Inventory Control with a Fixed Ordering Cost and Demand Learning*

Alain Bensoussan[1], Abinash Khandelwal[2], Viswanath Ramakrishna[2] and Suresh Sethi[3]

***Abstract*— We study an inventory problem involving a fixed ordering cost and Bayesian demand learning. We obtain the optimal feedback rule, which is complex to implement. So, we find a policy that is optimal among (s,S) policies.**

I. INTRODUCTION

We study a discrete-time, infinite-horizon inventory problem involving a fixed ordering cost and unknown stationary demand, which we learn through Bayesian updating. See [1] for a study of this problem with no fixed cost. The demands D_n are i.i.d. random variables on $\mathbb{R}^+$ with density $g(x)$ and C.D.F. $G(x)=\int_0^x g(\zeta)d\zeta$. We let $\bar{G}(x)=1-G(x)$. We call v_n and x_n the order and the beginning inventory level in period n, respectively. To allow for full observation of the realized demands, backlogging is allowed.

Thus, the x_n evolves according to

$$\begin{aligned} x_{n+1} &= x_n+v_n-D_n, \\ x_1 &= x. \end{aligned} \tag{1}$$

The order $v_n=v(x_n)$ is a function of the beginning inventory level at period n. It is a decision of the Inventory Manager (IM), known as feedback. It is also called a decision rule. Denoting the feedback to be chosen by $V=v(\cdot)$, the expected total cost, also called the payoff, is

$$J_x(V)=E\sum_{n=1}^{\infty}\alpha^{n-1}(hx_n^+ + px_n^- + cv_n + K\mathbf{1}_{v_n>0}). \tag{2}$$

Here, $x_n^+=\max\{x_n,0\}$ and $x_n^-=\max\{-x_n,0\}$. We let p,h,c denote the unit storage cost, the unit holding cost, and the unit variable cost, respectively. K is the fixed cost incurred only when $v_n>0$. $\alpha\in(0,1)$ is the discount factor. When the distribution of the demand depends on an unknown parameter ζ, the IM has the opportunity to learn about this parameter through observing the realized demands. We aim to find the ordering policy that minimizes the total discounted cost.

II. REVISITING THE CLASSICAL PROBLEM WITH NO UNKNOWN PARAMETER

A. Classical (s,S) Policy

We first consider the classical case, as studied in [2], in which the demand is assumed to stem from an i.i.d. sequence with a distribution known to the IM. Thus, there is no learning in this case. We define the value function

$$\Phi(x)=\inf_V J_x(V). \tag{3}$$

Then, $\Phi(x)$ is a solution of the Bellman equation

$$\Phi(x)=hx^+ + px^- + \inf_{v\geq 0}[cv+K\mathbf{1}_{v>0}+\alpha E\Phi(x+v-D)]. \tag{4}$$

Corresponding to the idea that the shortage cost must be sufficiently large to incentivize ordering, we assume

$$\alpha p - c(1-\alpha)>0. \tag{5}$$

The solution is obtained as follows. There are two numbers s and S, with $s<S$. The optimal feedback is then

$$\hat{v}(x)=\begin{cases} S-x & \text{if } x\leq s \\ 0 & \text{if } x>s \end{cases} \tag{6}$$

The solution of (4) is given by

$$\Phi(x)=hx^+ + px^- - cx + \frac{l(s)}{1-\alpha}, \text{ for } x\leq s \tag{7}$$

with $\overline{D}=ED$ and

$$l(x)=cx(1-\alpha)+\alpha c\overline{D}+\alpha hE(x-D)^+ + \alpha pE(x-D)^-. \tag{8}$$

We can then express $l(x)$ in terms of the function $\overline{G}(x)$ as

$$\begin{aligned} l(x) = {} & (c(1-\alpha)+\alpha h)x+\alpha(h+p)x^- + \\ & \alpha(h+p)\int_{x^+}^{+\infty}\overline{G}(z)dz+\alpha(c-h)\int_0^{+\infty}\overline{G}(z)dz. \end{aligned} \tag{9}$$

For $x>s$, $\Phi(x)$ is a solution of the linear integral equation

$$\Phi(x)=hx^+ + px^- + \alpha E\Phi(x-D). \tag{10}$$

Or, equivalently

$$\Phi(x)=hx^+ + px^- + \alpha\int_0^{+\infty}\Phi(x-\zeta)g(\zeta)d\zeta. \tag{11}$$

It is easy to check from the definition of the function $l(x)$ that the function $\Phi(x)$ is continuous at s. But it is not C^1, and we have

$$\Phi'(s+0)-\Phi'(s-0)=l'(s). \tag{12}$$

If we fix s, the function $\Phi(x)$, satisfying (7) and (11), is well defined. We then introduce $Z(x)$, defined by

$$Z(x)=\Phi(x)-(hx^+ + px^- - cx), \tag{13}$$

which satisfies the following equations

$$Z(x)=\frac{l(s)}{1-\alpha}, \text{ if } x\leq s, \tag{14}$$

$$Z(x)=l(x)+\alpha\int_0^{+\infty}Z(x-\zeta)g(\zeta)d\zeta, \text{ if } x>s. \tag{15}$$

To define s and S, we add the conditions

$$Z'(S)=0, \tag{16}$$

*Research supported by the NSF grant DMS-2204795

[1]ICDRiA, The University of Texas at Dallas, USA
[2]Dept. of Mathematical Sciences, The University of Texas at Dallas, USA
[3]Naveen Jindal School of Management, The University of Texas at Dallas, USA sethi@utdallas.edu

$$Z(s) = K + Z(S), s < S. \tag{17}$$

The claim is that we can solve the system $s, S, Z(x)$ satisfying (14), (15), (16), and (17). Then, the function $\Phi(x)$ defined by (13) is the solution of the Bellman equation (4) and the optimal feedback is defined by (6).

B. Basic Result

The following result is not new, but its proof is new and not based on the K-convexity theory introduced by H. Scarf [3].

Theorem 1 : We assume (5). There exists a triple, $s, S, Z(x)$, satisfying (14), (15), (16), and (17). The function $\Phi(x)$ defined by (13) is a solution of the Bellman equation (4) and also the value function (3).

Sketch of the proof : We show that the function $l(x)$ is convex and identify its minimum point. Then, we make use of a contraction mapping argument to solve an integral equation for a potential reorder point s. It leads to the following equation

$$Z_s(x) = \frac{l(s)}{1-\alpha} + H_s(x).$$

Here, $H_s(x)$ is expressed using a renewal-like function $\Gamma(x)$. We then prove and make use of the following important property.

Proposition 2 :

$$B_s(x) \leq B_s(s) \tag{18}$$

where

$$B_s(x) = H_s(x) - \inf_{y>x} H_s(y),$$

and the function $H_s(x)$ is the solution of

$$H_s(x) = \begin{cases} l(x) - l(s) + \alpha \int_0^{+\infty} H_s(x-\zeta) g(\zeta) d\zeta, \; x > s \\ 0, \; x \leq s \end{cases}$$

The proposition proves that the *gap* between $H_s(x)$ and its future infimum is bounded. This ensures that not ordering is optimal when the inventory level is above s. Next, we show that the specific values of s and S can be obtained by solving

$$0 = K + H_s(S(s)) = K - B_s(s)$$

where $S(s) = S$.

Finally, to complete the proof, we show that this constructed $Z(x)$ satisfies the Bellman equation.

III. LEARNING AND CONTROL

Let $g(x|\mu)$ denote the demand density depending on the unknown parameter μ, which is learned by observing the demand D_n. We introduce the filtration $\mathscr{F}^n = \sigma(D_1, \cdots, D_n)$, $n \geq 1$. The order at time n, v_n, is $\mathscr{F}^{n-1}$ measurable. At the beginning of the first time period, the only information available is that the inventory level $x_1 = x$. However, the IM has an initial belief on the parameter ζ, which we express by a probability density $f(\zeta)$, taken on $\mathbb{R}$ or $\mathbb{R}^+$ to fix ideas. This belief will evolve through the learning process. From the inventory evolution equation

$$\begin{aligned} x_{n+1} &= x_n + v_n - D_n, \\ x_1 &= x, \end{aligned} \tag{19}$$

we see that the inventory process x_n is also $\mathscr{F}^{n-1}$ measurable.

The updated belief $f_{n+1}(\zeta)$ after observing demand D_n is obtained via Bayes' theorem. Given the current belief $f_n(\zeta)$, the posterior density is given by

$$f_{n+1}(\zeta) = f_n(\zeta) \frac{g(D_n|\zeta)}{\int g(D_n|\mu) f_n(\mu)\, d\mu}, \tag{20}$$

with initial belief

$$f_1(\zeta) = f(\zeta).$$

The pair $(x_n, f_n(\cdot))$ is a stochastic process adapted to $\mathscr{F}^{n-1}$. Setting $y_n = x_n + v_n$, we have the following Controlled Markov Chain with state $(x_n, f_n(\cdot))$:

$$x_{n+1} = y_n - D_n \;, \; x_1 = x, \tag{21}$$

$$f_{n+1}(\zeta) = f_n(\zeta) \frac{g(D_n|\zeta)}{\int g(D_n|\mu) f_n(\mu) d\mu} \;, \; f_1(\zeta) = f(\zeta). \tag{22}$$

IV. DYNAMIC PROGRAMMING

A. Bounds on the Value Function

The pair $(x_n, f_n(\cdot))$ is the state of our dynamic system, and the state space is $\mathbb{R} \times \mathscr{L}(\mathbb{R})$, where $\mathscr{L}(\mathbb{R})$ is the set of probability densities on $\mathbb{R}$. Coming back to our problem, calling $V = \{y_{1,\dots}, y_n, \cdots\}$ with $y_1 = y \geq x$, $y_n \geq x_n$, we define the payoff as

$$J_{x,f(\cdot)}(V) = \sum_{n=1}^{+\infty} \alpha^{n-1} E[h x_n^+ + p x_n^- - c x_n + c y_n + K \mathbf{1}_{y_n > x_n}], \tag{23}$$

and we call the value function

$$\Phi(x, f(\cdot)) = \inf_V J_{x,f(\cdot)}(V). \tag{24}$$

Using dynamic programming, we write the functional Bellman equation for the value function as

$$\begin{aligned} &\Phi(x, f(\cdot)) = hx^+ + px^- - cx + \\ &\inf_{y \geq x} \left[cy + K\mathbf{1}_{y>x} + \alpha E \Phi\left(y - D, \frac{f(\cdot) g(D|.)}{\int f(\mu) g(D|\mu) d\mu} \right) \right]. \end{aligned} \tag{25}$$

As Bellman's equation may have many solutions, we obtain conditions that guarantee that the value function (24) becomes its unique solution:

$$\begin{aligned} &\left(\frac{hx^+}{1-\alpha} + px^- - \frac{h\alpha}{(1-\alpha)^2} \int \varphi(\mu) f(\mu) d\mu \right)^+ \\ &\qquad \leq \Phi(x, f(\cdot)) \leq \\ &\frac{hx^+}{1-\alpha} + (p+c)x^- + \frac{\alpha(p+c)}{1-\alpha} \int \varphi(\mu) f(\mu) d\mu + \frac{K}{1-\alpha}, \end{aligned} \tag{26}$$

where

$$\varphi(\mu) = \int_0^{+\infty} \bar{G}(\zeta|\mu)d\zeta. \quad (27)$$

Of course, the right-hand side is a real constraint if we assume

$$\int \varphi(\mu)f(\mu)d\mu < +\infty. \quad (28)$$

B. Bounds on the Decision and the Modified Bellman Equation

The functional $\Phi(x, f(\cdot))$ is a solution of the Bellman equation (25). For a solution that will be the value function (24), we must satisfy the inequalities (26). If a solution of (25) satisfies the upper bound in (26), we obtain the following bounds on the set of values of y:

$$x \le y \le x^+ + \frac{K}{c}\frac{\alpha}{(1-\alpha)} + \frac{\alpha}{c(1-\alpha)^2}(h+(p+c)(1-\alpha))\int \varphi(\mu)f(\mu)d\mu. \quad (29)$$

We call this interval the admissibility interval for y and denote it as $\mathrm{I}_{x,f(\cdot)}$. So we can expect the value function to be a solution of the following equation

$$\Phi(x, f(\cdot)) = hx^+ + px^- - cx + \inf_{y \in \mathrm{I}_{x,f(\cdot)}} \Bigg[cy + K\mathbf{1}_{y>x} + \alpha \int_0^{+\infty} \Phi\left(y-z, \frac{f(\cdot)g(z|.)}{\int f(\mu)g(z|\mu)d\mu}\right) \int f(\mu)g(z|\mu)d\mu dz \Bigg], \quad (30)$$

and we expect the solution to satisfy (26).

C. A Comparison Result

We state here an important comparison result between the value function $\inf_V J_{x,f(\cdot)}(V)$ and the solutions of (30) satisfying the inequalities (26). To avoid confusion, we temporarily write the value function as

$$\tilde{\Phi}(x, f(\cdot)) = \inf_V J_{x,f(\cdot)}(V).$$

Proposition 3 : Any solution $\Phi(x, f(\cdot))$ of (30) such that (26) holds, satisfies

$$\Phi(x, f(\cdot)) \le \tilde{\Phi}(x, f(\cdot)) = \inf_V J_{x,f(\cdot)}(V). \quad (31)$$

V. UNNORMALIZED PROBABILITY SETUP

A. Extension of the Functional $\Phi(x, f(\cdot))$

For computational convenience, we work with unnormalized probability to make the belief updating linear. We first consider the space of functions on $\mathbb{R}$, such that

$$||f||_\varphi = \int |f(\mu)|d\mu + \int \varphi(\mu)|f(\mu)|d\mu < +\infty, \quad (32)$$

and $||f||_\varphi$ is a norm. We denote this space $L^1_\varphi(\mathbb{R})$. It is a Banach space for this norm. The subset of positive functions is closed. We denote this subset $L^1_{\varphi+}(\mathbb{R})$. A functional on $\mathbb{R} \times f(\cdot)$ is extended to $\mathbb{R} \times L^1_{\varphi+}(\mathbb{R})$ by

$$W(x, f(\cdot)) = \Phi\left(x, \frac{f(\cdot)}{\int f(\mu)d\mu}\right) \int f(\mu)d\mu. \quad (33)$$

From (30), we get the Unnormalized Bellman Equation

$$W(x, f(\cdot)) = (hx^+ + px^- - cx)\int f(\mu)d\mu + \inf_{\mathscr{L}_{x,f(\cdot)}} \Bigg[(cy + K\mathbf{1}_{y>x}) \int f(\mu)d\mu + \alpha \int_0^{+\infty} W(y-z, f(\cdot)g(z|.))dz \Bigg], \quad (34)$$

where $\mathscr{L}_{x,f(\cdot)}$ is the new admissibility interval computed to be

$$\Bigg[x, x^+ + \frac{K}{c}\frac{\alpha}{(1-\alpha)} + \frac{1}{c}\frac{\alpha}{(1-\alpha)^2}(h+(p+c)(1-\alpha))\frac{\int \varphi(\mu)f(\mu)d\mu}{\int f(\mu)d\mu} \Bigg],$$

and from (26), we can bound $W(x, f(\cdot))$ as follows:

$$\left(\left(\frac{hx^+}{1-\alpha} + px^-\right)\int f(\mu)d\mu - \frac{h\alpha}{(1-\alpha)^2}\int \varphi(\mu)f(\mu)d\mu \right)^+ \le W(x, f(\cdot)) \le \left(\frac{hx^+}{1-\alpha} + (p+c)x^- + \frac{K}{1-\alpha} \right)\int f(\mu)d\mu + \frac{\alpha(p+c)}{1-\alpha}\int \varphi(\mu)f(\mu)d\mu. \quad (35)$$

B. An Equivalent Formulation

We discard the upper bound on the argument y in the infimum of the right side of equation (34) and define

$$Z(x, f(\cdot)) = W(x, f(\cdot)) - (hx + px^- - cx)\int f(\mu)d\mu. \quad (36)$$

Following simple calculations, we obtain the equation

$$Z(x, f(\cdot)) = \inf_{y \ge x} \left(\int (l(y,\mu) + K\mathbf{1}_{y>x})f(\mu)d\mu + \alpha \int_0^{+\infty} Z(y-z, f(\cdot)g(z|.))dz \right), \quad (37)$$

with

$$l(x,\mu) = (c(1-\alpha) + \alpha h)x^+ + (\alpha p - c(1-\alpha))x^- + \alpha(h+p)\int_{x^+}^{+\infty} \overline{G}(z|\mu)dz + \alpha(c-h)\int_0^{+\infty} \overline{G}(z|\mu)dz. \quad (38)$$

We next introduce the following two operators

$$T(Z)(x, f(\cdot)) = \int l(x,\mu)f(\mu)d\mu + \alpha \int_0^{+\infty} Z(x-z, f(\cdot)g(z|.))dz, \quad (39)$$

$$M(Z)(x, f(\cdot)) = K\int f(\mu)d\mu + \inf_{y>x} Z(y, f(\cdot)). \quad (40)$$

We can then rewrite (37) as

$$Z(x,f(\cdot)) = \min\left[T(Z)(x,f(\cdot)), M(T(Z))(x,f(\cdot))\right], \quad (41)$$

and show that it is equivalent to

$$Z(x,f(\cdot)) = \min\left[T(Z)(x,f(\cdot)), M(Z)(x.f(\cdot))\right]. \quad (42)$$

C. An Approximate Solution

The question we want to address is whether (42) has an (s,S) solution like in the case of fully known demand. The answer is no. We will discuss this assertion in the next section. So the natural question is what could be an (s,S) approximate solution? To emphasize the fact that it depends on $f(\cdot)$, we call it (s_f, S_f) solution. Given s, we define a function $Z_s(x,f(\cdot))$ such that

$$Z_s(x,f(\cdot)) = \begin{cases} T(Z_s)(x,f(\cdot)),\ x > s \\ M(Z_s)(x,f(\cdot)),\ x < s \end{cases} \quad (43)$$

Note that S denotes the smallest number greater than s such that

$$Z_s(S,f(\cdot)) = \inf_{y>s} Z_s(y,f(\cdot)). \quad (44)$$

The problem is to find (s_f,S_f) and a function $Z_{s_f}(x,f(\cdot))$ that is continuous in x at point s, such that (43) holds. To that end, we need to introduce some functions. Recalling $l(x,\mu)$ defined in (38), we have

$$\begin{aligned} l'(x,\mu) = (c(1-\alpha)+\alpha h - \alpha(h+p)\overline{G}(x,\mu))\mathbf{1}_{x>0} \\ -(\alpha p - c(1-\alpha))\mathbf{1}_{x<0}. \quad (45) \end{aligned}$$

We define $\overline{s}_f$ to be the solution of

$$\int l'(\overline{s}_f,\mu) f(\mu) d\mu = 0. \quad (46)$$

We introduce renewal-like functions

$$\Gamma(x,\mu) = 1 + \alpha \int_0^x \Gamma(x-\zeta,\mu) g(\zeta|\mu) d\zeta, x>0 \quad (47)$$

$$\Gamma^{'}(x,\mu) = \alpha g(x|\mu) + \alpha \int_0^x \Gamma^{'}(x-\zeta,\mu) g(\zeta|\mu) d\zeta, x>0 \quad (48)$$

with

$$1 \le \Gamma(x,\mu) \le \frac{1}{1-\alpha}, \Gamma^{'}(x,\mu) > 0. \quad (49)$$

Next, we define

$$\overline{L}(x,\mu) = \frac{1}{1-\alpha}(l')^+(x,\mu) - (l')^-(x,\mu), \quad (50)$$

$$\underline{L}(x,\mu) = (l')^+(x,\mu) - \frac{1}{1-\alpha}(l')^-(x,\mu), \quad (51)$$

which gives us the following relationship

$$\underline{L}(x,\mu) \le l'(x,\mu) \le \overline{L}(x,\mu). \quad (52)$$

Consider $\int \underline{L}(x,\mu) f(\mu) d\mu$ and $\int \overline{L}(x,\mu) f(\mu) d\mu$. There exist unique values s_f^* and s_f^{**}such that

$$\int \underline{L}(s_f^*,\mu) f(\mu) d\mu = 0,\ \int \overline{L}(s_f^{**},\mu) f(\mu) d\mu = 0. \quad (53)$$

From (52), we see that

$$0 < s_f^{**} \le \overline{s}_f \le s_f^*. \quad (54)$$

Theorem 4 : If we assume

$$K \int f(\mu) d\mu + \int_{\overline{s}_f}^{s_f^*} \int \underline{L}(\zeta,\mu) f(\mu) d\mu d\zeta > 0, \quad (55)$$

then there exists one and only one (s_f,S_f) such that (43) and (44) hold and $Z_{s_f}(x,f(\cdot))$ is continuous in x at point s_f.

D. The Convexity Gap

When complete information is available, an (s,S) policy is optimal (see Theorem 1). We discuss in this section why it does not carry over to our Bayesian control formulation. The key point is in Proposition 2, in the spirit of which, we introduce

$$B_s(x,f(\cdot)) = H_s(x,f(\cdot)) - \inf_{y>x} H_s(y,f(\cdot)). \quad (56)$$

We have

$$B_s(s,f(\cdot)) = -\inf_{y>s} H_s(y,f(\cdot)), \quad (57)$$

and for $s = s_f$, this quantity is equal to $K\int f(\mu)d\mu$. Therefore, if we could extend Proposition 2 and could assert, for $x > s$

$$B_s(x,f(\cdot)) \le B_s(s,f(\cdot)), \forall s, \quad (58)$$

then we would obtain

$$Z_{s_f}(x,f(\cdot)) \le M(Z_{s_f})(x,f(\cdot)), \forall x > s_f. \quad (59)$$

The inequality (58) holds when f is a Dirac function is because of the convexity property, which carries over from $l(x)$ to $H_s(x)$ for large x. However, when $f(\cdot)$ is a general distribution, this convexity is lost. So this line of proof cannot be extended to prove the optimality of (s,S) policy. Nevertheless we can somewhat extend Proposition 2 by finding a bound on $B_s(x,f(\cdot))$ which is higher than $B_s(s,f(\cdot))$. To that extent, we introduce the function

$$B_s(x,\mu) = H_s(x,\mu) - \inf_{y>x} H_s(y,\mu). \quad (60)$$

We can now apply Proposition 2 and assert that

$$B_s(x,\mu) < B_s(s,\mu),\ \forall x > s, \quad (61)$$

and we note that $f(\cdot) \to B_s(x,f(\cdot))$ is not linear, but we have

$$B_s(x,f(\cdot)) \le \int B_s(x,\mu) f(\mu) d\mu, \quad (62)$$

and thus

$$B_s(x,f(\cdot)) \le \int B_s(s,\mu) f(\mu) d\mu, \quad (63)$$

which is a higher bound, since

$$\int B_s(s,\mu) f(\mu) d\mu > B_s(s,f(\cdot)). \quad (64)$$

Clearly,

$$\begin{aligned} &\int B_s(s,\mu) f(\mu) d\mu - B_s(s,f(\cdot)) \\ &= \inf_{y>s} \int H_s(y,\mu) f(\mu) d\mu - \int \inf_{y>s} H_s(y,\mu) f(\mu) d\mu. \quad (65) \end{aligned}$$

We call this quantity a *convexity gap*.

VI. Ongoing and Future Work

Historically, problems of this nature are typically addressed using K-convexity arguments [3][4][5]. While we can show the optimality of an (s,S) policy when demand is known without invoking K-convexity, our approach hits a roadblock - the convexity gap in the case of an unknown demand distribution. We are now working on using the K-convexity theory to see if the optimality of an (s,S) policy extends to the case with demand learning. It is our expectation that this is the case.

References

[1] Bensoussan A, Sethi SP, Thiam A, and Turi J (2025) Optimality of Base-Stock Policy Under Unknown General Demand Distributions: New Methods, New Results, and Computations. Production and Operations Management xx(x): 1–22.

[2] A. Bensoussan, *Dynamic Programming and Inventory Control,* IOS Press, Amsterdam, (2011).

[3] H. Scarf, The Optimality of (s,S) Policies in the Dynamic Inventory Problem , *Mathematical Methods in the Social Sciences,* K. Arrow, S. Karlin, P. Suppes Eds, chapter 13, pp 196-202, Stanford University Press, Stanford, (1960).

[4] Iglehart, D. L. (1963). Optimality of (s,S) Policies in the Infinite Horizon Dynamic Inventory Problem. Management Science, 9(2), 259–267.

[5] Perera, S. C., and Sethi, S. P. (2023). A survey of stochastic inventory models with fixed costs: Optimality of (s,S) and (s,S)-type policies—Discrete-time case. Production and Operations Management, 32, 131–153.