 100\), \(\mu = 0.02\), \(C_b = 50\), \(\gamma = 0.1\), \(C_w = 5\), \(\alpha = [1, 0]\), \(C_l = 200\), \(T = \begin{bmatrix} -1 & 1 \\ 0 & -1 \end{bmatrix}\), \(C_o = 300\), \(C_p = 100\).

From Table 1, we determine that the optimal values of \((s, S)\) are \(s^* = 3\) and \(S^* = 8\) (choosing \(\epsilon = 0.01\) implies \(N = 162\)). The optimal cost rate is 164.21.

Next, we discuss the effects of \(\lambda\), \(\mu\), and \(\gamma\) on the current total cost rate function and the value of parameters. Figure 2 shows the trend of the optimal total cost rate.

| \(S\) | \(s\) | \(1\) | \(2\) | \(3\) | \(4\) | \(5\) | \(6\) | \(7\) | \(8\) |
|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| 2   | 214.85 |
| 3   | 193.68  | 192.23  |
| 4   | 182.28  | 180.09  |
| 5   | 175.77  | 171.17  | 170.76  | 174.86  |
| 6   | 172.06  | 167.51  | 166.41  | 168.46  | 173.94  |
| 7   | 170.09  | 165.83  | 164.59  | 165.86  | 169.41  | 175.65  |
| 8   | 169.28  | 165.37  | 164.21  | 165.19  | 167.94  | 172.36  | 178.93  |
| 9   | 169.85  | 166.67  | 165.86  | 166.79  | 169.01  | 172.21  | 176.28  | 181.34  |
| 10  | 170.85  | 168.02  | 167.40  | 168.40  | 170.55  | 173.54  | 177.21  | 181.51  |
| 11  | 172.19  | 169.67  | 169.24  | 170.33  | 172.48  | 175.36  | 178.79  | 184.45  |
| 12  | 173.78  | 171.55  | 171.31  | 172.49  | 174.66  | 177.49  | 180.79  | 184.45  |
| 13  | 175.57  | 173.60  | 173.54  | 174.83  | 177.02  | 179.83  | 183.06  | 186.58  |
| 14  | 177.53  | 175.80  | 175.90  | 177.28  | 179.53  | 182.34  | 185.52  | 188.96  |
with the value of $\lambda$. The optimal values of $(s, S)$ are shown in the supplementary column axis. In this graph, we see that the optimal total cost rate increases as the value of $\lambda$ increases. This same trend can be observed for the value of $(s, S)$ as well. The difference between $S$ and $s$ decreases as the value of $\lambda$ increases. Moreover, Fig. 2 shows that the base-stock policy ($S = s + 1$) is an optimal strategy for a moderate value of $\lambda$.

Figure 3 shows the relationship between the value of $\mu$ and the optimal total cost rate. Notice this relationship is also positive; however, the optimal value of $S$ decreases as the value of $\mu$ decreases. This phenomenon is quite natural since a product with a short lifetime is hard to keep in stock, thus increasing the ordering and penalty costs for unmet demands.

Finally, Fig. 4 shows that the value of $\gamma$ has minimal effects on the optimal total cost rate and optimal value of $(s, S)$. Furthermore, observe that total cost rate
5. **Conclusion.** In this paper, we introduced an \((s,S)\) inventory system for perishable items, impatient customers, and phase-type distributed lead times. We represented our model as a Markov process with a level-dependent \(G/M/1\)-type structure. We analyzed the level-dependent \(G/M/1\)-type process using an approximate method, and solved an optimization problem.

Both the finiteness of items and impatience of waiting customers are important and natural in an inventory model. In fact, a relatively small number of studies have addressed both of them. One major contribution of this paper is that we presented a method to study the impact of the perishability of items and impatience of waiting customers together when \((s,S)\) policy is adopted in an inventory system. Our second major contribution is that we provided an approximate method for analyzing a Markov process with level-dependent \(G/M/1\)-type structure.

In the future, we will extend this model with phase-type distributed perishable times and reneging times. This extended model may be easily applied in real world problems because of diversity of phase-type distribution. And we also consider an \((s,S)\) inventory model with repeated demands. In select real inventory systems, such as warehouse inventory, a demand not satisfied immediately does not wait, but returns at a later time.

**Appendix A. Level-dependent \(G/M/1\)-type structure.** In appendix, we derive several matrix analytic properties for Markov processes with a level-dependent \(G/M/1\)-type structure. This Markov process is a generalization of the Markov process with a level-dependent quasi-birth-and-death (QBD) structure [5] and the Markov process with a level-independent \(G/M/1\)-type structure [19, 20].

Let \(\{(N(t), J(t)) : t \geq 0\}\) be an irreducible, continuous-time Markov process with state space \(E = \{(n, i) : n \geq 0, i = 1, 2, \cdots, m_{\text{t}}\}\), where \(n\) denotes the level of the process and \(i\) denotes the phase. Suppose \(\{(N(t), J(t))\}\) is positive recurrent and
has the infinitesimal generator $Q$ of the level-dependent $G/M/1$-type structure:

$$Q = \begin{bmatrix}
A_1^{(0)} & A_0^{(0)} & 0 & 0 & 0 & \cdots \\
A_1^{(1)} & A_1^{(1)} & A_0^{(1)} & 0 & 0 & \cdots \\
A_1^{(2)} & A_2^{(2)} & A_1^{(2)} & A_0^{(2)} & 0 & \cdots \\
\vdots & \vdots & \vdots & \ddots & \ddots & \ddots \\
\vdots & \vdots & \vdots & \ddots & \ddots & \ddots \\
\end{bmatrix},$$

where $A_j^{(i)}$ is an $m_i \times m_{i+1-j}$ matrix.

The stationary distribution $\pi$ of a continuous Markov process with infinitesimal generator $Q$ is given by the system of equations $\pi Q = 0$ and $\pi e = 1$, where $e$ is a column vector of ones. Consider subvectors $\pi_0, \pi_1, \cdots$ of the stationary distribution $\pi$ such that $\pi_i$ is a row vector of length $m_i$. Then, the equations of the system $\pi Q = 0$ for the stationary distribution $\pi$ of a continuous-time Markov process can be expressed as

$$\sum_{n=1}^{\infty} \pi_n A_{n-i}^{(n)} = 0, \quad i = 0, 1, 2, \cdots.$$ 

In the general matrix analytic approach [15], there exists a matrix $\{R_n : n \geq 1\}$ such that $\pi_{n+1} = \pi_n R_{n+1}$ for all $n$, and $R_n$ can be interpreted as a rate matrix whose $(i,j)$th entry $(R_n)_{ij}$ ($1 \leq i \leq m_{n-1}, 1 \leq j \leq m_n$) is defined as $-(A_{ij}^{(n-1)})_{ii}$ times the expected time spent in state $(n,j)$ before the first return to level $n-1$, given that the Markov process starts at $(n-1,i)$. Formally,

$$(R_n)_{ij} = -\left(A_{ij}^{(n-1)}\right)_{ii} \cdot E\left[\int_{\tau_1}^{\tau_2} 1_{\{(N(t),J(t))=(n,j)\}} dt | (N(0),J(0)) = (n-1,i)\right],$$

where

$$\tau_1 = \inf\{t > 0 : (N(t),J(t)) \neq (n-1,i)\},$$

$$\tau_2 = \inf\{t > \tau_1 : N(t) = n-1\}.$$ 

Next, we introduce the first passage probability matrices $G_k^{(n)}$ whose $(i,j)$th entry $(G_k^{(n)})_{ij}$ is the probability that the first hitting state is $(k,j)$ in the class of states whose levels are less than or equal to $n-1$, given that the initial state is $(n,i)$ for $n > k$. That is,

$$(G_k^{(n)})_{ij} = P\{(N(\tau),J(\tau)) = (k,j) | (N(0),J(0)) = (n,i)\},$$

where

$$\tau = \inf\{t > 0 : N(t) \leq n-1\}.$$ 

Let $\{(N^{[0,n]}(t),J^{[0,n]}(t)) : t \geq 0\}$ be the restricted process of $\{(N(t),J(t)) : t \geq 0\}$ for the subset $\{(k,i) \in E : 0 \leq k \leq n\}$. For the restricted process of a Markov process, refer to pp. 13 in [2] or pp. 126 in [15]. Furthermore, let $Q^{[0,n]}$ be the generator for the restricted process $\{(N^{[0,n]}(t),J^{[0,n]}(t))\}$, which is partitioned as submatrices:

$$Q^{[0,n]} = \begin{bmatrix}
Q_{00}^{[0,n]} & Q_{01}^{[0,n]} & 0 & 0 & 0 \\
Q_{10}^{[0,n]} & Q_{11}^{[0,n]} & Q_{12}^{[0,n]} & 0 & 0 \\
\vdots & \vdots & \ddots & \ddots & \ddots \\
Q_{n-1,0}^{[0,n]} & Q_{n-1,1}^{[0,n]} & \cdots & Q_{n-1,n}^{[0,n]} \\
Q_{n0}^{[0,n]} & Q_{n1}^{[0,n]} & \cdots & Q_{nn}^{[0,n]} \\
\end{bmatrix},$$
Theorem A.1. \( \pi_n, R_n, U_n, \) and \( G^{(n)}_k \) satisfy the following relations where \( k < n \):

\[
\begin{align*}
\pi_{n+1} &= \pi_n R_{n+1}, \\
\pi_0 U_0 &= 0, \quad \pi_0 \left( e_0 + \sum_{i=1}^{n} \prod_{i=1}^{n} R_i e_n \right) = 1, \\
U_n &= A_1^{(n)} + A_0^{(n)} G^{(n+1)}_n, \\
&= A_1^{(n)} + \sum_{i=n+1}^{\infty} \left( \prod_{i=n+1}^{n} R_i \right) A_{i-n+1}^{(l)}, \\
G^{(n)}_k &= -\sum_{i=n}^{\infty} U_n^{-1} \left( \prod_{i=n+1}^{n} R_i \right) A_{i-k+1}^{(l)}, \\
R_n &= -A_0^{(n-1)} U_n^{-1}, \\
&= -A_0^{(n-1)} \left( A_1^{(n)} + A_0^{(n)} G^{(n+1)}_n \right)^{-1}, \\
&= -A_0^{(n-1)} \left( A_1^{(n)} + \sum_{i=n+1}^{\infty} \left( \prod_{i=n+1}^{n} R_i \right) A_{i-n+1}^{(l)} \right)^{-1}.
\end{align*}
\]

Proof. Equation (16) can be deduced from (5.33) in [15] by letting \( T \) and \( D \) be \( \{(n, i) : 1 \leq i \leq m_n\} \) and \( \{(n + 1, i) : 1 \leq i \leq m_{n+1}\} \), respectively.

The stationary probability vector of the restricted process \( \{(N^{[0,0]}(t), J^{[0,0]}(t))\} \) is proportional to \( \pi_0 \) (refer to pp. 13 in [2] or Theorem 5.5.3 in [15]). Hence, \( \pi_0 Q^{[0,0]} = 0 \). Since \( U_0 = Q^{[0,0]} \), we have \( \pi_0 U_0 = 0 \). From (16), we have \( \pi_n = \pi_0 \prod_{i=1}^{n} R_i \) when \( n \geq 1 \). Hence, the normalization condition of the stationary probability vector \( \pi \) gives the second equality in (17).

From the right skip-free property of level \( N(t) \), we deduce that

\[
Q_{kl}^{[0,n]} = \begin{cases} Q_{kl}, & \text{if } k < n, l \leq n, \\
Q_{nl} + Q_{n,n+1} G^{(n+1)}_l, & \text{if } k = n, l \leq n,
\end{cases}
\]

where \( Q_{kl} \) is the submatrix of \( Q \) corresponding to the transition rates from level \( k \) to level \( l \). Hence, this yields (18). Equation (19) is obtained by substituting (20) and (21) into the first equality in (17).

For the derivation of (20), let \( n_1 P_{nl}(t) \) (for \( n \leq l \) be the \( m_n \times m_l \) matrix whose \((i, i')\)th entry is defined as

\[
(n_1 P_{nl}(t))_{ii'} = P\{(N(t), J(t)) = (l, i'), N(s) \neq n-1 \text{ for all } s \in [0,t] | (N(0), J(0)) = (n, i)\}.
\]

Then, for \( 0 \leq k \leq n - 1 \),

\[
P\{(N(\tau), J(\tau)) = (k, j), \tau \in [t, t + \Delta t], (N(\tau-), J(\tau-)) = (l, i')\} = (n_1 P_{nl}(t))_{ii'} (A^{(l)}_{i-k+1})_{ij} \Delta t + o(\Delta t),
\]

where \( Q^{[0,n]}_{kl} \) (\( 0 \leq k, l \leq n \)) is an \( m_k \times m_l \) matrix corresponding to the transition rates from level \( k \) to level \( l \). Next, let \( U_n = Q^{[0,n]}_{nn} \). Below, we use the convention that \( \prod_{i=n}^{n-1} x_i = 1 \).
where \( \tau \) is given by (15). Hence,

\[
(G^{(n)}_k)_{ij} = \sum_{l=n}^{\infty} \sum_{i'=1}^{m_l} \int_0^\infty (n-1)P_{nl}(t)(A^{(l)}_{i'-k+1})_{ii'}dt.
\]

Thus,

\[
G^{(n)}_k = \sum_{l=n}^{\infty} \left( \int_0^\infty (n-1)P_{nl}(t)dt \right) A^{(l)}_{l-k+1}.
\] (24)

It should be noted that \( \int_0^\infty (n-1)P_{nl}(t)dt \) is the expected sojourn time of \((N(t), J(t))\) in state \((l, j)\) before the first visit to level \(n - 1\), given that \((N(0), J(0)) = (n, i)\). On the other hand, \(-U^{-1}_{n-1}ij\) is the expected sojourn time of \((N(t), J(t))\) in state \((n, j)\) before the first visit to level \(n - 1\), given that \((N(0), J(0)) = (n, i)\). Thus, by the probabilistic interpretation of \(R_{n+1}, \cdots, R_l\), it is easy to see that

\[
\int_0^\infty (n-1)P_{nl}(t)dt = -U^{-1}_{n-1} \prod_{i=n+1}^{l} R_i.
\] (25)

Substituting (25) into (24) leads to (20). From (12), we have

\[
(R_n)_{ij} = -(A^{(n-1)}_1)_{ii} \cdot E \left[ \int_{\tau_1}^{\tau_2} 1_{\{(N(t), J(t))=(n,j)\}} |(N(0), J(0)) = (n - 1, i) dt \right]
\]

\[
= -(A^{(n-1)}_1)_{ii} \sum_{i'=1}^{m_n} P \left\{ (N(\tau_1), J(\tau_1)) = (n, i')(N(0), J(0)) = (n - 1, i) \right\}
\]

\[
\cdot E \left[ \int_{\tau_1}^{\tau_2} 1_{\{(N(t), J(t))=(n,j)\}} |(N(\tau_1), J(\tau_1)) = (n, i') dt \right]
\]

\[
= -(A^{(n-1)}_1)_{ii} \sum_{i'=1}^{m_n} \left( A^{(n-1)}_0 \right)_{ii'} (-U^{-1}_{n-1})_{i'j}
\]

\[
= -(A^{(n-1)}_0 U^{-1}_{n-1})_{ij},
\]

where \( \tau_1 \) and \( \tau_2 \) are given by (13) and (14), respectively. Therefore by substituting (19) into (21), we obtain (21), (22), and (23).

\[\square\]

REFERENCES

[1] E. Altman and A. A. Borovkov, On the stability of retrial queues, Queueing Syst., 26 (1997), 343–363.
[2] S. Asmussen, Applied Probability and Queues, John Wiley & Sons, 1987.
[3] A. Brandt and M. Brandt, Asymptotic results and a markovian approximation for the M(n)/M(n)/s queue with impatient calls, Perform. Eval., 35 (1999), 1–18.
[4] A. Brandt and M. Brandt, Asymptotic results and a markovian approximation for the M(n)/M(n)/s + GI system, Queueing Syst., 41 (2002), 73–94.
[5] L. Bright and P. G. Taylor, Calculating the equilibrium distribution in level dependent quasi-birth-and-death processes, Commun. Statist. - Stochastic Models, 11 (1995), 497–525.
[6] S. Charkravarthy and J. Daniel, A markovian inventory system with random shelf time and back orders, Computers and Industrial Engineering, 47 (2004), 315–337.
[7] G. I. Falin, On sufficient conditions for ergodicity of multichannel queueing systems with repeated calls, Adv. Appl. Prob., 16 (1984), 447–448.
[8] Qi-Ming He, E. M. Jewkes and J. Buzzacott, The value of information used in inventory control of a make-to-order inventory-production system, IIE Transactions, 34 (2002), 999–1013.
[9] S. Ioannidis, O. Jouini, A. A. Economopoulos and V. S. Kouikoglou, Control policies for single-stage production systems with perishable inventory and customer impatience, *Annals of Operations Research*, (2012), 1–24.

[10] S. Kalpakam and K. P. Sapna, Continuous review \((s, S)\) inventory system with random lifetimes and positive leadtimes, *Operations Research Letters*, 16 (1994), 115–119.

[11] S. Kalpakam and K. P. Sapna, \((S - 1, S)\) perishable systems with stochastic lead times, *Mathematical and Computer Modelling*, 21 (1995), 95–104.

[12] I. Karaesmen, A. Scheller-Wolf and B. Deniz, Managing perishable and aging inventories: Review and future research directions, In *Planning Production and Inventories in the Extended Enterprise*, Springer, (2011), 393–438.

[13] A. Krishnamoorthy, K. P. Jose and V. C. Narayanan, Numerical investigation of a \(PH/PH/1\) inventory system with positive service time and shortage, *Neural Parallel & Scientific Comp.*, 16 (2008), 579–591.

[14] S. Kumaraswamy and E. Sankarasubramanian, A continuous review of \((S - s)\) inventory systems in which depletion is due to demand and failure of units, *Journal of Operational Research Society*, 32 (1981), 997–1001.

[15] G. Latouche and V. Ramaswami, *Introduction to Matrix Analytic Methods in Stochastic Modeling*, ASA-SIAM series on statistics and applied probability, 1999.

[16] L. Liu, \((s, S)\) continuous review models for inventory with random lifetimes, *Operations Research Letters*, 9 (1990), 161–167.

[17] L. Liu and T. Yang, An \((s, S)\) random lifetime inventory model with a positive lead time, *European Journal of Operational Research*, 112 (1999), 52–63.

[18] S. Nahmias, *Perishable inventory theory: A review*, *Operational Research*, 30 (1982), 680–708.

[19] M. F. Neuts, *Matrix-Geometric Solutions in Stochastic Models: An Algorithmic Approach*, The Johns Hopkins University Press, Baltimore, MD., 1981.

[20] M. F. Neuts, *Structured Stochastic Matrices of \(M/G/1\) Type and Their Applications*, Marcel Dekker, Inc., 1989.

[21] D. Perry and W. Stadje, Perishable inventory systems with impatient demands, *Math. Meth. of OR*, 50, (1999), 77–90.

[22] G. P. Prestacos, Blood inventory management, *Management Science*, 30 (1984), 777–801.

[23] M. Raafat, Survey of literature on continuously deteriorating inventory models, *Journal of Operational Research Society*, 42 (1991), 27–37.

[24] N. Ravichandran, Stochastic analysis of a continuous review perishable inventory system with positive lead time and Poisson demand, *European Journal of Operational Research*, 84 (1995), 444–457.

[25] C. P. Schmidt and S. Nahmias, \((S - 1, S)\) policies for perishable inventory, *Management Science*, 31 (1985), 719–728.

[26] A. R. Ward and P. W. Glynn, A diffusion approximation for a markovian queue with reneging, *Queueing Syst.*, 43 (2003), 103–128.

[27] S. Zeltyn and A. Mandelbaum, Call centers with impatient customers: Many-server asymptotics of the \(M/M/n + G\) queue, *Queueing Syst.*, 51 (2005), 361–402.

Received July 2014; 1st revision January 2015; 2nd revision March 2015.

E-mail address: janghakang@chosun.ac.kr