

Markov Decision Processes with Censored Observations: Analysis and Applications in Dynamic Pricing

Li Chen

The Fuqua School of Business, Duke University, Durham, NC 27708

li.chen@duke.edu

Censored (or truncated) observations occur commonly in practice. For example, the transaction price of a product may be a censored observation of the customer's true valuation, and inventory stockouts may cause retail point-of-sales data to contain censored observations of the true customer demand. How do imperfect observations of this sort affect optimal decisions? We consider this problem in the setting of a finite-horizon Markov decision process with two-sided censoring. We first show a general expression for the problem that quantifies the loss of informational value due to censoring. Based on this result, we derive a sufficient condition for comparing the optimal decision under censored observations to the optimal decision under exact observations. We then provide applications in dynamic pricing that involves two-sided censoring. In these applications, we show that the comparison result is state-dependent. This insight generalizes the previous findings obtained in one-sided censoring problems such as the Bayesian inventory control problem with unobserved lost sales.

Key words: Markov decision processes, incomplete information, censored observations, two-sided censoring, Bayesian dynamic programming, dynamic pricing.

History: Revised in September, 2011.

1. Introduction

Censored (or truncated) observations occur commonly in practice. Tobin (1958) first noted that zero points in household durable goods expenditure data are censored observations of the actual willingness-to-consume of low-income households. Similar censoring effects have been reported in many other situations, such as in unemployment, welfare receipt, and inheritance data, to name a few (see Amemiya 1984). The same effect is also present in retail point-of-sales data: when inventory runs out, the actual total demand is not observed—all we know is that it is greater than or equal to the sales observed (Nahmias 1994; Lariviere and Porteus 1999). Given the prevalence of censored data, it is important for managers to understand how the existence of imperfect observations affects optimal decisions. Mathematically, it presents an interesting but nontrivial research problem: Can we obtain general structural results given the censored observations?

In this paper, we seek to address this challenging problem by analyzing a general finite-horizon Markov decision process (MDP) with two-sided censoring. In particular, we derive a general expression for the problem that quantifies the loss of informational value due to censoring. Based on this result, we derive a sufficient condition for comparing the optimal decision under censored observations to the optimal decision under exact observations. We then apply this result to analyze problems in dynamic pricing with demand learning that involves two-sided censoring. To the best of our knowledge, our paper is the first in the literature that attempts to analyze and provide application examples for MDP problems involving two-sided censoring. In these dynamic pricing applications, we show that the comparison result is state-dependent. This insight generalizes the previous findings obtained in one-sided censoring problems such as the Bayesian inventory control problem with unobserved lost sales (e.g., Harpaz et al. 1982, Lariviere and Porteus 1999, Ding et al. 2002, Lu et al. 2005, 2008, Bensoussan et al. 2005, 2007, 2008, 2009, Bisi and Dada 2007, Chen and Plambeck 2008, Chen 2010, and Bisi et al. 2011).

To give a two-sided censoring example, let us consider a finite-horizon dynamic pricing problem for a single product. The goal is to maximize the expected total revenue during the selling horizon. Assume that only one customer arrives during a period. The arriving customer type follows an unknown distribution. After the customer makes the purchase decision, the customer type can be inferred from the purchase quantity according to the (known) Engel curve (e.g., Figure 1 in Section 4). Here we allow for multiple-unit purchases from the customer, which was usually assumed away in the dynamic pricing literature—where, instead, each customer’s demand is simply assumed to be binary (e.g., Aviv and Pazgal 2005ab, Araman and Caldentey 2009, and Farias and Van Roy 2010).

Because of the relaxation of the binary demand assumption, the data-censoring effect emerges: a zero purchase quantity from the customer is now a “left-censored” observation of the actual customer type. Furthermore, when there is a limited amount of inventory, if a customer’s preferred purchase quantity exceeds this amount, the excess demand becomes lost sales. This lost sales effect leads to a “right-censored” observation of the actual customer type—all we know is that the customer type is greater than or equal to a threshold value determined by the available inventory amount (see Figure 1 in Section 4 for a graphical illustration).

Let us now consider two scenarios in this dynamic pricing problem. In the first scenario, we assume the product is regularly replenished and the system inventory is restored to a predetermined capacity level at the beginning of each period. This capacity level can be viewed as, for example, a fixed shelf size or display capacity. Intuitively, there are two opposing forces for demand learning in this case: the left censoring due to customers not purchasing motivates one to lower the price, but the right censoring due to the finite system capacity induces one to raise the price. Analytically, we show that there exist two state-dependent thresholds. If the optimal price for the exact-observation problem is less than the lower threshold, then the optimal price for the two-sided censoring problem is greater than that of the exact-observation problem. On the other hand, if the optimal price for the exact-observation problem is greater than the upper threshold, then the optimal price for the two-sided censoring problem is less than that of the exact-observation problem. The intuition behind this result is the following: when price is low, the left censoring effect becomes less severe and thus one should increase price to mitigate the right-censoring effect; on the other hand, when price is already high, the right censoring becomes less of an issue and thus one should lower price to alleviate the left-censoring effect instead. In other words, when there is two-sided censoring, one should avoid pricing either too low or too high, so as to increase chances of observing the exact customer type information *in the middle*. This insight generalizes those obtained from the one-sided censoring problems in the literature. A numerical illustration based on the gamma-exponential conjugate prior distribution is given in Section 4.1.

In the second scenario, we assume that there is no inventory replenishment during the selling horizon. This scenario applies to certain products with a short selling season and limited replenishment opportunities, such as high-tech products and fashion apparel. In this case, if the available inventory is sold out, the problem ends. Thus, the right-censoring problem becomes a non-issue here. However, there is an added complexity of inventory state interaction between periods in this case. Analytically, we show that there exists a state-dependent threshold. If the optimal price for the exact-observation problem is greater than the threshold, then the optimal price for the censoring

problem is less than that of the exact-observation problem. In particular, this threshold reduces to zero if there is an ample supply of inventory or if the decision horizon is two periods—meaning the optimal price for the censoring problem is always less than that of the exact-observation problem in these cases. In a sense, this state-dependent result generalizes the “price low to obtain information” insight obtained by Braden and Oren (1994), which was based on the ample inventory supply assumption.

The MDP problem we consider in this paper belongs to the family of models with incomplete information (see Dynkin and Yushkevich 1979, Chapter 8). Specifically, there is an unknown system parameter, but the decision maker can use observations (which can be censored) to update the probabilistic distribution of the parameter by Bayes’ rule. This problem is closely related to the classic partially-observed Markov decision process (POMDP). The classic POMDP formulation involves only one state variable, i.e., the probabilistic distribution of the unknown parameter (see Monahan 1982). Here we allow for additional observable state variables, such as a capacity limit, in the decision process (see also Lovejoy 1993).

Censored observations in the Bayesian updating process complicate the analysis significantly. To deal with them, we use the unnormalized prior to create an equivalent dynamic programming formulation. The unnormalized prior is simply the product of the likelihood functions and the initial prior distribution (without performing the normalization step). This technique was first introduced in a series of papers by Bensoussan et al. (2005, 2007, 2008, 2009) to linearize the state transition equations in various Bayesian inventory control problems with unobserved lost sales (a one-sided censoring process). In this paper, we apply the technique to simplify our analysis for a general finite-horizon Markov decision process with two-sided censoring.

For ease of exposition, throughout the paper we assume that random events are independent and identically-distributed. In Section 5, we comment that all the results derived in this paper can be carried over to a more general Markov-modulated random process (e.g., Iglehardt and Karlin 1962, Song and Zipkin 1993).

The rest of the paper is organized as follows. We introduce the Bayesian dynamic programming formulation for our problem in Section 2. We then present our analysis and main results in Section 3. In Section 4, we present two applications in dynamic pricing with demand learning. Section 5 contains a discussion and our concluding remarks. All proofs are presented in the Appendix.

2. Dynamic Programming Formulation

Let us consider a Markov decision process with incomplete information (Dynkin and Yushkevich 1979, Chapter 8). There are a total of T decision periods, with the initial period indexed by $t = 1$. Throughout the paper, we shall use the subscript “ t ” to denote the period index for a variable or function, but suppress it whenever appropriate to reduce notation.

The state of the system at period t is given by a pair (s_t, θ_t) , where s_t is a fully-observable state (such as system capacity or inventory level) and θ_t is an unknown and unobservable state (such as the parameter of the underlying demand or customer type distribution). We assume the state space is the Borel space on $[0, \infty) \times \Theta$, where Θ can be a finite or countable set or a (half) real line. A prior distribution of θ_1 , denoted by $\pi_1(\theta_1)$, is assumed to be available on Θ at the beginning of the first period. In what follows, we will assume that $\pi_1(\theta_1)$ is a continuous density on Θ (e.g., $\Theta = [0, \infty)$); the results for the finite and countable Θ case can be obtained by treating $\pi_1(\theta_1)$ as a probability mass.

The unobservable state θ_t determines the underlying probability distribution of the random event X_t in period t . Specifically, we assume X_t is a nonnegative real-valued random variable, with a continuous density $f(x|\theta_t)$ for $x \in [0, \infty)$. For ease of exposition, we assume θ_t is an (unknown) constant and henceforth drop the subscript t ; as a result, the random event X_t is independently identically-distributed (i.i.d.). The more general case where θ_t is time-varying will be discussed in section 5.

Let a be a nonnegative real-valued action taken in a period. We assume the action space in each period is the Borel space on $[0, \infty)$. Furthermore, there exists an observation process Z_t (of X_t) defined as follows.

$$Z_t = \begin{cases} l(a, s) & \text{if } 0 \leq X_t \leq l(a, s), \\ X_t & \text{if } l(a, s) < X_t < r(a, s), \\ r(a, s) & \text{if } r(a, s) \leq X_t < \infty, \end{cases}$$

Following Braden and Freimer (1991), we call $l(a, s)$ the left-censoring point and $r(a, s)$ the right-censoring point, where both points are dependent on the current-period action a and the observable state s , with $l(a, s) \leq r(a, s)$. We have a left-censored observation if $Z_t = l(a, s)$, an exact observation if $l(a, s) < Z_t < r(a, s)$, and a right-censored observation if $Z_t = r(a, s)$. We further assume that $l(a, s)$ and $r(a, s)$ are differentiable in a and s . Note that this two-sided censoring process can be easily reduced to a single-sided censoring process by setting either $r(a, s) \equiv \infty$ (for a left-censored only process) or $l(a, s) \equiv 0$ (for a right-censored only process).

Based on the above definition, the likelihood of an observation $Z_t = z$ can be written as

$$k(Z_t = z|\theta) = \begin{cases} F(l(a, s)|\theta) & \text{if } z = l(a, s), \\ f(z|\theta) & \text{if } l(a, s) < z < r(a, s), \\ \bar{F}(r(a, s)|\theta) & \text{if } z = r(a, s), \end{cases} \quad (1)$$

where $F(z|\theta) = \int_0^z f(x|\theta)dx$ and $\bar{F}(z|\theta) = 1 - F(z|\theta)$.

The state transitions between periods are defined as follows:

$$s_{t+1} = h(s_t, a_t, Z_t), \quad (2)$$

with $l(a_t, s_t) \leq Z_t \leq r(a_t, s_t)$, and $\theta_{t+1} = \theta_t$ for $t = 1, \dots, T - 1$ (because θ_t is assumed to be a constant). We assume $h(s_t, \cdot, \cdot)$ is differentiable in s_t . Given the observation $Z_t = z$, the updated prior of θ at the beginning of period $t + 1$ is obtained via Bayes' rule:

$$\pi_{t+1}(\theta) = \frac{k(Z_t = z|\theta) \cdot \pi_t(\theta)}{\int_{\Theta} k(Z_t = z|\theta) \cdot \pi_t(\theta)d\theta}, \quad (3)$$

where $\pi_t(\theta)$ is the prior at the beginning of period t . We shall use π_t and $\pi_t(\theta)$ interchangeably whenever appropriate within context.

The expected single-period reward in the decision process is defined as

$$R(a, s, \pi) = \int_0^\infty \int_{\Theta} u(a, s, x)f(x|\theta)\pi(\theta)d\theta dx,$$

where $u(a, s, x)$ is the nonnegative reward function given action a , state s , and random outcome x . We assume $R(a, s, \pi)$ is bounded above and differentiable in a and s . Furthermore, we assume that the problem is well-defined such that there exists a finite solution that maximizes $R(a, s, \pi)$ for any given state s and prior π . This assumption can be satisfied, for example, by requiring $\lim_{a \rightarrow \infty} R(a, s, \pi) = 0$ for any given state s and prior π . Finally, we assume the terminal reward at the period $T + 1$ is zero.

According to Dynkin and Yushkevich (1979, pp. 214-217), the above problem with incomplete information (in terms of θ) can be reduced to an equivalent model with complete information by replacing the state θ_t with its prior distribution $\pi_t(\theta)$ (given by (3)). Furthermore, based on our problem definition, it can be verified that the resulting complete-information model satisfies the conditions of the general Borel model (Dynkin and Yushkevich 1979, pp. 46-47, conditions (α) – (ε)). Thus, the problem can be formulated by the following dynamic programming optimality equations: For $t = 1, \dots, T$,

$$V_t(s, \pi)$$

$$\begin{aligned}
&= \max_{a \geq 0} \left\{ G_t(a, s, \pi) \right\} \\
&= \max_{a \geq 0} \left\{ R(a, s, \pi) + \int_{l(a, s)}^{r(a, s)} V_{t+1} \left(h(s, a, z), \frac{f(z|\cdot)\pi}{\int_{\Theta} f(z|\theta)\pi(\theta)d\theta} \right) \cdot \left(\int_{\Theta} f(z|\theta)\pi(\theta)d\theta \right) dz \right. \\
&\quad + V_{t+1} \left(h(s, a, l(a, s)), \frac{F(l(a, s)|\cdot)\pi}{\int_{\Theta} F(l(a, s)|\theta)\pi(\theta)d\theta} \right) \cdot \int_{\Theta} F(l(a, s)|\theta)\pi(\theta)d\theta \\
&\quad \left. + V_{t+1} \left(h(s, a, r(a, s)), \frac{\bar{F}(r(a, s)|\cdot)\pi}{\int_{\Theta} \bar{F}(r(a, s)|\theta)\pi(\theta)d\theta} \right) \cdot \int_{\Theta} \bar{F}(r(a, s)|\theta)\pi(\theta)d\theta \right\}, \quad (4)
\end{aligned}$$

with $V_{T+1} = 0$. In the above expression, the first term is the current-period reward and the rest of the terms are the expected total future rewards when the current-period observation is exact, left-censored, or right-censored, respectively. Note that we assume the problem is well-defined such that finite (though not necessarily unique) optimal solution exists ; hence, the maximum is attainable. In cases where the optimal solution is not unique, let $a_t^*(s, \pi)$ denote the solution with the smallest value.

To facilitate the subsequent analysis, from (3), let us define an unnormalized prior $\tilde{\pi}_t$ as

$$\tilde{\pi}_t(\theta) = k(Z_{t-1} = z|\theta) \cdot \tilde{\pi}_{t-1}(\theta) \quad (5)$$

for $t = 2, \dots, T$, with $\tilde{\pi}_1(\theta) = \pi_1(\theta)$. It is straightforward to show that $\tilde{\pi}_t = \pi_t \cdot \int_{\Theta} \tilde{\pi}_t(\theta)d\theta$. Thus, $\tilde{\pi}_t(\theta)$ is a derived measure on Θ . Based on this measure, let us redefine the single-period reward as

$$R(a, s, \tilde{\pi}) = \int_0^{\infty} \int_{\Theta} u(a, s, x) f(x|\theta) \tilde{\pi}(\theta) d\theta dx, \quad (6)$$

and the optimality equations as: for $t = 1, \dots, T$,

$$\begin{aligned}
\tilde{V}_t(s, \tilde{\pi}) &= \max_{a \geq 0} \left\{ \tilde{G}_t(a, s, \tilde{\pi}) \right\} \\
&= \max_{a \geq 0} \left\{ R(a, s, \tilde{\pi}) + \int_{l(a, s)}^{r(a, s)} \tilde{V}_{t+1}(h(s, a, z), f(z|\cdot)\tilde{\pi}) dz \right. \\
&\quad \left. + \tilde{V}_{t+1}(h(s, a, l(a, s)), F(l(a, s)|\cdot)\tilde{\pi}) + \tilde{V}_{t+1}(h(s, a, r(a, s)), \bar{F}(r(a, s)|\cdot)\tilde{\pi}) \right\}, \quad (7)
\end{aligned}$$

with $\tilde{V}_{T+1} = 0$. The following lemma establishes the equivalence between (7) and the original program (4):

Lemma 1. For $t = 1, \dots, T$, $\tilde{G}_t(a, s, \tilde{\pi}) = G_t(a, s, \pi) \cdot \int_{\Theta} \tilde{\pi}(\theta)d\theta$ and $\tilde{V}_t(s, \tilde{\pi}) = V_t(s, \pi) \cdot \int_{\Theta} \tilde{\pi}(\theta)d\theta$.

This lemma shows that dynamic program (7), defined on the derived measure $\tilde{\pi}$, is a scaled version of the original program (4). Thus, both programs share the same optimal solution, and

we only need to focus on (7) for our subsequent analysis. We note that this unnormalized-prior technique was first introduced in a series of papers by Bensoussan et al. (2005, 2007, 2008, 2009) to solve various Bayesian inventory control problems that involves right-censored observations. Here, by Lemma 1, we formally establish the equivalence between the original dynamic program and its counterpart under the unnormalized prior for a general two-sided censoring process.

In what follows, we shall focus only on the problem defined by (7), and henceforth drop the tilde above $\tilde{\pi}$ and all functions of (7) unless otherwise noted.

3. Analysis and Main Results

Let us first define the following function:

$$\begin{aligned} \Delta_t(s, \pi', \pi) &= V_t(s, \pi) - G_t(\hat{a}, s, \pi) + \int_{l(\hat{a}, s)}^{r(\hat{a}, s)} \Delta_{t+1}(h(s, \hat{a}, z), f(z|\cdot)\pi', f(z|\cdot)\pi) dz \\ &\quad + \Delta_{t+1}(h(s, \hat{a}, l(\hat{a}, s)), F(l(\hat{a}, s)|\cdot)\pi', F(l(\hat{a}, s)|\cdot)\pi) \\ &\quad + \Delta_{t+1}(h(s, \hat{a}, r(\hat{a}, s)), \bar{F}(r(\hat{a}, s)|\cdot)\pi', \bar{F}(r(\hat{a}, s)|\cdot)\pi), \end{aligned} \quad (8)$$

with $\Delta_{T+1} = 0$ and $\hat{a} = a_t^*(s, \pi')$.

The function $\Delta_t(s, \pi', \pi)$ defined above can be interpreted as the difference between the optimal value function $V_t(s, \pi)$ and the expected total rewards under a policy based on an altered initial prior π' (note the system is evolving according to the prior π). Thus, it is straightforward to obtain the following lemma:

Lemma 2. *For any given s, π' , and π , $\Delta_t(s, \pi', \pi) \geq 0$ and the equality holds when $\pi' = \pi$.*

By applying a generalized envelope theorem of Milgrom and Segal (2002, Theorem 2) to our dynamic programming problem, with a backward induction, we can obtain the following result:

Theorem 1. *The value function $V_t(s, F(z|\cdot)\pi_{t-1})$ is absolutely continuous in z (implying differentiable in z almost everywhere) and is given by*

$$V_t(s, F(z|\cdot)\pi_{t-1}) = \int_0^z V_t(s, f(\zeta|\cdot)\pi_{t-1}) d\zeta - \int_0^z \Delta_t(s, F(\zeta|\cdot)\pi_{t-1}, f(\zeta|\cdot)\pi_{t-1}) d\zeta,$$

where Δ_t is defined by (8). Symmetrically, $V_t(s, \bar{F}(z|\cdot)\pi_{t-1})$ is absolutely continuous in z and is given by

$$V_t(s, \bar{F}(z|\cdot)\pi_{t-1}) = \int_z^\infty V_t(s, f(\zeta|\cdot)\pi_{t-1}) d\zeta - \int_z^\infty \Delta_t(s, \bar{F}(\zeta|\cdot)\pi_{t-1}, f(\zeta|\cdot)\pi_{t-1}) d\zeta.$$

The term $\int_0^z \Delta_t(s, F(\zeta|\cdot)\pi_{t-1}, f(\zeta|\cdot)\pi_{t-1}) d\zeta$ in the above result can be viewed as the loss of informational value when z is a left-censored observation. In other words, it quantifies the informational value of having an exact observation as compared to having a left-censored observation. Similarly, the term $\int_z^\infty \Delta_t(s, \bar{F}(\zeta|\cdot)\pi_{t-1}, f(\zeta|\cdot)\pi_{t-1}) d\zeta$ quantifies the loss of informational value when a right-censored observation occurs. In particular, if z is sent to ∞ in the first equation of the above theorem, we have

$$V_t(s, \pi_{t-1}) = \int_0^\infty V_t(s, f(\zeta|\cdot)\pi_{t-1}) d\zeta - \int_0^\infty \Delta_t(s, F(\zeta|\cdot)\pi_{t-1}, f(\zeta|\cdot)\pi_{t-1}) d\zeta.$$

Thus, the term $\int_0^\infty \Delta_t(s, F(\zeta|\cdot)\pi_{t-1}, f(\zeta|\cdot)\pi_{t-1}) d\zeta$ captures the informational gain between Bayesian updating with an exact observation and no information updating at all.

Substituting the result of Theorem 1 into the original problem, we obtain

$$\begin{aligned} G_t(a, s, \pi) &= R(a, s, \pi) + \int_0^\infty V_{t+1}(h(s, a, Z_t(x)), f(x|\cdot)\pi) dx \\ &\quad - \int_0^{l(a,s)} \Delta_{t+1}(h(s, a, l(a, s)), F(z|\cdot)\pi, f(z|\cdot)\pi) dz \\ &\quad - \int_{r(a,s)}^\infty \Delta_{t+1}(h(s, a, r(a, s)), \bar{F}(z|\cdot)\pi, f(z|\cdot)\pi) dz, \end{aligned}$$

where Δ_{t+1} is defined by (8). The above expression can be interpreted as follows. The second term on the right-hand side represents the expected future rewards if the current-period observation is exact. The last two terms represent the expected loss of informational value due to a left-censored observation and a right-censored observation, respectively.

Now let us consider a decision process in which one can fully observe X_t . We add the superscript “e” to the dynamic programming value functions to denote this exact observation case. The optimality equations for this problem are given below: for $t = 1, \dots, T$,

$$\begin{aligned} V_t^e(s, \pi) &= \max_{a \geq 0} \{G_t^e(a, s, \pi)\} \\ &= \max_{a \geq 0} \left\{ R(a, s, \pi) + \int_0^\infty V_{t+1}^e(h(s, a, Z_t(x)), f(x|\cdot)\pi) dx \right\}, \end{aligned} \quad (9)$$

with $V_{T+1}^e = 0$. By comparing the objective function $G_t^e(a, s, \pi)$ with the original censoring problem $G_t(a, s, \pi)$, we arrive at the following expression:

$$\begin{aligned} G_t^\Delta(a, s, \pi) &= G_t(a, s, \pi) - G_t^e(a, s, \pi) \\ &= \int_0^\infty [V_{t+1}(h(s, a, Z_t(x)), f(x|\cdot)\pi) - V_{t+1}^e(h(s, a, Z_t(x)), f(x|\cdot)\pi)] dx \\ &\quad - \int_0^{l(a,s)} \Delta_{t+1}(h(s, a, l(a, s)), F(z|\cdot)\pi, f(z|\cdot)\pi) dz \\ &\quad - \int_{r(a,s)}^\infty \Delta_{t+1}(h(s, a, r(a, s)), \bar{F}(z|\cdot)\pi, f(z|\cdot)\pi) dz. \end{aligned}$$

Proposition 1. *Suppose that the exact-observation problem $G_t^e(a, s, \pi)$ has a unique optimal solution. If the function $G_t^\Delta(a, s, \pi)$ is increasing (decreasing) in a , then the optimal solution to the censoring problem $G_t(a, s, \pi)$ is greater (less) than the optimal solution to the exact-observation problem.*

The above proposition provides a sufficient condition for comparing optimal decisions under censored observations to those under exact observations. It turns out that for a class of problems this condition can be satisfied. For example, assume that $s_{t+1} = h(s, a, Z_t) = (a - Z_t)^+$, $l(a, s) = 0$ and $r(a, s) = a$ (i.e., a right-censored only process). Thus, if $V_{t+1}((a - Z_t)^+, \cdot) - V_{t+1}^e((a - Z_t)^+, \cdot)$ is increasing in a , then $G_t^\Delta(a, s, \pi)$ is increasing in a . This can be verified to be true in the Bayesian inventory control problem with nonperishable inventory and unobserved lost sales (Chen and Plambeck 2008), and thus their “stock more” result can be obtained by a simple application of the above proposition.

Moreover, if the state transition function $s_{t+1} = h(s, a, Z_t)$ is independent of a , then we have $dG_t^e(a, s, \pi)/da = dR(a, s, \pi)/da$. Thus, the optimal solution to the exact observation case is the same as the myopic optimal solution. For a special case in which $s_{t+1} \equiv 0$, if we further assume that $l(a, s) = 0$ and $r(a, s) = a$ (i.e., a right-censored only process), then it is straightforward to show that $G_t^\Delta(a, s, \pi)$ is increasing in a and we can conclude that the optimal decision for this right-censored only process is greater than the myopic optimal decision. This is precisely the insight of “stock more than myopic inventory level” in the Bayesian inventory control problem with perishable inventory and unobserved lost sales (e.g., Harpaz et al. 1982, Lariviere and Porteus 1999, Ding et al. 2002; Lu et al. 2005, 2008; Bensoussan et al. 2007, 2009).

However, in more complex problems with two-sided censoring, the monotonicity of $G_t^\Delta(a, s, \pi)$ cannot always be guaranteed. As a result, the comparison result between optimal decisions under censored observations and those under exact observations becomes state-dependent. To further illustrate this point, below we present two applications in dynamic pricing with demand learning.

4. Applications in Dynamic Pricing with Demand Learning

Consider a finite-horizon dynamic pricing problem for a single product. At the beginning of a period, the seller determines the unit price a for the product. A customer arrives during the period and makes a purchase decision (we assume that only one customer arrives in a period). The type of customer arriving in each period is randomly drawn from an i.i.d. distribution $f(x|\theta)$ in each period, with x ($x \geq 0$) representing the customer type and θ representing the unknown parameter.

The seller has a prior belief concerning the value of the parameter θ at the beginning of each period, denoted by $\pi(\theta)$. The goal is to maximize expected total revenue over the selling horizon.

Specifically, we assume that the demand curve of each customer type is distinct so that there is a one-to-one correspondence between the customer type and the customer's preferred (positive) purchase-quantity decision. Because the mapping is one-to-one, we can infer the exact customer type from the quantity purchased and use this information to update the prior of the unknown parameter θ .

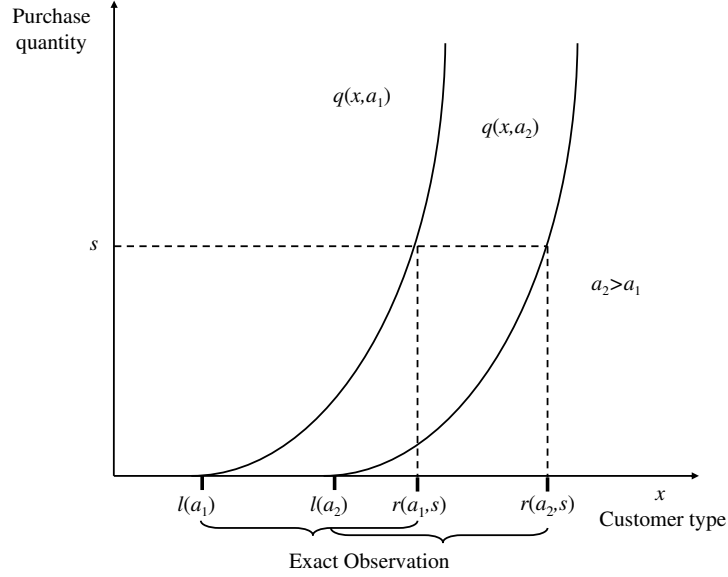


Figure 1: Customer Type and Purchase Quantity at Two Price Points a_1 and a_2

An example of the purchase-quantity curves at different prices is shown in Figure 1. Let $q(x, a)$ denote the purchase quantity, a function of the customer type x and the price a . Let us further assume $\partial q(x, a)/\partial x > 0$ (i.e., a high customer type buys more) and $\partial q(x, a)/\partial a < 0$ (i.e., a high price reduces one's purchase quantity). Note that this only applies to a positive purchase quantity. There are customers who choose not to purchase because their type is below a threshold $l(a)$ (these customers are effectively "priced out" by the price point a). Therefore, when a customer does not make a purchase, all we know is that the customer's type is less than or equal to $l(a)$; in other words, we have a left-censored observation of the customer type at $l(a)$. Furthermore, when a customer's preferred purchase quantity exceeds the available inventory amount, the excess demand becomes lost sales. Thus, if the available inventory for sale at the beginning of a period is s and the price is a , let us define $r(a, s) = q^{-1}(s; a)$, i.e., the customer type who would choose to purchase s units of

the product. Then $r(a, s)$ is a right-censored observation of the actual customer type—given the sales of s units, all we know is that the customer type is greater than or equal to $r(a, s)$.

For ease of exposition, for the rest of this paper, we shall assume that the purchase quantity curve $q(x, a)$ follows a simple linear form of $q(x, a) = \max\{x - a, 0\}$; the qualitative insights derived under this simplification nevertheless remain true for a general purchase quantity curve $q(x, a)$ as described above. Under the linear curve, the left-censoring point is given by $l(a) = a$ and the right-censoring point is given by $r(a, s) = a + s$.

Given price a , inventory capacity s , and prior $\pi(\theta)$, the expected single-period revenue can be written as

$$R(a, s, \pi) = a \int_a^{a+s} \int_{\Theta} (x - a) f(x|\theta) \pi(\theta) d\theta dx + a \int_{a+s}^{\infty} \int_{\Theta} s f(x|\theta) \pi(\theta) d\theta dx$$

Clearly, $R(0, s, \pi) = 0$. We require $R(a, s, \pi)$ to be unimodal to ensure the revenue-maximizing price is finite and unique (we note that this is a common assumption used in the revenue management literature; see Ziya et al. 2004 for a discussion).

4.1 The Regular Replenishment Case

Let us consider the case in which the product inventory is regularly replenished. In particular, we assume that the system inventory is restored to a predetermined capacity level S at the beginning of each period. This capacity level S can be viewed as, for example, a fixed shelf size or display capacity.

Because the starting inventory at each period is always S , we can suppress the state variable of available inventory in this case. The optimality equations for this revenue-maximizing problem are given as follows: for $t = 1, \dots, T$,

$$\begin{aligned} V_t(\pi) &= \max_{a \geq 0} \{G_t(a, \pi)\} \\ &= \max_{a \geq 0} \left\{ R(a, S, \pi) + \int_a^{a+S} V_{t+1}(f(x|\cdot)\pi) dx + V_{t+1}(F(a|\cdot)\pi) + V_{t+1}(\bar{F}(a+S|\cdot)\pi) \right\}, \end{aligned}$$

with $V_{T+1} = 0$. Thus, by Theorem 1, we immediately have

$$\begin{aligned} G_t(a, \pi) &= R(a, S, \pi) + \int_0^{\infty} V_{t+1}(f(x|\cdot)\pi) dx - \int_0^a \Delta_{t+1}(F(z|\cdot)\pi, f(z|\cdot)\pi) dz \\ &\quad - \int_{a+S}^{\infty} \Delta_{t+1}(\bar{F}(z|\cdot)\pi, f(z|\cdot)\pi) dz. \end{aligned}$$

Taking the difference between $G_t(a, \pi)$ and $G_t^e(a, \pi)$, we obtain the following:

$$G_t^{\Delta}(a, \pi) = G_t(a, \pi) - G_t^e(a, \pi)$$

$$\begin{aligned}
&= \int_0^\infty [V_{t+1}(f(x|\cdot)\pi) - V_{t+1}^e(f(x|\cdot)\pi)] dx - \int_0^a \Delta_{t+1}(F(z|\cdot)\pi, f(z|\cdot)\pi) dz \\
&\quad - \int_{a+S}^\infty \Delta_{t+1}(\bar{F}(z|\cdot)\pi, f(z|\cdot)\pi) dz.
\end{aligned}$$

Note that the first term on the right-hand side is independent of a . Thus, we have

$$\frac{d}{da} G_t^\Delta(a, \pi) = \Delta_{t+1}(\bar{F}(a+S|\cdot)\pi, f(a+S|\cdot)\pi) - \Delta_{t+1}(F(a|\cdot)\pi, f(a|\cdot)\pi). \quad (10)$$

Using Lemma 2, we can show that $\lim_{a \rightarrow 0} dG_t^\Delta(a, \pi)/da \geq 0$ and $\lim_{a \rightarrow \infty} dG_t^\Delta(a, \pi)/da \leq 0$ (the proof is given in the Appendix). Hence, depending the problem parameters, the optimal decisions for the two-sided censoring problem can be either greater or less than those for the exact-observation problem.

Proposition 2. *In the regular replenishment case, given the same prior π , there exist two state-dependent thresholds: $0 \leq \underline{a}(\pi) \leq \bar{a}(\pi)$. If the optimal price for the exact-observation problem (which is the same as the myopic optimal price) is less than $\underline{a}(\pi)$, then the optimal price for the two-sided censoring problem is greater than that of the exact-observation problem. On the other hand, if the optimal price for the exact-observation problem is greater than $\bar{a}(\pi)$, then the optimal price for the two-sided censoring problem is less than that of the exact-observation problem.*

The intuition behind the above result is the following: when price is low, the left censoring effect becomes less severe and thus one should increase price to mitigate the right-censoring effect; on the other hand, when price is already high, the right censoring becomes less of an issue and thus one should lower price to alleviate the left-censoring effect instead. In other words, when there is two-sided censoring, one should avoid pricing either too low or too high, so as to increase chances of observing exact customer type information in the middle. This insight generalizes those findings obtained in the one-sided censoring problems as discussed in the previous section. To further illustrate this insight, below we present a numerical example based on the gamma-exponential conjugate prior distribution.

The Gamma-Exponential Example

Let us consider a two-period problem. Specifically, let us assume that the customer type follows an exponential distribution with an unknown parameter θ , i.e., $f(x|\theta) = \theta e^{-\theta x}$. The initial prior of θ follows a unnormalized gamma distribution given by $\pi_1(\theta) = \theta^{\alpha-1} e^{-\beta\theta}$ (where α is the shape parameter and β the scale parameter). With some calculus, it is straightforward to show that

$$R(a, S, \pi_1) = \frac{\Gamma(\alpha-1)a}{(a+\beta)^{\alpha-1}} - \frac{\Gamma(\alpha-1)a}{(a+\beta+S)^{\alpha-1}},$$

$$\begin{aligned}
R(a, S, f(z|\cdot)\pi_1) &= \frac{\Gamma(\alpha)a}{(a + \beta + z)^\alpha} - \frac{\Gamma(\alpha)a}{(a + \beta + z + S)^\alpha}, \\
R(a, S, F(z|\cdot)\pi_1) &= \frac{\Gamma(\alpha - 1)a}{(a + \beta)^{\alpha-1}} - \frac{\Gamma(\alpha - 1)a}{(a + \beta + S)^{\alpha-1}} - \frac{\Gamma(\alpha - 1)a}{(a + \beta + z)^{\alpha-1}} + \frac{\Gamma(\alpha - 1)a}{(a + \beta + z + S)^{\alpha-1}}, \\
R(a, S, \bar{F}(z|\cdot)\pi_1) &= \frac{\Gamma(\alpha - 1)a}{(a + \beta + z)^{\alpha-1}} - \frac{\Gamma(\alpha - 1)a}{(a + \beta + z + S)^{\alpha-1}}.
\end{aligned}$$

Therefore, by equation (10) and the definition of (8), we obtain

$$\begin{aligned}
\frac{d}{da}G_1^\Delta(a, \pi_1) &= \Delta_2(\bar{F}(a + S|\cdot)\pi_1, f(a + S|\cdot)\pi_1) - \Delta_2(F(a|\cdot)\pi_1, f(a|\cdot)\pi_1) \\
&= \max_{a \geq 0}\{R(a, S, f(a + S|\cdot)\pi_1)\} - R(\hat{a}, S, f(a + S|\cdot)\pi_1) \\
&\quad - \max_{a \geq 0}\{R(a, S, f(a|\cdot)\pi_1)\} + R(\hat{a}', S, f(a|\cdot)\pi_1),
\end{aligned}$$

where $\hat{a} = \arg \max_{a \geq 0}\{R(a, S, \bar{F}(a + S|\cdot)\pi_1)\}$ and $\hat{a}' = \arg \max_{a \geq 0}\{R(a, S, F(a|\cdot)\pi_1)\}$. All these functions and quantities can be easily computed.

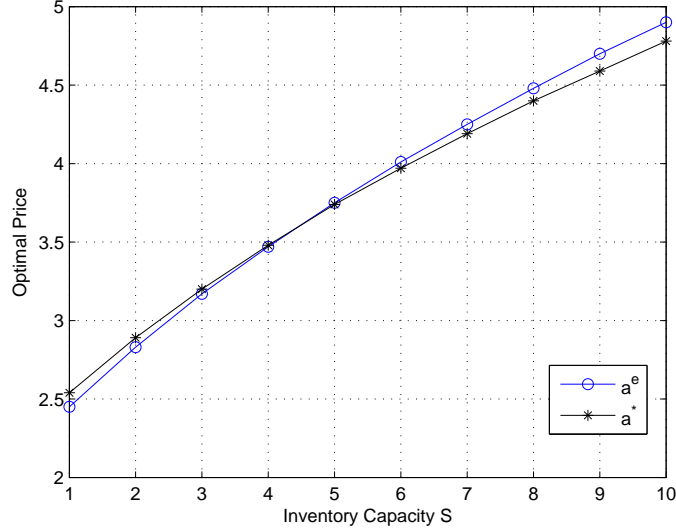


Figure 2: Optimal Prices a^* and a^e versus Inventory Capacity S (with $\alpha = 2$ and $\beta = 2$)

Figure 2 shows the optimal prices for the censoring problem (a^*) and the optimal prices for the exact-observation problem (a^e) by varying the inventory capacity S from 1 to 10 (with $\alpha = 2$ and $\beta = 2$). In particular, when $\alpha = 2$, the optimal solution to the exact-observation problem (which is the same as the optimal solution to $R(a, S, \pi_1)$) has a close-form expression given by $a^e = \sqrt{\beta(\beta + S)}$. As predicted in Proposition 2, the comparison result of a^* and a^e is indeed state-dependent: when a^e is small, we have $a^* > a^e$; and when a^e is large, we have $a^* < a^e$.

4.2 The No Replenishment Case

Now let us consider the case in which there is no inventory replenishment during the selling horizon. This scenario applies to certain products with a short selling season and limited replenishment opportunities, such as high-tech products and fashion apparel.

Since there is no replenishment, the inventory will keep depleting as more sales accrue. The optimality equations for this revenue-maximizing problem are given as follows: for $t = 1, \dots, T$,

$$\begin{aligned} V_t(s, \pi) &= \max_{a \geq 0} \{G_t(a, s, \pi)\} \\ &= \max_{a \geq 0} \left\{ R(a, s, \pi) + \int_a^{a+s} V_{t+1}(s+a-x, f(x|\cdot)\pi) dx + V_{t+1}(s, F(a|\cdot)\pi) \right\}, \end{aligned}$$

with $V_{T+1} = 0$ and $V_t(0, \cdot) = 0$ for all t . In this case, if the available inventory is sold out, the problem ends. Thus, the right-censoring has no effect on the value function here.

Applying Theorem 1, we can rewrite $G_t(a, s, \pi)$ as

$$G_t(a, s, \pi) = R(a, s, \pi) + \int_0^{a+s} V_{t+1}(s - (x-a)^+, f(x|\cdot)\pi) dx - \int_0^a \Delta_{t+1}(s, F(z|\cdot)\pi, f(z|\cdot)\pi) dz.$$

By taking the difference between $G_t(a, s, \pi)$ and $G_t^e(a, s, \pi)$, we obtain the following:

$$\begin{aligned} G_t^\Delta(a, s, \pi) &= G_t(a, s, \pi) - G_t^e(a, s, \pi) \\ &= \int_0^{a+s} [V_{t+1}(s - (x-a)^+, f(x|\cdot)\pi) - V_{t+1}^e(s - (x-a)^+, f(x|\cdot)\pi)] dx \\ &\quad - \int_0^a \Delta_{t+1}(s, F(z|\cdot)\pi, f(z|\cdot)\pi) dz. \end{aligned}$$

In a two-period problem, we have $V_2(s, \cdot) = V_2^e(s, \cdot) = \max_{a \geq 0} \{R(a, s, \cdot)\}$. Therefore, by Proposition 1, we conclude that the optimal price for the censoring problem is less than that of the exact-observation problem. Similarly, if there is an ample supply of supply (as a result, the state s can be suppressed), we reach the same conclusion. However, in more general cases, we have the following state-dependent result (the proof is given in the Appendix):

Proposition 3. *In the no replenishment case, there exists a state-dependent threshold $\bar{a}(s, \pi) \geq 0$. If the optimal price for the exact-observation problem is greater than $\bar{a}(s, \pi)$, then the optimal price for the censoring problem is less than that of the exact-observation problem. In particular, $\bar{a}(s, \pi) = 0$ if there is an ample supply of inventory or if the decision horizon is two periods.*

The above result suggests that when there is no replenishment, if the price is high, one should lower price so as to mitigate the left-censoring effect. This state-dependent result is derived based

on the finite inventory capacity assumption. In this sense, it generalizes the “price low to obtain information” insight obtained by Braden and Oren (1994), which was based on the ample inventory supply assumption.

5. Discussion and Concluding Remarks

In this paper, we have studied how censored observations affect optimal decisions in a general finite-horizon Markov decision process. So far we have derived all our results assuming that the random event X_t is i.i.d. In many real situations, however, random events, such as demands, are often correlated across different periods. This temporal correlation can be modeled by a Markov-modulated process (e.g., Iglehardt and Karlin 1962, Song and Zipkin 1993). Below we note that all our results can be carried over to the Markov-modulated process.

Let us assume that the unknown parameter θ evolves between periods according to a *known* stochastic transition function $\tau(\theta|\theta')$ (given the value of θ' , $\tau(\cdot|\theta')$ is a probability distribution). Hence, before observing Z_{t-1} , $\pi_{t-1}(\theta)$ is first transitioned to a new distribution by

$$(\tau \circ \pi_{t-1})(\theta) = \int_{\Theta} \tau(\theta|\theta')\pi_{t-1}(\theta')d\theta'.$$

The dependency of the random events across different periods is thus captured by the evolution of the Markov-modulating state θ . After observing $Z_{t-1} = z$, the posterior π_t is given by Bayes' rule as

$$\pi_t(\theta) = \frac{k(Z_{t-1} = z|\theta) \cdot (\tau \circ \pi_{t-1})(\theta)}{\int_{\Theta} k(Z_{t-1} = z|\theta) \cdot (\tau \circ \pi_{t-1})(\theta)d\theta},$$

where the likelihood function $k(Z_{t-1} = z|\theta)$ is defined in (1). In the i.i.d. case, $\tau(\theta|\theta')$ is the Dirac delta function $\delta(\theta - \theta')$; hence, $(\tau \circ \pi)(\theta) = \pi(\theta)$ and the above formula reduces to (3).

Analogous to (5), define the unnormalized prior as

$$\tilde{\pi}_t(\theta) = k(Z_{t-1} = z|\theta) \cdot (\tau \circ \tilde{\pi}_{t-1})(\theta),$$

for $t = 2, \dots, T$, with $\tilde{\pi}_1(\theta) = \pi_1(\theta)$. It is easy to show that $\tilde{\pi}_t = \pi_t \cdot \int_{\Theta} \tilde{\pi}_t(\theta)d\theta$. Furthermore, we have

$$\begin{aligned} \frac{\partial(\tau \circ (F(z|\cdot)(\tau \circ \tilde{\pi}_{t-1})))}{\partial z} &= \frac{\partial}{\partial z} \int_{\Theta} \tau(\theta|\theta')F(z|\theta')(\tau \circ \tilde{\pi}_{t-1})(\theta')d\theta' \\ &= \int_{\Theta} \tau(\theta|\theta')f(z|\theta')(\tau \circ \tilde{\pi}_{t-1})(\theta')d\theta' \\ &= (\tau \circ (f(z|\cdot)(\tau \circ \tilde{\pi}_{t-1}))) (\theta), \end{aligned}$$

where in the second equality, the derivative passes through the integration because $f(z|\theta')$ is a continuous function in z . Symmetrically, we also have

$$\frac{\partial(\tau \circ (\bar{F}(z|\cdot)(\tau \circ \tilde{\pi}_{t-1})))}{\partial z} = -(\tau \circ (f(z|\cdot)(\tau \circ \tilde{\pi}_{t-1})))'(\theta).$$

With these observations, by a similar proof argument, we can extend Theorem 1 to the Markov-modulated process, and, thus, all the subsequent results follow.

Finally, we comment that there is a related topic concerning the consistency of the Bayesian updating process under the guided policy (see, for example, Scarf 1959). Given the finite-horizon nature of our problem setting, it is not clear whether the Bayesian updating process will lead to a close approximation to the true distribution. To study this consistency problem, we would need to consider an infinite-horizon Markov decision process. The analysis lies beyond the scope of the present research, but merits investigation in future studies.

Appendix

Proof (Lemma 1) It suffices to show that $\tilde{G}_t(a, s, \tilde{\pi}) = G_t(a, s, \pi) \cdot \int_{\Theta} \tilde{\pi}(\theta) d\theta$. We prove the result by backward induction. For period $t = T$, $\tilde{G}_T(a, s, \tilde{\pi}) = R(a, s, \tilde{\pi})$. By definition (6), it is straightforward to show that $R(a, s, \tilde{\pi}) = R(a, s, \pi) \cdot \int_{\Theta} \tilde{\pi}(\theta) d\theta$. Now, assume the result holds for period $t + 1$. For period t , we have

$$\begin{aligned} \tilde{G}_t(a, s, \tilde{\pi}) &= R(a, s, \tilde{\pi}) + \int_{l(a,s)}^{r(a,s)} \tilde{V}_{t+1}(h(s, a, z), f(z|\cdot)\tilde{\pi}) dz + \tilde{V}_{t+1}(h(s, a, l(a, s)), F(l(a, s)|\cdot)\tilde{\pi}) \\ &\quad + \tilde{V}_{t+1}(h(s, a, r(a, s)), \bar{F}(r(a, s)|\cdot)\tilde{\pi}) \\ &= R(a, s, \tilde{\pi}) + \int_{l(a,s)}^{r(a,s)} V_{t+1} \left(h(s, a, z), \frac{f(z|\cdot)\tilde{\pi}}{\int_{\Theta} f(z|\theta)\tilde{\pi}(\theta) d\theta} \right) \cdot \left(\int_{\Theta} f(z|\theta)\tilde{\pi}(\theta) d\theta \right) dz \\ &\quad + V_{t+1} \left(h(s, a, l(a, s)), \frac{F(l(a, s)|\cdot)\tilde{\pi}}{\int_{\Theta} F(l(a, s)|\theta)\tilde{\pi}(\theta) d\theta} \right) \cdot \int_{\Theta} F(l(a, s)|\theta)\tilde{\pi}(\theta) d\theta \\ &\quad + V_{t+1} \left(h(s, a, r(a, s)), \frac{\bar{F}(r(a, s)|\cdot)\tilde{\pi}}{\int_{\Theta} \bar{F}(r(a, s)|\theta)\tilde{\pi}(\theta) d\theta} \right) \cdot \int_{\Theta} \bar{F}(r(a, s)|\theta)\tilde{\pi}(\theta) d\theta \\ &= \left\{ R(a, s, \pi) + \int_{l(a,s)}^{r(a,s)} V_{t+1} \left(h(s, a, z), \frac{f(z|\cdot)\pi}{\int_{\Theta} f(z|\theta)\pi(\theta) d\theta} \right) \cdot \left(\int_{\Theta} f(z|\theta)\pi(\theta) d\theta \right) dz \right. \\ &\quad + V_{t+1} \left(h(s, a, l(a, s)), \frac{F(l(a, s)|\cdot)\pi}{\int_{\Theta} F(l(a, s)|\theta)\pi(\theta) d\theta} \right) \cdot \int_{\Theta} F(l(a, s)|\theta)\pi(\theta) d\theta \\ &\quad \left. + V_{t+1} \left(h(s, a, r(a, s)), \frac{\bar{F}(r(a, s)|\cdot)\pi}{\int_{\Theta} \bar{F}(r(a, s)|\theta)\pi(\theta) d\theta} \right) \cdot \int_{\Theta} \bar{F}(r(a, s)|\theta)\pi(\theta) d\theta \right\} \cdot \int_{\Theta} \tilde{\pi}(\theta) d\theta \\ &= G_t(a, s, \pi) \cdot \int_{\Theta} \tilde{\pi}(\theta) d\theta, \end{aligned}$$

where the second equality follows from the induction assumption, the third equality follows from the fact that $\tilde{\pi} = \pi \cdot \int_{\Theta} \tilde{\pi}(\theta) d\theta$, and the last equality follows from definition (4). \square

Proof (Lemma 2) The result follows by a simple backward induction and the fact that \hat{a} is not necessarily the optimal solution to the original problem with the starting state (s, π) . \square

Proof (Theorem 1) Let us prove the result by backward induction. For period $t = T$,

$$V_T(s, F(z|\cdot)\pi_{T-1}) = \max_{a \geq 0} \left\{ R(a, s, F(z|\cdot)\pi_{T-1}) \right\}.$$

Let $\hat{a}(z)$ be an optimal solution to the above problem. Note that

$$R(a, s, F(z|\cdot)\pi_{T-1}) = \int_0^\infty \int_{\Theta} u(a, s, x) F(z|\theta) \pi_{T-1}(\theta) d\theta dx,$$

which is differentiable in z because $u(a, s, x) f(z|\theta) \pi_{T-1}(\theta)$ is continuous in z for all $a \geq 0$. Thus, $dR(a, s, F(z|\cdot)\pi_{T-1})/dz = R(a, s, f(z|\cdot)\pi_{T-1})$, which is bounded above by assumption. Also, note that $V_T(s, F(0|\cdot)\pi_{T-1}) = 0$. Then, by Theorem 2 of Milgrom and Segal (2002), we immediately have $V_T(s, F(z|\cdot)\pi_{T-1})$ is absolutely continuous in z for all s . In other words, $V_T(s, F(z|\cdot)\pi_{T-1})$ is differentiable in z almost everywhere, and when it is differentiable, $dV_T(s, F(z|\cdot)\pi_{T-1})/dz = R(\hat{a}(z), s, f(z|\cdot)\pi_{T-1})$. Utilizing the definition of Δ_T given by (8), we have

$$R(\hat{a}(z), s, f(z|\cdot)\pi_{T-1}) = V_T(s, f(z|\cdot)\pi_{T-1}) - \Delta_T(s, F(z|\cdot)\pi_{T-1}, f(z|\cdot)\pi_{T-1}).$$

Therefore, we arrive at

$$V_T(s, F(z|\cdot)\pi_{T-1}) = \int_0^z V_T(s, f(\zeta|\cdot)\pi_{T-1}) d\zeta - \int_0^z \Delta_T(s, F(\zeta|\cdot)\pi_{T-1}, f(\zeta|\cdot)\pi_{T-1}) d\zeta.$$

Now assume this holds for period $t + 1$. For period t , we have

$$\begin{aligned} G_t(a, s, F(z|\cdot)\pi_{t-1}) &= R(a, s, F(z|\cdot)\pi_{t-1}) + \int_{l(a,s)}^{r(a,s)} V_{t+1}(h(s, a, z'), f(z'|\cdot)F(z|\cdot)\pi_{t-1}) dz' \\ &\quad + V_{t+1}(h(s, a, l(a, s)), F(l(a, s)|\cdot)F(z|\cdot)\pi_{t-1}) \\ &\quad + V_{t+1}(h(s, a, r(a, s)), \bar{F}(r(a, s)|\cdot)F(z|\cdot)\pi_{t-1}). \end{aligned}$$

By the induction assumption, we know that $G_t(a, s, F(z|\cdot)\pi_{t-1})$ is differentiable in z almost everywhere, and when it is differentiable at z , we have for all a and s ,

$$\begin{aligned} &\frac{d}{dz} G_t(a, s, F(z|\cdot)\pi_{t-1}) \\ &= R(a, s, f(z|\cdot)\pi_{t-1}) + \int_{l(a,s)}^{r(a,s)} V_{t+1}(h(s, a, z'), f(z'|\cdot) f(z|\cdot)\pi_{t-1}) dz' \end{aligned}$$

$$\begin{aligned}
& - \int_{l(a,s)}^{r(a,s)} \Delta_{t+1}(h(s, a, z'), f(z'|\cdot)F(z|\cdot)\pi_{t-1}, f(z'|\cdot)f(z|\cdot)\pi_{t-1})dz' \\
& + V_{t+1}(h(s, a, l(a, s)), F(l(a, s)|\cdot)f(z|\cdot)\pi_{t-1}) \\
& - \Delta_{t+1}(h(s, a, l(a, s)), F(l(a, s)|\cdot)F(z|\cdot)\pi_{t-1}, F(l(a, s)|\cdot)f(z|\cdot)\pi_{t-1}) \\
& + V_{t+1}(h(s, a, r(a, s)), \bar{F}(r(a, s)|\cdot)f(z|\cdot)\pi_{t-1}) \\
& - \Delta_{t+1}(h(s, a, r(a, s)), \bar{F}(r(a, s)|\cdot)F(z|\cdot)\pi_{t-1}, \bar{F}(r(a, s)|\cdot)f(z|\cdot)\pi_{t-1}) \\
= & G_t(a, s, f(z|\cdot)\pi_{t-1}) - \int_{l(a,s)}^{r(a,s)} \Delta_{t+1}(h(s, a, z'), f(z'|\cdot)F(z|\cdot)\pi_{t-1}, f(z'|\cdot)f(z|\cdot)\pi_{t-1})dz' \\
& - \Delta_{t+1}(h(s, a, l(a, s)), F(l(a, s)|\cdot)F(z|\cdot)\pi_{t-1}, F(l(a, s)|\cdot)f(z|\cdot)\pi_{t-1}) \\
& - \Delta_{t+1}(h(s, a, r(a, s)), \bar{F}(r(a, s)|\cdot)F(z|\cdot)\pi_{t-1}, \bar{F}(r(a, s)|\cdot)f(z|\cdot)\pi_{t-1}),
\end{aligned}$$

where the last equality follows from the definition of G_t . Let $\hat{a}(z) = a^*(s, F(z|\cdot)\pi_{t-1})$. Then, by applying Theorem 2 of Milgrom and Segal (2002), we have $V_t(s, F(z|\cdot)\pi_{t-1})$ is absolutely continuous in z , and if $G_t(a, s, F(z|\cdot)\pi_{t-1})$ is differentiable at z (for all a and s), we have $V_t(s, F(z|\cdot)\pi_{t-1})$ is also differentiable at z and is given by

$$\begin{aligned}
& \frac{d}{dz} V_t(s, F(z|\cdot)\pi_{t-1}) \\
= & G_t(\hat{a}(z), s, f(z|\cdot)\pi_{t-1}) - \int_{l(\hat{a}(z), s)}^{r(\hat{a}(z), s)} \Delta_{t+1}(h(s, \hat{a}(z), z'), f(z'|\cdot)F(z|\cdot)\pi_{t-1}, f(z'|\cdot)f(z|\cdot)\pi_{t-1})dz' \\
& - \Delta_{t+1}(h(s, \hat{a}(z), l(\hat{a}(z), s)), F(l(\hat{a}(z), s)|\cdot)F(z|\cdot)\pi_{t-1}, F(l(\hat{a}(z), s)|\cdot)f(z|\cdot)\pi_{t-1}) \\
& - \Delta_{t+1}(h(s, \hat{a}(z), r(\hat{a}(z), s)), \bar{F}(r(\hat{a}(z), s)|\cdot)F(z|\cdot)\pi_{t-1}, \bar{F}(r(\hat{a}(z), s)|\cdot)f(z|\cdot)\pi_{t-1}) \\
= & V_t(s, f(z|\cdot)\pi_{t-1}) - \Delta_t(s, F(z|\cdot)\pi_{t-1}, f(z|\cdot)\pi_{t-1}),
\end{aligned}$$

where the last equality follows from the definition of Δ_t given by (8). Since $G_t(a, s, F(z|\cdot)\pi_{t-1})$ is differentiable in z almost everywhere and note that $V_t(s, F(0|\cdot)\pi_{t-1}) = 0$, by the absolute continuity of $V_t(s, F(z|\cdot)\pi_{t-1})$, we have

$$V_t(s, F(z|\cdot)\pi_{t-1}) = \int_0^z V_t(s, f(\zeta|\cdot)\pi_{t-1})d\zeta - \int_0^z \Delta_t(s, F(\zeta|\cdot)\pi_{t-1}, f(\zeta|\cdot)\pi_{t-1})d\zeta.$$

This completes the induction proof. Similarly, we can show the corresponding result holds for $V_T(s, \bar{F}(z|\cdot)\pi_{T-1})$. \square

Proof (Proposition 1) Let $a^e(s, \pi)$ be the unique optimal solution to $G_t^e(a, s, \pi)$. Thus, for any $a < a^e(s, \pi)$, $G_t^e(a, s, \pi) < G_t^e(a^e, s, \pi)$. If $G_t^\Delta(a, s, \pi)$ is increasing in a , then for any $a < a^e(s, \pi)$, $G_t^\Delta(a, s, \pi) \leq G_t^\Delta(a^e, s, \pi)$, or $G_t(a, s, \pi) \leq G_t(a^e, s, \pi) + G_t^e(a, s, \pi) - G_t^e(a^e, s, \pi) < G_t(a^e, s, \pi)$. Thus, the optimal solution to $G_t(a, s, \pi)$ must be greater than or equal to $a^e(s, \pi)$. By a similar

reasoning, we can show that if $G_t^\Delta(a, s, \pi)$ is decreasing in a , then the optimal solution to $G_t(a, s, \pi)$ must be less than or equal to $a^e(s, \pi)$. \square

Proof (Proposition 2) By Lemma 2, we have

$$\begin{aligned} \frac{d}{da} G_t^\Delta(a, \pi) &= \Delta_{t+1}(\bar{F}(a + S|\cdot)\pi, f(a + S|\cdot)\pi) - \Delta_{t+1}(F(a|\cdot)\pi, f(a|\cdot)\pi) \\ &\geq -\Delta_{t+1}(F(a|\cdot)\pi, f(a|\cdot)\pi). \end{aligned}$$

Note that $G_{t+1}(a, F(z|\cdot)\pi) = z \cdot G_{t+1}(a, (F(z|\cdot)/z)\pi)$. Hence, $a_{t+1}^*(F(z|\cdot)\pi) = a_{t+1}^*((F(z|\cdot)/z)\pi)$ for all $z > 0$. Because $\lim_{z \rightarrow 0} F(z|\cdot)/z = f(0|\cdot)$, we have

$$\lim_{z \rightarrow 0} a_{t+1}^*(F(z|\cdot)\pi) = \lim_{z \rightarrow 0} a_{t+1}^*((F(z|\cdot)/z)\pi) = a_{t+1}^*(f(0|\cdot)\pi).$$

Therefore, $\lim_{z \rightarrow 0} [V_{t+1}(f(z|\cdot)\pi) - G_{t+1}(a_{t+1}^*(F(z|\cdot)\pi), f(z|\cdot)\pi)] = 0$. By a backward induction, we can show $\lim_{a \rightarrow 0} \Delta_{t+1}(F(a|\cdot)\pi, f(a|\cdot)\pi) = 0$. Thus, letting a go to 0 in the above inequality, we conclude that $\lim_{a \rightarrow 0} dG_t^\Delta(a, \pi)/da \geq 0$. Similarly, we have

$$\begin{aligned} \frac{d}{da} G_t^\Delta(a, \pi) &= \Delta_{t+1}(\bar{F}(a + S|\cdot)\pi, f(a + S|\cdot)\pi) - \Delta_{t+1}(F(a|\cdot)\pi, f(a|\cdot)\pi) \\ &\leq \Delta_{t+1}(\bar{F}(a + S|\cdot)\pi, f(a + S|\cdot)\pi). \end{aligned}$$

Performing a transform of $y = 1/x$, we have $\bar{F}(a + S|\cdot) = \int_0^{1/(a+S)} f(1/y)y^2 dy$. By an analogous argument as shown above, we can show $\lim_{a \rightarrow \infty} \Delta_{t+1}(\bar{F}(a + S|\cdot)\pi, f(a + S|\cdot)\pi) = 0$. Thus, we conclude that $\lim_{a \rightarrow \infty} dG_t^\Delta(a, \pi)/da \leq 0$.

Therefore, there must exist two thresholds: $0 \leq \underline{a}(\pi) \leq \bar{a}(\pi)$, such that $dG_t^\Delta(a, \pi)/da \geq 0$ for $0 \leq a \leq \underline{a}(\pi)$ and $dG_t^\Delta(a, \pi)/da \leq 0$ for $a \geq \bar{a}(\pi)$. Since $R(a, S, \pi)$ is unimodal in a and $dG_t^e(a, \pi)/da = dR(a, S, \pi)/da$, it follows that $G_t^e(a, \pi)$ has a unique optimal solution. Hence, by Proposition 1, the results follow. \square

Proof (Proposition 3) With some calculus and using Lemma 2, it is straightforward to show that

$$\begin{aligned} \frac{d}{da} G_t^\Delta(a, s, \pi) &= \int_a^{a+s} \left[\frac{d}{da} V_{t+1}(s + a - x, f(x|\cdot)\pi) - \frac{d}{da} V_{t+1}^e(s + a - x, f(x|\cdot)\pi) \right] dx \\ &\quad - \Delta_{t+1}(s, F(a|\cdot)\pi, f(a|\cdot)\pi) \\ &\leq \int_a^{a+s} \left[\frac{d}{ds} V_{t+1}(s + a - x, f(x|\cdot)\pi) - \frac{d}{ds} V_{t+1}^e(s + a - x, f(x|\cdot)\pi) \right] dx \\ &= \int_0^s \left[\frac{d}{ds} V_{t+1}(s - y, f(y + a|\cdot)\pi) - \frac{d}{ds} V_{t+1}^e(s - y, f(y + a|\cdot)\pi) \right] dy. \end{aligned}$$

Because $\lim_{a \rightarrow \infty} f(a|\cdot) = 0$, we have $\lim_{a \rightarrow \infty} V_{t+1}(s, f(a|\cdot)\pi) = \lim_{a \rightarrow \infty} V_{t+1}^e(s, f(a|\cdot)\pi) = 0$ for all $s \geq 0$. Therefore, we conclude that

$$\lim_{a \rightarrow \infty} \int_0^s \left[\frac{d}{ds} V_{t+1}(s-y, f(y+a|\cdot)\pi) - \frac{d}{ds} V_{t+1}^e(s-y, f(y+a|\cdot)\pi) \right] dy = 0.$$

Hence, $\lim_{a \rightarrow \infty} dG_t^\Delta(a, s, \pi)/da \leq 0$. Therefore, there exist a threshold: $\bar{a}(s, \pi) \geq 0$, such that $dG_t^\Delta(a, \pi)/da \leq 0$ for all $a \geq \bar{a}(s, \pi)$. Hence, by Proposition 1, the result follows. If there is an ample supply of inventory or if the decision horizon is two periods, it is straightforward to show that $dG_t^\Delta(a, s, \pi)/da \leq 0$ for all a . Thus, we have $\bar{a}(s, \pi) \equiv 0$ in these cases. \square

Acknowledgements

The author would like to thank Paul Zipkin, Jeannette Song, Marty Lariviere, Damian Beil, Erica Plambeck, Kaijie Zhu, the seminar participants of Duke University, University of North Carolina, Michigan University, and Shanghai Jiao Tong University.

References

- Amemiya, T. 1984. Tobit models: A survey. *J. of Econometrics*. **24** 3–61.
- Araman, V., R. Caldentey. 2009. Dynamic pricing for non-perishable products with demand learning. *Oper. Res.* **57**(5) 1169–1188.
- Aviv, Y., A. Pazgal. 2005a. A partially observed Markov decision process for dynamic pricing. *Management Sci.* **51**(9) 1400–1416.
- Aviv, Y., A. Pazgal. 2005b. Dynamic pricing of short life-cycle products through active learning. Working Paper. Washington University, St Louis.
- Bensoussan, A., M. Cakanyildirim, S.P. Sethi. 2005. On the optimal control of partially observed inventory systems. *Comptes Rendus Mathematique*. **341** 419–426.
- Bensoussan, A., M. Cakanyildirim, S.P. Sethi. 2007. A multi-period newsvendor problem with partially observed demand. *Math. of Oper. Res.* **32**(2) 322–344.
- Bensoussan, A., M. Cakanyildirim, S.P. Sethi. 2008. Inventory problems with partially observed demands and lost sales. *J. Optim. Theory and Applications*. **136**(2) 321–340.
- Bensoussan, A., M. Cakanyildirim, S.P. Sethi. 2009. A note on “The censored newsvendor and the optimal acquisition of information”. *Oper. Res.* **57**(3) 791–794.

- Bisi, A., M. Dada. 2007. Dynamic learning, pricing, and ordering by a censored newsvendor. *Naval Res. Logist. Quart.* **54** 448–461.
- Bisi, A., M. Dada, S. Tokdar. 2011. A censored-data multiperiod inventory problem with newsvendor demand distributions. *Manufacturing & Service Oper. Management*. Forthcoming.
- Braden, D.J., M. Freimer. 1991. Informational dynamics of censored observations. *Management Sci.* textbf37(11) 1390–1404.
- Braden, D.J., S.S. Oren. 1994. Nonlinear pricing to produce information. *Marketing Sci.* **13**(3) 310–326.
- Chen, L. 2010. Bounds and heuristics for optimal Bayesian inventory control with unobserved lost sales. *Oper. Res.* Forthcoming.
- Chen, L., E.L. Plambeck. 2008. Dynamic inventory management with learning about the demand distribution and substitution probability. *Manufacturing & Service Oper. Management.* **10**(2) 236–256.
- Ding, X., M.L. Puterman, A. Bisi. 2002. The censored newsvendor and the optimal acquisition of information. *Oper. Res.* **50**(3) 517–527.
- Dynkin, E., A. Yushkevich. 1979. *Controlled Markov Processes*. Springer, New York.
- Farias, V., B. Van Roy. 2010. Dynamic pricing with a prior on market response. *Oper. Res.* **58**(1) 16–29.
- Harpaz, G., W.Y. Lee, R.L. Winkler. 1982. Learning, experimentation, and the optimal output decisions of a competitive firm. *Management Sci.* **28**(6) 589–603.
- Iglehardt, D., S. Karlin. 1962. Optimal policy for dynamic inventory process with nonstationary stochastic demands. *Studies in Applied Probability and Management Science*. K. Arrow, S. Karlin, H. Scarf (eds.). Chapter 8, Stanford University Press, Stanford, California.
- Lariviere, M.A., E.L. Porteus. 1999. Stalking information: Bayesian inventory management with unobserved lost sales. *Management Sci.* **45**(3) 346–363.
- Lovejoy, W.S. 1993. Suboptimal policies, with bounds, for parameter adaptive decision processes. *Oper. Res.* **41**(3) 583–599.
- Lu, X., J.-S. Song and K. Zhu. 2005. On “The censored newsvendor and the optimal acquisition of information”. *Oper. Res.* **53**(6) 1024–1026.
- Lu, X., J.S. Song, K. Zhu. 2008. Analysis of perishable inventory systems with censored data.

- Oper. Res.* **56**(4) 1034–1038.
- Milgrom, P., I. Segal. 2002. Envelope theorems for arbitrary choice sets. *Econometrica.* **70**(2) 583–601.
- Monahan, G.E. 1982. A survey of partially observable Markov decision processes: model, theory and algorithms. *Management Sci.* **28**(1) 1–16.
- Nahmias, S. 1994. Demand estimation in lost sales inventory systems. *Naval Res. Logist.* **41** 739–757.
- Scarf, H.E. 1959. Bayes solution of the statistical inventory problem. *Ann. Math. Statist.* **30** 490–508.
- Song, J.-S., P. Zipkin. 1993. Inventory control in a fluctuating demand environment. *Oper. Res.* **41**(2) 351–370.
- Tobin, J. 1958. Estimation of relationships for limited dependent variables. *Econometrica.* **26**(1) 24–36.
- Ziya, S., H. Ayhan, R.D. Foley. 2004. Relationships among three assumptions in revenue management. *Oper. Res.* **52**(5) 804–809.