Chapter

Density Estimation in Inventory Control Systems under a Discounted Optimality Criterion

Jesús Adolfo Minjárez-Sosa

Abstract

This chapter deals with a class of discrete-time inventory control systems where the demand process $\{D_t\}$ is formed by independent and identically distributed random variables with unknown density. Our objective is to introduce a suitable density estimation method which, combined with optimal control schemes, defines a procedure to construct optimal policies under a discounted optimality criterion.

Keywords: discounted optimality, density estimation, inventory systems, optimal policies, Markov decision processes

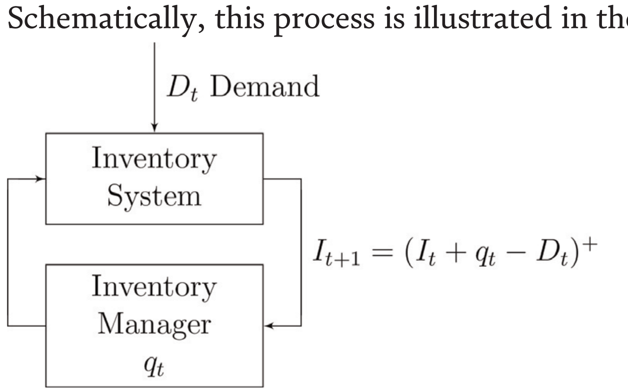
AMS 2010 subject classifications: 93E20, 62G07, 90B05

1. Introduction

Inventory systems are one of the most studied sequential decision problems in the fields of operation research and operation management. Its origin lies in the problem of determining how much inventory of a certain product should be kept in existence to meet the demand of buyers, at a cost as low as possible. Specifically, the question is: How much should be ordered, or produced, to satisfy the demand that will be presented during a certain period? Clearly, the behavior of the inventory over time depends on the ordered quantities and the demand of the product in successive periods. Indeed, let I_t and q_t be the inventory level and the order quantity at the beginning of period t, respectively, and D_t be the random demand during period t. Then $\{I_t\}_{t\geq 0}$ is a stochastic process whose evolution in time is given as

$$I_{t+1} = \max\{0, I_t + q_t - D_t\} =: (I_t + q_t - D_t)^+, \quad t = 0, 1, \ldots$$

Schematically, this process is illustrated in the following figure.

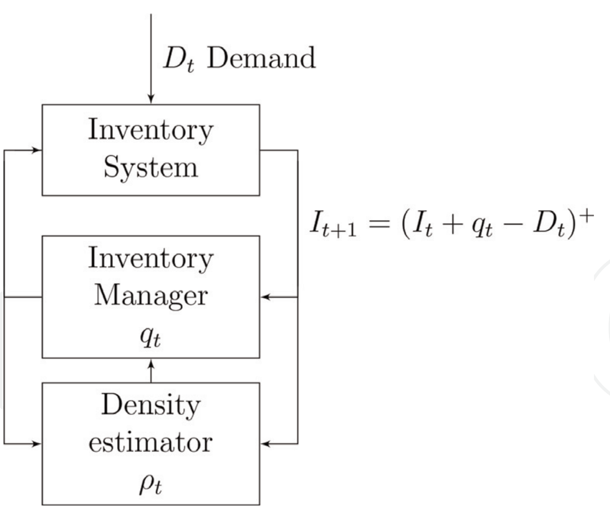


(Standard inventory system)

In this case, the inventory manager (IM) observes the inventory level I_t and then selects the order quantity q_t as a function of I_t. The order quantity process causes costs in the operation of the inventory system. For instance, if the quantity ordered is relatively small, then the items are very likely to be sold out, but there will be unmet demand. In this case the holding cost is reduced, but there is a significant cost due to shortage. Otherwise, if the size of the order is large, there is a risk of having surpluses with a high holding cost. These facts give rise to a stochastic optimization problem, which can be modeled as a Markov decision process (MDP). That is, the inventory system can be analyzed as a stochastic optimal control problem whose objective is to find the optimal ordering policy that minimizes a total expected cost.

The analysis of the control problem associated to inventory systems has been done under several scenarios: discrete-time and continuous-time systems with finite or infinite capacity, inventory systems considering bounded and unbounded one-stage cost, as well as partially observable models, among others (see, e.g., [1–5, 7]). Moreover, such scenarios have their own methods and techniques to solve the corresponding control problem. However, in most cases, it has been assumed that all the components that define the behavior of the inventory system are known to the IM, which, in certain situations, can be too strong and unrealistic. Hence it is necessary to implement schemes that allow learning or collecting information about the unknown components during the evolution of the system to choose a decision with as much information as possible.

In this chapter we study a class of inventory control systems where the density of the demand is unknown by the IM. In this sense, our objective is to propose a procedure that combines density estimation methods and control schemes to construct optimal policies under a total expected discounted cost criterion. The estimation and control procedure is illustrated in the following figure:



(Estimation and control procedure)

In this case, unlike the standard inventory system, before choosing the order quantity q_t, the IM implements a density estimation method to get an estimate ρ_t, and, possibly, combines this with the history of the system $h_t = (I_0, q_0, D_0, \ldots, I_{t-1}, q_{t-1}, D_{t-1}, I_t)$ to select $q_t = q_t(h_t, \rho_t)$. Specifically, the density of the demand is estimated by the projection of an arbitrary estimator on an appropriate set, and its convergence is stated with respect to a norm which depends on the components of the inventory control model.

In general terms, our approach consists in to show that the inventory system can be studied under the weighted-norm approach, widely studied by several authors in the field of Markov decision processes (see, e.g., [11] and references therein) and in adaptive control (see, e.g. [9, 12–14]). That is, we prove the existence of a weighted function W which imposes a growth condition on the cost functions. Then, applying the dynamic programming algorithm, the density estimation method is adapted to such a condition to define an estimation and control procedure for the construction of optimal policies.

The chapter is organized as follows. In Section 2 we describe the inventory model and define the corresponding optimal control problem. In Section 3 we introduce the dynamic programming approach under the true density. Next, in Section 4 we present the density estimation method which will be used to state, in Section 5, an estimation and control procedure for the construction of optimal policies. The proofs of the main results are given in Section 6. Finally, in Section 7, we present some concluding remarks.

2. The inventory model

We consider an inventory system evolving according to the difference equation

$$I_{t+1} = \left(I_t + q_t - D_t\right)^+, \quad t = 0, 1, \ldots, \tag{1}$$

where I_t and q_t are the inventory level and the order quantity at the beginning of period t, taking values in $\mathbb{I} := [0, \infty)$ and $\mathbb{Q} := [0, \infty)$, respectively, and D_t represents the random demand during period t. We assume that $\{D_t\}$ is an observable sequence of nonnegative independent and identically distributed (i.i.d.) random variables with a common density $\rho \in L_1[0, \infty)$ which is unknown by the inventory manager. In addition, we assume finite expectation

$$\overline{D} := E(D_t) < \infty. \tag{2}$$

Moreover, there exists a measurable function $\overline{\rho} \in L_1[0, \infty)$ such that

$$\rho(s) \leq \overline{\rho}(s) \tag{3}$$

almost everywhere with respect to the Lebesgue measure. In addition

$$\int_0^\infty s^2 \overline{\rho}(s) ds < \infty. \tag{4}$$

For example, if $\overline{\rho}(s) := K \min\{1, 1/s^{1+r}\}, s \in [0, \infty)$, for some positive constants K and r, then there are plenty of densities that satisfy (3)–(4).

The one-stage cost function is defined as

$$\tilde{c}(I, q, D) = cq + h(I + q - D)^+ + b(D - I - q)^+, \qquad (I, q) \in \mathbb{I} \times \mathbb{Q}, \tag{5}$$

where h, c, and b are, respectively, the holding cost per unit, the ordering cost per unit, and the shortage cost per unit, satisfying $b > c$.

The order quantities applied by the IM are selected according to rules known as ordering control policies defined as follows. Let $\mathcal{H}_t$ be the space of histories of the inventory system up to time t. That is, a typical element of $\mathcal{H}_t$ is written as

$$h_t = \left(I_0, q_0, D_0, \dots, I_{t-1}, q_{t-1}, D_{t-1}, I_t\right).$$

An ordering policy (or simply a policy) $\gamma = \{\gamma_t\}$ is a sequence of measurable functions $\gamma_t : \mathcal{H}_t \to \mathbb{Q},$ such that $\gamma_t(h_t) = q_t, t \geq 0$. We denote by Γ the set of all policies. A feedback policy or Markov policy is a sequence $\gamma = \{g_t\}$ of functions $g_t : \mathbb{I} \to \mathbb{Q},$ such that $g_t(I_t) = q_t$. A feedback policy $\gamma = \{g_t\}$ is stationary if there exists a function $g : \mathbb{I} \to \mathbb{Q}$ such that $g_t = g$ for all $t \geq 0$.

When using a policy $\gamma \in \Gamma,$ given the initial inventory level $I_0 = I$, we define the total expected discounted cost as

$$V(\gamma, I) := E\left[\sum_{t=0}^{\infty} \alpha^t \tilde{c}\left(I_t, q_t, D_t\right)\right], \tag{6}$$

where $\alpha \in (0,1)$ is the so-called discount factor. The inventory control problem is then to find an optimal feedback policy γ^* such that $V(\gamma^*, I) = V^*(I)$ for all $I \in \mathbb{I}$, where

$$V^*(I) := \inf_{\gamma \in \Gamma} V(\gamma, I), \quad I \in \mathbb{I}, \tag{7}$$

is the optimal discounted cost, which we call value function.

We define the mean one-stage cost as

$$\begin{aligned} c(I,q) &= cq + hE(I+q-D)^+ + bE(D-I-q)^+ \\ &= cq + h\int_0^{I+q} (I+q-s)^+ \rho(s)ds + b\int_{I+q}^{\infty} (s-I-q)^+ \rho(s)ds, \quad (I,q) \in \mathbb{I} \times \mathbb{Q}. \end{aligned} \tag{8}$$

Then, by using properties of conditional expectation, we can rewrite the total expected discounted cost (6) as

$$V(\gamma, I) = E_I^{\gamma}\left[\sum_{t=0}^{\infty} \alpha^t c\left(I_t, q_t\right)\right], \tag{9}$$

where E_I^{γ} denotes the expectation operator with respect to the probability P_I^{γ} induced by the policy $\gamma,$ given the initial inventory level $I_0 = I$ (see, e.g., [8, 10]).

The sequence of events in our model is as follows. Since the density ρ is unknown, the one-stage cost (8) is also unknown by the IM. Then if at stage t the inventory level is $I_t = I \in \mathbb{I},$ the IM implements a suitable density estimation method to get an estimate ρ_t of ρ. Next, he/she combines this with the history of the system to select an order quantity $q_t = q = \gamma_t^{\rho_t}(h_t) \in \mathbb{Q}$. Then a cost $c(I,q)$ is incurred, and the system moves to a new inventory level $I_{t+1} = I' \in \mathbb{I}$ according to the transition law

$$\begin{aligned} Q(B|I,q) &:= \operatorname{Prob}\left[I_{t+1} \in B | I_t = I, q_t = q\right] \\ &= \int_0^{\infty} 1_B\left((I+q-s)^+\right)\rho(s)ds \end{aligned} \tag{10}$$

where $1_B(.)$ denotes the indicator function of the set $B \in \mathcal{B}(\mathbb{I}),$ and $\mathcal{B}(\mathbb{I})$ is the Borel $\sigma-$algebra on $\mathbb{I}$. Once the transition to the inventory level I' occurs, the process is repeated. Furthermore, the costs are accumulated according to the discounted cost criterion (9).

3. Dynamic programming equation under the true density ρ

The study of the inventory control problem will be done by means of the well-known dynamic programming (DP) approach, which we now introduce in terms of the unknown density ρ. In order to establish precisely the ideas, we first present some preliminary and useful facts.

The set of order quantities in which we can find the optimal ordering policy should be $\mathbb{Q}^* = [0, Q^*] \subset \mathbb{Q}$,

where

$$Q^* = \frac{b\overline{D}}{c(1-\alpha)}.$$

Thus, we can restrict the range of q so that $q \in \mathbb{Q}^*$. Specifically we have the following result.

Lemma 3.1 *Let $\gamma^0 \in \Gamma$ be the policy defined as $\gamma^0 = \{0, 0, \ldots\}$, and let $\overline{\gamma} = \{\overline{\gamma}_t\}$ be a policy such that $\overline{\gamma}_k(h_k) = \overline{q}_k > Q^*$, for at least a $k = 0, 1, \ldots$. Then*

$$V(\gamma^0, I) \le V(\overline{\gamma}, I), \quad I \in \mathbb{I}. \tag{11}$$

That is, γ^0 is a better solution than $\overline{\gamma}$.

Proof. Let I_t^0, $t = 0, 1, \ldots,$ be the inventory levels generated by the application of γ^0, and $(\overline{I}_t, \overline{q}_t)$ be the sequence of inventory levels and order quantities generated by $\overline{\gamma}$, where $I_0^0 = \overline{I}_0 = I$, $I_{t+1}^0 = (I_t^0 - D_t)^+$, and $\overline{I}_{t+1} = (\overline{I}_t + \overline{q}_t - D_t)^+$, $t \ge 0$. Without loss of generality, we suppose that for a $\overline{q} > Q^*$ we have $\overline{q}_0 = \overline{q}$. Note that $I_t^0 \le \overline{I}_t$, for all $t \ge 0$. Then observing that $c\overline{q} > b\overline{D}/(1-\alpha)$,

$$\begin{aligned}
V(\gamma^0, I) &= E\left[\sum_{t=0}^{\infty} \alpha^t c(I_t^0, 0, D_t)\right] = E\left[\sum_{t=0}^{\infty} \alpha^t \left(h(I_t^0 - D_t)^+ + b(D_t - I_t^0)^+\right)\right] \\
&\le E\sum_{t=0}^{\infty} \alpha^t h(\overline{I}_t - D_t)^+ + b\sum_{t=0}^{\infty} \alpha^t E(D_t) \\
&\le E\left[\sum_{t=0}^{\infty} \alpha^t \left(h(\overline{I}_t + \overline{q}_t - D_t)^+ + b(D_t - \overline{I}_t - \overline{q}_t)^+ + \frac{b\overline{D}}{1-\alpha}\right)\right] \\
&\le E\left[\sum_{t=0}^{\infty} \alpha^t \left(h(\overline{I}_t + \overline{q}_t - D_t)^+ + b(D_t - \overline{I}_t - \overline{q}_t)^+ + c\overline{q}\right)\right] \\
&\le E\left[\sum_{t=0}^{\infty} \alpha^t \left(c\overline{q}_t + h(\overline{I}_t + \overline{q}_t - D_t)^+ + b(D_t - \overline{I}_t - \overline{q}_t)^+\right)\right] \\
&= V(\overline{\gamma}, I), \quad I \in \mathbb{I}. \blacksquare
\end{aligned}$$

Remark 3.2 *Observe that for $(I, q) \in \mathbb{I} \times \mathbb{Q}^*$ we have*

$$c(I, q) = cq + L(I + q),$$

where, by writing $y = I + q$,

$$L(y) := hE(y - D)^{+} + bE(D - y)^{+}.$$

In addition, observe that for any fixed $s \in [0, \infty)$, *the functions* $y \to (y - s)^{+}$ *and* $y \to (s - y)^{+}$ *are convex, which implies that* $L(y)$ *is convex. Moreover*

$$\lim_{y \to \infty} L(y) = \infty.$$

The following lemma provides a growth property of the one-stage cost function (8).

Lemma 3.3 *There exist a number* β *and a function* $W : \mathbb{I} \to [1, \infty)$ *such that* $0 < \alpha\beta < 1$,

$$\sup_{(I,q,s) \in \mathbb{I} \times \mathbb{Q}^{*} \times [0,\infty)} \frac{W\left((I + q - s)^{+}\right)}{W(I)} := \varphi < \infty, \tag{12}$$

and for all $(I, q) \in \mathbb{I} \times \mathbb{Q}^{*}$

$$c(I, q) \leq W(I). \tag{13}$$

In addition, for any density μ *on* $[0, \infty)$ *such that* $\int_0^\infty s\mu(s) < \infty$,

$$\int_0^\infty W\left((I + q - s)^{+}\right)\mu(s)ds \leq \beta W(I), \quad (I, q) \in \mathbb{I} \times \mathbb{Q}^{*}. \tag{14}$$

The proof of Lemma 3.3 is given in Section 6.

We denote by $\mathbb{B}_W$ the normed linear space of all measurable functions $u : \mathbb{I} \to \mathfrak{R}$ with finite weighted-norm ($W-$norm) $\|\cdot\|_W$ defined as

$$\|u\|_W := \sup_{I \in \mathbb{I}} \frac{|u(I)|}{W(I)}. \tag{15}$$

Essentially, Lemma 3.3 proves that the inventory system (1) falls within of the weighted-norm approach used to study general Markov decision processes (see, e.g., [11]). Hence, we can formulate, on the space $\mathbb{B}_W$, important results as existence of solutions of the DP-equation, convergence of the value iteration algorithm, as well as existence of optimal policies, in the context of the inventory system (1). Indeed, let

$$V^{(n)}(\gamma, I) = E_I^{\gamma}\left[\sum_{t=0}^{n-1} \alpha^t c(I_t, q_t)\right]$$

be the n-stage discounted cost under the policy $\gamma \in \Gamma$ and the initial inventory level $I \in \mathbb{I}$, and

$$V^{(n)}(I) = \inf_{\gamma \in \Gamma} V^{(n)}(\gamma, I); \quad V^{(0)}(I) = 0, \quad I \in \mathbb{I}$$

the corresponding value function. Then, for all $n \geq 0$ and $I \in \mathbb{I}$, (see, e.g., [6, 10, 11]),

$$V^{(n)}(I) = \min_{q \in \mathbb{Q}^*} \left\{ c(I,q) + \alpha \int_0^\infty V^{(n-1)}\left((I+q-s)^+\right)\rho(s)ds \right\} \tag{16}$$

Moreover, from [11, Theorem 8.3.6], by making the appropriate changes, we have the following result.

Theorem 3.4 (Dynamic programming) *(a) The functions* $V^{(n)}$ *and* V^* *belong to* $\mathbb{B}_W$. *Moreover*

$$V^{(n)}(I) \le \frac{W(I)}{1-\alpha\beta}, V^*(I) \le \frac{W(I)}{1-\alpha\beta}, I \in \mathbb{I}. \tag{17}$$

(b) As $n \to \infty$, $\left\| V^{(n)} - V^* \right\|_W \to 0$.

(c) V^* *is convex.*

(d) V^* *satisfies the dynamic programming equation:*

$$\begin{aligned} V^*(I) &= \min_{q \in \mathbb{Q}^*} \left\{ c(I,q) + \alpha \int_0^\infty V^*\left((I+q-s)^+\right)\rho(s)ds \right\} \\ &= \min_{I \le y \le Q^*+I} \left\{ cy + L(y) + \alpha \int_0^\infty V^*\left((y-s)^+\right)\rho(s)ds \right\} - cI, I \in \mathbb{I}. \end{aligned} \tag{18}$$

(e) There exists a function $g^* : \mathbb{I} \to \mathbb{Q}$ *such that* $g^*(I) \in \mathbb{Q}^*$ *and, for each* $I \in \mathbb{I}$,

$$V^*(I) = c(I, g^*(I)) + \alpha \int_0^\infty V^*\left((I + g^*(I) - s)^+\right)\rho(s)ds, I \in \mathbb{I}.$$

Moreover, $\gamma^* = \{g^*\}$ *is an optimal control policy.*

4. Density estimation

As the density ρ is unknown, the results in Theorem 3.4 are not applicable, and therefore they are not accessible to the IM. In this section we introduce a suitable density estimation method with which we can obtain an estimated DP-equation. This will allow us to define a scheme for the construction of optimal policies. To this end, let $D_0, D_1, \ldots, D_t, \ldots$ be independent realizations of the demand whose density is ρ.

Theorem 4.1 *There exists an estimator* $\rho_t(s) := \rho_t(s; D_0, D_1, \ldots, D_{t-1}),\ \ s \in 0, \infty)$, *of* ρ, *such that (see (2) and (3))*:

D.1. $\rho_t \in L_1[0,\infty)$ is a density.

D.2. $\rho_t \le \overline{\rho}(\cdot)$ a.e. with respect to the Lebesgue measure.

D.3. $\int_0^\infty s\rho_t(s)ds \le \overline{D}$.

D.4. $E\int_0^\infty |\rho_t - \rho(s)|ds \to 0, as\ \ t \to \infty$.

D.5. $E\|\rho_t - \rho\| \to 0,\ as\, t \to \infty,$ where

$$\|\mu\| := \sup_{(I,q) \in \mathbb{I} \times \mathbb{Q}^*} \frac{1}{W(I)} \int_0^\infty W\left((I+q-s)^+\right)\mu(s)ds \tag{19}$$

for measurable functions μ on $[0,\infty)$.

It is worth noting that for any density μ on $[0, \infty)$ satisfying (14), the norm $\|\mu\|$ is finite. The remainder of the section is devoted to prove Theorem 4.1.

We define the set $\mathcal{D} \subset L_1([0, \infty))$ as:

$$\mathcal{D} := \left\{ \mu : \mu \text{ is a density}, \quad \int_0^\infty s\mu(s)ds \leq \overline{D}, \quad \mu(s) \leq \overline{\rho}(s) \text{ a.s.} \right\}.$$

Observe that $\rho \in \mathcal{D}$.

Lemma 4.2 *The set $\mathcal{D}$ is closed and convex in $L_1([0, \infty))$.*

Proof. The convexity of $\mathcal{D}$ follows directly. To prove that $\mathcal{D}$ is closed, let $\mu_t \in \mathcal{D}$ be a sequence in $\mathcal{D}$ such that $\mu_t \xrightarrow{L_1} \mu \in L_1([0, \infty))$. First, we prove

$$\mu(s) \leq \overline{\rho}(s) \quad a.e. \tag{20}$$

We assume that there is $A \subset [0, \infty)$ with $m(A) > 0$ such that $\mu(s) > \overline{\rho}(s), s \in A,$ m being the Lebesgue measure on $\mathfrak{R}$. Then, for some $\varepsilon > 0$ and $A' \subset A$ with $m(A') > 0$,

$$\mu(s) > \overline{\rho}(s) + \varepsilon, s \in A'. \tag{21}$$

Now, since $\mu_t \in \mathcal{D}, t \geq 0,$ there exists $B_t \subset 0, \infty)$ with $m(B_t) = 0,$ such that

$$\mu_t(s) \leq \overline{\rho}(s), \quad s \in [0, \infty) \backslash B_t, t \geq 0. \tag{22}$$

Combining (21) and (22) we have

$$|\mu_t(s) - \mu(s)| \geq \varepsilon, \quad s \in A' \cap ([0, \infty) \backslash B_t), \quad t \geq 0.$$

Using the fact that $m(A' \cap ([0, \infty) \backslash B_t) = m(A') > 0$, we obtain that μ_t does not converge to μ in measure, which is a contradiction to the convergence in L_1. Therefore $\mu(s) \leq \overline{\rho}(s)$ a.e.

On the other hand, applying Holder's inequality and using the fact that $\overline{\rho} \in L_1[0, \infty)$, from (20),

$$\begin{aligned} \left|1 - \int_0^\infty \mu(s)ds\right| &= \left|\int_0^\infty \mu_t(s)ds - \int_0^\infty \mu(s)ds\right| = \int_0^\infty |\mu_t(s) - \mu(s)|^{\frac{1}{2}} |\mu_t(s) - \mu(s)|^{\frac{1}{2}} ds \\ &\leq \left(\int_0^\infty 2\overline{\rho}(s)\right)^{1/2} \left(\int_0^\infty |\mu_t(s) - \mu(s)|\right)^{1/2} \to 0 \text{ as } t \to \infty, \end{aligned} \tag{23}$$

which implies $\int_0^\infty \mu(s)ds = 1$. Now, as $\mu \geq 0$ $a.e.,$ we have that μ is a density. Similarly, from (4),

$$\begin{aligned} \int_0^\infty s|\mu_t(s) - \mu(s)|ds &= \int_0^\infty s|\mu_t(s) - \mu(s)|^{\frac{1}{2}} |\mu_t(s) - \mu(s)|^{\frac{1}{2}} ds \\ &\leq \left(\int_0^\infty s^2 2\overline{\rho}(s)ds\right)^{1/2} \left(\int_0^\infty |\mu_t(s) - \mu(s)|ds\right)^{1/2} \\ &\leq 2^{\frac{1}{2}} M' \left(\int_0^\infty |\mu_t(s) - \mu(s)|ds\right)^{1/2}, \end{aligned} \tag{24}$$

for some constant $M' < \infty$. Letting $t \to \infty$ we obtain

$$\int_0^\infty s\mu_t(s)ds \to \int_0^\infty s\mu(s)ds$$

which, in turn, implies that

$$\int_0^\infty s\mu(s)ds \le \overline{D}.$$

This proves that $\mathcal{D}$ is closed.∎

Let $\hat{\rho}_t(s) := \hat{\rho}_t(s; D_0, D_1, \ldots, D_t), \quad s \in [0, \infty)$, be an arbitrary estimator of ρ such that

$$E\|\rho - \hat{\rho}_t\|_{L_1} = E\int_0^\infty |\rho(s) - \hat{\rho}_t(s)| \to 0 \quad \text{as } t \to \infty. \tag{25}$$

Lemma 4.2 ensures the existence of the estimator ρ_t which is defined by the projection of $\hat{\rho}_t$ on the set of densities $\mathcal{D}$. That is, the density $\rho_t \in \mathcal{D}$, expressed as

$$\rho_t := \arg\min_{\sigma \in \mathcal{D}} \|\sigma - \hat{\rho}_t\|_{L_1},$$

is the "best approximation" of the estimator $\hat{\rho}_t$ on the set $\mathcal{D}$, that is,

$$\|\rho_t - \hat{\rho}_t\|_{L_1} = \inf_{\mu \in \mathcal{D}} \|\mu - \hat{\rho}_t\|_{L_1}. \tag{26}$$

Now observe that ρ_t satisfies the properties D.1, D.2, and D.3. Hence, Theorem 4.1 will be proved if we show that ρ_t satisfies D.4 and D.5. To this end, since $\rho \in \mathcal{D}$, from (26) observe that

$$\|\rho_t - \rho\|_{L_1} \le \|\rho_t - \hat{\rho}_t\|_{L_1} + \|\hat{\rho}_t - \rho\|_{L_1} \le 2\|\hat{\rho}_t - \rho\|_{L_1}, \ t \ge 0,$$

which implies that, from (25),

$$E\int_0^\infty |\rho(s) - \rho_t(s)|ds \le 2E\|\hat{\rho}_t - \rho\|_{L_1} \to 0, \text{ as } t \to \infty. \tag{27}$$

That is, ρ_t satisfies Property D.4. In fact, since $\int_0^\infty |\rho(s) - \rho_t(s)|ds \le 2$ a.s., from (27) it is easy to see that

$$E\left(\int_0^\infty |\rho(s) - \rho_t(s)|ds\right)^q \to 0, \text{ as } t \to \infty, \text{ for any } q > 0. \tag{28}$$

Now, to obtain property D.5, observe that from (12)

$$\|\rho_t - \rho\| = \sup_{(I,q) \in \mathbb{I} \times \mathbb{Q}^*} \frac{1}{W(I)} \int_0^\infty W\left((I + q - s)^+\right)|\rho(s) - \rho_t(s)|ds = \varphi \int_0^\infty |\rho(s) - \rho_t(s)|ds. \tag{29}$$

Therefore, property D.4 yields

$$E\|\rho_t - \rho\| \to 0, \quad \text{as } t \to \infty, \tag{30}$$

which proves the property D.5.

5. Estimation and control

Having defined the estimator ρ_t, we will now introduce an estimate dynamic programming procedure with which we can construct optimal policies for the inventory systems.

Observe that for each $t \geq 0$, from (14),

$$\int_0^\infty W\left((I+q-s)^+\right)\rho_t(s)ds \leq \beta W(I), \; (I,q) \in \mathbb{I} \times \mathbb{Q}^*. \tag{31}$$

Now, we define the estimate one-stage cost function:

$$\begin{aligned} c_t(I,q) &= cq + h\int_0^{I+q} (I+q-s)^+\rho_t(s)ds + b\int_{I+q}^\infty (s-I-q)^+\rho_t(s)ds \\ &= cq + L_t(I+q), \quad (I,q) \in \mathbb{I} \times \mathbb{Q}^*, \end{aligned} \tag{32}$$

where (see Remark 3.2) for $y = I + q$,

$$L_t(y) := h\int_0^y (y-s)^+\rho_t(s)ds + b\int_y^\infty (s-y)^+\rho_t(s)ds.$$

In addition, observe that for each $t \geq 0$, $L_t(y)$ is convex and

$$\lim_{y \to \infty} L_t(y) = \infty. \tag{33}$$

We define the sequence of functions $\{V_t\}$ as $V_0 \equiv 0$, and for $t \geq 1$

$$\begin{aligned} V_t(I) &= \min_{q \in \mathbb{Q}^*}\left\{c_t(I,q) + \alpha\int_0^\infty V_{t-1}\left((I+q-s)^+\right)\rho_t(s)ds\right\} \\ &= \min_{I \leq y \leq Q^*+I}\left\{cy + L_t(y) + \alpha\int_0^\infty V_{t-1}\left((y-s)^+\right)\rho_t(s)ds\right\} - cI, \quad I \in \mathbb{I}. \end{aligned} \tag{34}$$

We can state our main results as follows:

Theorem 5.1 *(a) For $t \geq 0$ and $I \in \mathbb{I}$,*

$$V_t(I) \leq \frac{W(I)}{1-\alpha\beta}. \tag{35}$$

Therefore, $V_t \in \mathbb{B}_W$.

(b) As $t \to \infty$, $E\left[\sup_{(I,q)\in \mathbb{I}\times\mathbb{Q}^*} \frac{|c_t(I,q)-c(I,q)|}{W(I)}\right] \to 0.$

(c) As $t \to \infty$, $E\|V_t - V^*\|_W \to 0.$

(d) For each $t \geq 0$, *there exists* $K_t \geq 0$ *such that the selector* $g_t : \mathbb{I} \to \mathbb{Q}$ *defined as*

$$q_t^* = g_t(I) := \begin{cases} K_t - I & \text{if} \quad 0 \leq I \leq K_t \\ 0 & \text{if} \quad I > K_t \end{cases}$$

attains the minimum in (34).

Remark 5.2 *From [10, Proposition D.7], for each* $I \in \mathbb{I}$, *there is an accumulation point* $g_\infty(I) \in \mathbb{Q}^*$ *of the sequence* $\{g_t(I)\}$. *Hence, there exists a constant* K^* *such that*

$$g_\infty(I) := \begin{cases} K^* - I & \text{if} \quad 0 \leq I \leq K^* \\ 0 & \text{if} \quad I > K^* \end{cases} \tag{36}$$

Theorem 5.3 *Let* g_∞ *be the selector defined in (36). Then the stationary policy* $\gamma^* := \{g_\infty\}$ *is an optimal base stock policy for the inventory problem.*

6. Proofs

6.1 Proof of Lemma 3.3

Note that, for each $(I,q) \in \mathbb{I} \times \mathbb{Q}^*$,

$$\begin{aligned} c(I,q) &\leq cQ^* + h(I+Q^*) + b\overline{D} \\ &\leq (c+h)Q^* + hI + b\overline{D} \leq MG(I), \end{aligned} \tag{37}$$

where $M := \max\{(c+h)Q^* + b\overline{D}, h\}$ and $G(I) = I + 1$. Moreover, for every density function μ on $[0,\infty)$ and $(I,q) \in \mathbb{I} \times \mathbb{Q}^*$,

$$\int_0^\infty G\big((I+q-s)^+\big)\mu(s)ds \leq G(I) + Q^*. \tag{38}$$

On the other hand, we define the sequence of functions $\{w_t\}$, $w_t : \mathbb{I} \to \mathfrak{R}$, as

$$w_0(I) := 1 + MG(I) \tag{39}$$

and for $t \geq 1$ and any density function μ on $[0,\infty)$

$$w_t(I) := \sup_{q \in \mathbb{Q}^*} \int_0^\infty w_{t-1}\big((I+q-s)^+\big)\mu(s)ds.$$

Observe that, for each $I \in \mathbb{I}$,

$$\begin{aligned} w_1(I) &= \sup_{q \in \mathbb{Q}^*} \int_0^\infty \big[1 + MG\big((I+q-s)^+\big)\big]\mu(s)ds \\ &\leq 1 + MG(I) + MQ^*. \end{aligned}$$

Thus,

$$w_2(I) = \sup_{q \in \mathbb{Q}^*} \int_0^\infty \left[1 + MG\left((I+q-s)^+\right) + MQ^*\right]\mu(s)ds$$
$$\leq 1 + MG(I) + MQ^* + MQ^*, \quad I \in \mathbb{I}.$$

In general, it is easy to see that for each $I \in \mathbb{I}$,

$$w_t(I) \leq MG(I) + 1 + \sum_{j=0}^{t-1} MQ^* = MG(I) + 1 + MQ^* t. \tag{40}$$

Let $\alpha_0 \in (\alpha, 1)$ be arbitrary, and define

$$W(I) := \sum_{t=0}^{\infty} \alpha_0^t w_t(I). \tag{41}$$

Then, from (40),

$$\begin{aligned} W(I) &\leq \sum_{t=0}^{\infty} \alpha_0^t [MG(I) + 1 + MQ^* t] \\ &= \sum_{t=0}^{\infty} \alpha_0^t (MG(I) + 1) + MQ^* \sum_{t=0}^{\infty} t\alpha_0^t \leq \frac{MG(I)+1}{1-\alpha_0} + \frac{MQ^*\alpha_0}{(1-\alpha_0)^2}. \end{aligned} \tag{42}$$

Therefore, $W(I) < \infty$ for each $I \in \mathbb{I}$, and since $w_0 > 1$, from (41),

$$W(I) > 1. \tag{43}$$

Furthermore, using (42) and the fact that $W(\cdot) \geq w_0(\cdot)$, a straightforward calculation shows that

$$\varphi := \sup_{(I,q,s) \in \mathbb{I} \times \mathbb{Q}^* \times 0,\, \infty)} \frac{W\left((I+q-s)^+\right)}{W(I)} < \infty. \tag{44}$$

Now, from (37) and (39), $c(I,q) \leq w_0(I)$, which yields, for all $(I,q) \in \mathbb{I} \times \mathbb{Q}^*$,

$$c(I,q) \leq W(I). \tag{45}$$

In addition, for every density function μ on $[0, \infty)$ and $(I,q) \in \mathbb{I} \times \mathbb{Q}^*$,

$$\begin{aligned} \int_0^\infty W\left((I+q-s)^+\right)\mu(s)ds &= \int_0^\infty \sum_{t=0}^{\infty} \alpha_0^t w_t\left((I+q-s)^+\right)\mu(s)ds \\ &= \sum_{t=0}^{\infty} \alpha_0^t \int_0^\infty w_t\left((I+q-s)^+\right)\mu(s)ds \\ &\leq \sum_{t=0}^{\infty} \alpha_0^t w_{t+1}(I) = \alpha_0^{-1}\left[\sum_{t=0}^{\infty} \alpha_0^t w_t(I) - w_0(I)\right] \\ &= \alpha_0^{-1}[W(I) - w_0(I)] \leq \alpha_0^{-1} W(I). \end{aligned}$$

Therefore, defining $\beta := \alpha_0^{-1}$, we have that $0 < \alpha\beta < 1$, and

$$\int_0^\infty W\big((I+q-s)^+\big)\mu(s)ds \le \beta W(I),\ (I,q) \in \mathbb{I} \times \mathbb{Q}^*,$$

which, together with (43), (44), and (45), proves Lemma 3.3.∎

6.2 Proof of Theorem 5.1

(a) Since $\int_0^\infty s\rho_t(s)ds \le \overline{D}$, from (32) (see (37)) $c_t(I,q) \le MG(I)$ for each $t \ge 0$, $(I,q) \in \mathbb{I} \times \mathbb{Q}^*$. Hence, it is easy to see that $c_t(I,q) \le W(I)$ for each $(I,q) \in \mathbb{I} \times \mathbb{Q}^*$ (see (45)). Then we have $V_1(I) \le W(I)$, and from (31), and by applying induction arguments, we get

$$V_t(I) \le \frac{W(I)}{1-\alpha\beta},\quad t \ge 0,\quad I \in \mathbb{I}. \tag{46}$$

(b) Observe that from (39), for each $I \in \mathbb{I}$,

$$W(I) \ge w_0(I) = 1 + MG(I),$$

which implies that (see (43))

$$\frac{MG(I)}{W(I)} \le 1 - \frac{1}{W(I)} < \infty. \tag{47}$$

In addition, from (37),

$$h(I + Q^*) \le MG(I). \tag{48}$$

On the other hand, similarly as (24), from (4), it is easy to see that

$$\int_0^\infty s|\rho_t(s) - \rho(s)|ds \le 2^{\frac{1}{2}}M'\left(\int_0^\infty |\rho_t(s)-\rho(s)|ds\right)^{1/2}, \tag{49}$$

for some constant $M' < \infty$. Hence, combining (47)–(49), from the definition of $c_t(I,q)$ and $c(I,q)$, we have

$$\begin{aligned}\frac{|c_t(I,q)-c(I,q)|}{W(I)} &\le \frac{h}{W(I)}\int_0^\infty (I+Q^*)|\rho_t(s)-\rho(s)|ds + \frac{b}{W(I)}\int_0^\infty s|\rho_t(s)-\rho(s)|ds \\ &\le \frac{MG(I)}{W(I)}\int_0^\infty |\rho_t(s)-\rho(s)|ds + b2^{\frac{1}{2}}M'\left(\int_0^\infty |\rho_t(s)-\rho(s)|ds\right)^{1/2}.\end{aligned}$$

Finally, taking expectation, (28) and Property D.4 prove the result.

(c) For each $I \in \mathbb{I}$ and $t \geq 0$, by adding and subtracting the term $\alpha \int_0^\infty V_{t-1}((I+q-s)^+)\rho(s)ds$, we have

$$|V_t(I) - V^*(I)| \leq \sup_{q \in \mathbb{Q}^*} |c_t(I,q) - (I,q)| + \sup_{q \in \mathbb{Q}^*} \alpha \int_0^\infty V_{t-1}((I+q-s)^+)|\rho_t(s) - \rho(s)|ds$$

$$+ \alpha \int_0^\infty \left| V_{t-1}((I+q-s)^+) - V^*((I+q-s)^+) \right| \rho(s)ds$$

$$\leq \sup_{q \in \mathbb{Q}^*} |c_t(I,q) - (I,q)| + \frac{\alpha}{1-\alpha\beta} \sup_{q \in \mathbb{Q}^*} \int_0^\infty W((I+q-s)^+)|\rho_t(s) - \rho(s)|ds$$

$$+ \alpha\beta \|V_{t-1} - V^*\|_W W(I),$$

where the last inequality is due to (35), (17), (14), and (15). Therefore, from (15) and (19) and by taking expectation,

$$E\|V_t - V^*\|_W \leq E \sup_{q \in \mathbb{Q}^*} |c_t(I,q) - c(I,q)| + \frac{\alpha}{1-\alpha\beta} E\|\rho_t - \rho\| + \alpha\beta E\|V_{t-1} - V^*\|_W. \tag{50}$$

Finally, from (17) and (35), $\eta := \limsup_{t\to\infty} E\|V^* - V_t\|_W < \infty$. Hence, taking limsup in both sides of (50), from part (a) and property D.5 in Theorem 4.1, we get $\eta \leq \alpha\beta\eta$, which yields $\eta = 0$ (since $0 < \alpha\beta < 1$). This proves (c).

(d) For each $t \geq 0$, let $H_t : \mathbb{I} \to \mathfrak{R}$ be the function defined as

$$H_t(y) := cy + L_t(y) + \alpha \int_0^\infty V_{t-1}((y-s)^+)\rho_t(s)ds.$$

Hence, (34) is equivalent to

$$V_t(I) = \min_{q \in \mathbb{Q}^*} H_t(I+q) - cI, \qquad I \in \mathbb{I}. \tag{51}$$

Moreover (see (33)), observe that H_t is convex and $\lim_{y\to\infty} H_t(y) = \infty$. Thus, there exist a constant $K_t \geq 0$ such that

$$H_t(K_t) = \min_{I \leq y \leq Q^* + I} H_t(y),$$

and

$$g_t(I) = \begin{cases} K_t - I & \text{if} \quad 0 \leq I \leq K_t \\ 0 & \text{if} \quad I > K_t \end{cases}$$

attains the minimum in (51).∎

6.3 Proof of Theorem 5.3

We fix an arbitrary $I \in \mathbb{I}$. Since $g_\infty(I)$ is an accumulation point of $\{g_t(I)\}$ (see Remark 5.2), there exists a subsequence $\{t_m(I)\}$ of $\{t\}$ ($t_m = t_m(I)$) such that

$$g_{t_m}(I) \to g_\infty(I) \quad \text{as } m \to \infty.$$

Moreover, from (34) and Theorem 5.1(d), letting $t_m = m$, we have

$$V_m(I) = c_m(I, g_m) + \alpha \int_0^\infty V_{m-1}\left(\left(I + g_m - s\right)^+\right)\rho_m(ds). \tag{52}$$

On the other hand, following similar arguments as the proof of Theorem 5.1(c), for each $m \geq 0$ and $(I, q) \in \mathbb{I} \times \mathbb{Q}$, we have

$$\begin{aligned}
&\left|\alpha\int_0^\infty V_{m-1}\left((I+q-s)^+\right)\rho_m(s)ds - \alpha\int_0^\infty V^*\left((I+q-s)^+\right)\rho(s)ds\right| \\
&\leq \alpha\int_0^\infty \left|V_{m-1}\left((I+q-s)^+\right) - V^*\left((I+q-s)^+\right)\right|\rho_m(s)ds + \alpha\int_0^\infty V^*\left((I+q-s)^+\right)|\rho_m(s)-\rho(s)|ds \\
&\leq \alpha\beta\|V_{m-1} - V^*\|_W W(I) + \frac{\alpha}{1-\alpha\beta}\|\rho_m - \rho\|.
\end{aligned}$$

Then, for each $I \in \mathbb{I}$,

$$E \sup_{q \in \mathbb{Q}^*}\left|\alpha\int_0^\infty V_{m-1}\left((I+q-s)^+\right)\rho_m(s)ds - \alpha\int_0^\infty V^*\left((I+q-s)^+\right)\rho(s)ds\right| \to 0, \quad \text{as } m \to \infty. \tag{53}$$

Now,

$$\begin{aligned}
&\alpha\int_0^\infty V_{m-1}\left(\left(I+g_m-s\right)^+\right)\rho_m(ds) \\
&\qquad = \left[\alpha\int_0^\infty V_{m-1}\left(\left(I+g_m-s\right)^+\right)\rho_m(ds) - \alpha\int_0^\infty V^*\left(\left(I+g_m-s\right)^+\right)\rho(s)ds\right] \\
&\qquad\quad + \alpha\int_0^\infty V^*\left(\left(I+g_m-s\right)^+\right)\rho(s)ds.
\end{aligned} \tag{54}$$

Taking expectation and liminf as $m \to \infty$ on both sides of (54), from (53) we obtain

$$\begin{aligned}
\liminf_{m\to\infty} E\alpha\int_0^\infty V_{m-1}\left(\left(I+g_m-s\right)^+\right)\rho_m(ds) &= \liminf_{m\to\infty} E\alpha\int_0^\infty V^*\left(\left(I+g_m-s\right)^+\right)\rho(s)ds \\
&\geq \int_0^\infty V^*\left(\left(I+g_\infty-s\right)^+\right)\rho(s)ds,
\end{aligned}$$

where the last inequality follows by applying Fatou's Lemma and because the function $q \to (I+q-s)^+$ is continuous. Hence, taking expectation and liminf in (52), we obtain

$$c(I,g_\infty) + \alpha \int_0^\infty V^*\left(\left(I + g_\infty - s\right)^+\right)\rho(s)ds \le V^*(I), \quad I \in \mathbb{I}. \tag{55}$$

As I was arbitrary, by (18), the equality holds in (55) for all $I \in \mathbb{I}$. To conclude, standard arguments on stochastic control literature (see, e.g., [10]) show that the policy $\gamma^* = \{g_\infty\}$ is optimal.■

7. Concluding remarks

In this chapter we have introduced an estimation and control procedure in inventory systems when the density of the demand is unknown by the inventory manager. Specifically we have proposed a density estimation method defined by the projection to a suitable set of densities, which, combined with control schemes relative to the inventory systems, defines a procedure to construct optimal ordering policies.

A point to highlight is that our results include the most general scenarios of an inventory system, e.g., state and control spaces either countable or uncountable, possibly unbounded costs, finite or infinite inventory capacity. This generality entailed the need to develop new estimation and control techniques, accompanied by a suitable mathematical analysis. For example, the simple fact of considering possibly unbounded costs led us to formulate a density estimation method that was related to the weight function W, which, in turn, defines the normed linear space $\mathbb{B}_W$ (see (15)), all this through the projection estimator. Observe that if the cost function c is bounded, we can take $W \equiv 1$ and we have $\|\cdot\| = \|\cdot\|_{L_1}$ (see (19) and (25)). Thus, any L_1−consistent density estimator ρ_t can be used for the construction of optimal ordering policies.

Finally, the theory presented in this chapter lays the foundations to develop estimation and control algorithms in inventory systems considering other optimality criteria, for instance, the average cost or discounted criteria with random state-action-dependent discount factors (see [14, 15] and references therein).

Author details

Jesús Adolfo Minjárez-Sosa
Departamento de Matemáticas, Universidad de Sonora, Hermosillo, Sonora, Mexico

*Address all correspondence to: aminjare@gauss.mat.uson.mx

References

[1] Arrow KJ, Karlin S, Scarf H. Studies in the Mathematical Theory of Inventory and Production. CA: Stanford University Press; 1958

[2] Bensoussan A, Çakanyıldırım M, Sethi SP. Partially observed inventory systems: The case of zero balance walk. SIAM Journal on Control and Optimization. 2007;**46**:176-209

[3] Bensoussan A, Çakanyıldırım M, Minjárez-Sosa JA, Royal A, Sethi SP. Inventory problems with partially observed demands and lost sales. Journal of Optimization Theory and Applications. 2008;**136**:321-340

[4] Bensoussan A, Çakanyıldırım M, Minjárez-Sosa JA, Sethi SP, Shi R. Partially observed inventory systems: The case of rain checks. SIAM Journal on Control and Optimization. 2008; **47**(5):2490-2519

[5] Bensoussan A, Çakanyıldırım M, Minjárez-Sosa JA, Sethi SP, Shi R. An incomplete information inventory model with presence of inventories or backorders as only observations. Journal of Optimization Theory and Applications. 2010;**146**(3):544-580

[6] Bertsekas DP. Dynamic Programming: Deterministic and Stochastic Models. Englewood Cliffs, N. J: Prentice-Hall; 1987

[7] Beyer D, Cheng F, Sethi SP, Taksar MI. Markovian Demand Inventory Models. New York: Springer; 2008

[8] Dynkin EB, Yushkevich AA. Controlled Markov Processes. New York: Springer-Verlag; 1979

[9] Gordienko EI, Minjárez-Sosa JA. Adaptive control for discrete-time Markov processes with unbounded costs: Discounted criterion. Kybernetika. 1998;**34**:217-234

[10] Hernández-Lerma O, Lasserre JB. Discrete-Time Markov Control Processes: Basic Optimality Criteria. New York: Springer-Verlag; 1996

[11] Hernández-Lerma O, Lasserre JB. Further Topics on Discrete-Time Markov Control Processes. New York: Springer-Verlag; 1999

[12] Hilgert N, Minjárez-Sosa JA. Adaptive policies for time-varying stochastic systems under discounted criterion. Mathematical Methods of Operations Research. 2001;**54**(3): 491-505

[13] Minjárez-Sosa JA. Approximation and estimation in Markov control processes under discounted criterion. Kybernetika. 2004;**6**(40):681-690

[14] Minjárez-Sosa JA. Empirical estimation in average Markov control processes. Applied Mathematics Letters. 2008;**21**:459-464

[15] Minjárez-Sosa JA. Markov control models with unknown random state-action-dependent discount factors. TOP. 2015;**23**:743-772