Optimal Inventory Policies when the Demand Distribution is not Known

C. Erik Larson, Lars J. Olson, and
Sunil Sharma

IMF Working Paper

IMF Institute

Optimal Inventory Policies when the Demand Distribution is not Known

Prepared by C. Erik Larson, Lars J. Olson, and Sunil Sharma[1]

Authorized for distribution by Eric V. Clifton

November 2000

Abstract

The views expressed in this Working Paper are those of the author(s) and do not necessarily represent those of the IMF or IMF policy. Working Papers describe research in progress by the author(s) and are published to elicit comments and to further debate.

This paper analyzes the stochastic inventory control problem when the demand distribution is not known. In contrast to previous Bayesian inventory models, this paper adopts a non-parametric Bayesian approach in which the firm's prior information is characterized by a Dirichlet process prior. This provides considerable freedom in the specification of prior information about demand and it permits the accommodation of fixed order costs. As information on the demand distribution accumulates, optimal history-dependent (s,S) rules are shown to converge to an (s,S) rule that is optimal when the underlying demand distribution is known.

---

[1] Erik Larson is a staff member of the US Treasury's Office of the Comptroller of the Currency, Lars Olson is on the faculty of the Department of Agricultural and Resource Economics, University of Maryland, College Park, and Sunil Sharma is a staff member of the IMF Institute. A substantial portion of the research was carried out when the authors were at USC, UC Riverside and UCLA, respectively. The authors are grateful to Ken Burdett, Thomas Ferguson, Nicholas Kiefer, Steven Lippman, Bruce Miller and Yaw Nyarko for comments. A version of this paper is forthcoming in the *Journal of Economic Theory*.

- 2 -

Contents

I. INTRODUCTION

Since its introduction by Arrow, Harris and Marschak [1951], the (s, S) inventory model has been used in a variety of ways. It has been used to characterize the dynamics of inventory adjustment by firms (see the survey by Porteus [1990]), understand the demand for money by different agents, analyze price-setting behavior when there are fixed costs associated with a decision to change prices, and examine the macroeconomic implications of inventory policies.

Inventory adjustment models generally assume that agents (hereafter, firms) know the distribution of demand. It seems more plausible that firms do not know the true distribution of demand, but that they learn about it as they accumulate information over time. The purpose of this paper is to develop more fully the theory of the (s, S) inventory model under learning.

Previous research on this problem has focussed on a parametric Bayesian approach that assumes the true demand distribution belongs to some parametric family characterized by a finite number of unknown parameters (Scarf [1959a, 1960], Karlin [1960], Iglehart [1964], Azoury and Miller [1984], Azoury [1985], and Lovejoy [1990]; see also Dvoretzky, Kiefer and Wolfowitz [1952]). The firm's information about demand is specified by some conjugate prior on the unknown parameters and updated via Bayes' Rule. Typically, the choice of a conjugate family of distributions places restrictions on the prior information that can be accommodated and the true distributions of demand that can be allowed. For example, it is difficult under standard conjugate family specifications to allow for bi-modal priors or bi-modal true demand distributions. Another key assumption underlying parametric Bayesian models in the inventory literature is that purchasing costs are convex. This rules out the possibility of fixed order costs and it implies optimal policies are degenerate in the sense that $s = S$.

This paper adopts a non-parametric Bayesian approach that provides greater flexibility in modeling prior information about demand, allows for a larger set of true (but unknown) demand distributions, and accommodates fixed order costs that make the cost function non-convex. The firm's information about the demand distribution is characterized by a Dirichlet process prior on the space of distributions (Ferguson [1973]). Given a Dirichlet process prior, any distribution whose support is included in the support of the measure characterizing the Dirichlet process prior can be approximated as a posterior. Our approach leads to results on time-dependent (s, S) policies under learning that incorporate fixed order costs and are analogous to those obtained when the distribution is known.

Section II introduces the Dirichlet process and discusses how it is used to model prior beliefs and how those beliefs are updated as the firm learns about demand. Section III develops a dynamic programming formulation of the problem. The state space is defined on the beginning-of-period inventory level and the measure characterizing the firm's beliefs

about the demand distribution. For both the finite and infinite planning horizon formulations of the model, a history-dependent (s,S) policy is optimal at each stage. We prove that as information on the demand distribution accumulates, these history-dependent (s,S) rules converge to the optimal (s,S) rule for the case where the underlying demand distribution is known.

Comparative dynamic results for the case where there is no fixed ordering cost are given in Section IV. If the expected demand distributions under two Dirichlet process priors are ordered by first-order stochastic dominance, then the critical stock levels are ordered for any common history of demand observations. This extends results obtained by Karlin [1960] and Scarf [1959a]. We also show that if in any period the current demand realization is less than (greater than) the minimum (maximum) of those observed to date, the optimal order-to-level S decreases or remains the same (increases or remains the same) for the following period. With no fixed ordering cost the time-dependent optimal policy is degenerate in the sense that $s = S$ in each period. Even so, if the current demand realization is sufficiently small it is possible that the optimal policy under learning is to refrain from ordering.

Section V offers some concluding remarks. Proofs of all results are given in the Appendix.

II. THE DIRICHLET PROCESS

The firm's initial beliefs about the true demand distribution are characterized by a prior distribution over the space of probability measures on $(\Omega,\mathcal{B})$ where Ω is taken to be a compact subset of $\mathfrak{R}_+$, the nonnegative real numbers and $\mathcal{B}$ is the σ-algebra of Borel subsets of Ω. The prior distributions used in this paper are the Dirichlet process priors (Ferguson [1973]). This nonparametric approach to Bayesian learning has received relatively little attention in the literature on dynamic programming models under uncertainty.[2]

Formally, let $\mathcal{F}(\Omega)$ be the space of probability measures on $(\Omega,\mathcal{B})$. The true, but unknown, demand distribution is denoted by F^* and the support of F^* is assumed to be a compact subset of Ω. Let $\mathcal{B}$ be the σ-algebra of Borel subsets of $\mathcal{F}(\Omega)$ with respect to the topology of weak convergence and let $\mathcal{F}(\mathcal{F}(\Omega))$ denote the space of probability measures on

---

[2] Rothschild's [1974] early work on optimal search for the lowest price from an unknown, discrete price distribution utilizes the Dirichlet distribution (see also Talman [1992]). The more general problem of optimal search under a Dirichlet process is analyzed by Christensen [1986] and Bikhchandani and Sharma [1990]. Clayton and Berry [1985] develop a nonparametric Bayesian analysis of the two-armed bandit problem using the Dirichlet process.

$(\mathcal{F}(\Omega), \mathcal{B})$. The firm's beliefs about the distribution of demand are represented by a measure $\mathcal{P} \in \mathcal{F}(\mathcal{F}(\Omega))$. Let ν be a random probability distribution on Ω chosen according to $\mathcal{P}$ so that both $\int \omega \, \nu(d\omega)$ and $\nu(B)$ for $B \in \mathcal{B}$ are random variables. Let $\mathcal{E}[.]$ denote the expectation operation with respect to $\mathcal{P}$. Then $\mathcal{E}[\nu]$ is a probability measure on $(\Omega, \mathcal{B})$. If F is the distribution function associated with ν, $F(\omega) = \nu((-\infty, \omega])$, then the distribution function associated with $\mathcal{E}[\nu]$ is denoted by $\mathcal{E}[F]$. Let $\omega_1, ..., \omega_n$ denote a random sample of size n chosen according to ν. The information contained in $\omega_1, ..., \omega_n$ is used to update the firm's beliefs about the true demand distribution.

Definition (Ferguson [1973]). *Let $\alpha(\cdot)$ be a finite non-null measure on $(\Omega, \mathcal{B})$ and let ν be a random process indexed by elements of $\mathcal{B}$. ν is a **Dirichlet process** with parameter α, written $\nu \in \mathfrak{D}(\alpha)$, or equivalently $F \in \mathfrak{D}(\alpha)$, if for every finite measurable partition $\{I_1, ..., I_m\}$ of Ω, the random vector $(\nu(I_1), ..., \nu(I_m))$ has a Dirichlet distribution with parameter $(\alpha(I_1), ..., \alpha(I_m))$.*

It is easy to see that if g is any measurable function on $(\Omega, \mathcal{B})$ then

$$\mathcal{E}\left(\int g(\omega) \, dF(\omega)\right) = \int g(\omega) \, d\mathcal{E}F(\omega).$$

Ferguson [1973, 1974] discusses three properties of the Dirichlet process. First, under the topology of weak convergence the support of $\mathfrak{D}(\alpha)$ is the set of all probability distributions whose support is contained in the support of α (Ferguson [1973, p. 216]). This means that the support of any firm's beliefs can be approximated arbitrarily close by the Dirichlet process. It does not imply the stronger property that the beliefs of any firm can be approximated arbitrarily close by the Dirichlet process.[3] Hence, while the Dirichlet process is rather flexible as a model of the underlying demand distribution, it is less so as a model of beliefs. The second useful property noted by Ferguson is that for any non-negative measurable function g, if $\int g(\omega) \, d\alpha(\omega) < \infty$ then $\int g(\omega) \, dF(\omega) < \infty$ almost surely (Ferguson [1973, Theorem 3]). This establishes a strong connection between properties of the

---

[3] We thank Thomas Ferguson for the following simple counter-example. Let $\Omega = \{0, 1\}$ and define $\theta_0 = \Pr(\omega = 0)$ and $\theta_1 \equiv \Pr(\omega = 1) = 1 - \theta_0$. The Dirichlet processes are then Beta distributions on the space of Binomial distributions with parameter θ_0, and the Dirichlet process cannot approximate beliefs characterized by an initial prior over θ_0 of, say, $\Pr(\theta_0 = 1/4) = (1/2)$ and $\Pr(\theta_0 = 3/4) = (1/2)$.

measure α of the Dirichlet process and those of the random distribution function F selected by the Dirichlet process: if a kth moment of α exists then a kth moment of F exists with probability one. Finally, if $F \in \mathfrak{D}(\alpha)$, then with probability one F is discrete. This has the limitation that if the set of possible distribution functions is a subset of the continuous distributions, the Dirichlet process assigns probability zero to the true set.

Given a Dirichlet process prior, and a sample $\omega_1, \ldots, \omega_n$ from the true (but unknown) demand distribution, the posterior distribution on the space of distributions is also a Dirichlet process.

Fact 1 (Ferguson [1973]). *If* $F \in \mathfrak{D}(\alpha)$ *and if* $\omega_1, \ldots, \omega_n$ *is a sample from* F, *then the posterior distribution of* F *given* $\omega_1, \ldots, \omega_n$, *denoted* $F | \omega_1, \ldots, \omega_n$, *is* $\mathfrak{D}\left(\alpha + \sum_i \delta_{\omega_i}\right)$, *a Dirichlet process with parameter* $\alpha + \sum_i \delta_{\omega_i}$ *where* δ_ω *is the measure on* $(\mathfrak{R}_+, \mathcal{B})$ *that assigns mass one to* ω.

The measure characterizing the updated Dirichlet process after n observations is given by $\alpha_n(\omega_1, \ldots, \omega_n) = \alpha + \sum_{i=1}^{n} \delta_{\omega_i}$. For brevity we will denote this by α_n, it being understood that α_n depends on the values of realized demands $\omega_1, \ldots, \omega_n$ and not just on the number of observations, n. In addition, let $\alpha(\omega)$ denote $\alpha((-\infty, \omega])$. Fact 1 implies that

$$\hat{F}_n \equiv \mathcal{E}\left[F(\omega) | \omega_1, \ldots, \omega_n\right] = \frac{\alpha(\omega) + \sum_{i=1}^{n} 1_{[\omega_i, \infty)}(\omega)}{\alpha(\Omega) + n}$$

$$= \left[\frac{\alpha(\Omega)}{\alpha(\Omega) + n}\right] \frac{\alpha(\omega)}{\alpha(\Omega)} + \left[\frac{n}{\alpha(\Omega) + n}\right] \frac{\sum_{i=1}^{n} 1_{[\omega_i, \infty)}(\omega)}{n} = \lambda_n \cdot F_0(\omega) + (1 - \lambda_n) \Gamma_{\omega_1, \ldots, \omega_n}(\omega)$$

where $\mathcal{E}$ is the expectation operator with respect to $\mathcal{P}$, $1_{[\omega_i, \infty)}$ is the indicator function of the set $[\omega_i, \infty)$, $\lambda_n = \alpha(n) / [\alpha(\Omega) + n]$, $F_0(\omega) = \alpha(\omega) / \alpha(\Omega)$, and $\Gamma_{\omega_1, \ldots, \omega_n}$ is the empirical distribution function after observing $\omega_1, \ldots, \omega_n$. Note that the expected distribution under the Dirichlet process is the expectation or mean of the distribution on the space of distributions, and hence is an element in the space of distributions; it is a weighted average of the expected distribution under the initial Dirichlet prior and the empirical distribution.

The updating of beliefs prescribed by the Dirichlet process is special in that

information provided by an observation is completely local. If a particular ω is observed, then the expected distribution under the updated Dirichlet process assigns a higher probability to any subset of the support of α that contains ω and uniformly decreases the probability of all subsets that do not contain ω (see also Blackwell and MacQueen (1973)).

$\alpha(\Omega)$ can be interpreted as a measure of confidence in the prior. If $\alpha(\Omega)$ is large relative to the sample size n, then greater weight or confidence is placed on the prior and the firm's beliefs respond more slowly to new information about demand. However, as the firm observes demand (and n becomes large) the expected distribution under the updated Dirichlet process becomes closer to the empirical distribution. More specifically, as $n \to \infty$ the expected distribution places greater weight on the empirical distribution $\Gamma_{\omega_1,...,\omega_n}$ and, by the Glivenko-Cantelli theorem, the expected distribution converges uniformly almost surely to the true demand distribution (Ferguson [1973, p. 223]).[4] As a consequence, the Dirichlet process provides a consistent estimator of the true demand distribution (e.g., Fabius [1964, Theorem 2.2]).

It is worthwhile to contrast Bayesian learning under the Dirichlet process with Bayesian updating in conjugate families characterized by finite-dimensional parameters. In the latter, a single observation provides information about the relative probabilities of other possible outcomes so that observing a high outcome typically implies that other high outcomes are also more likely. The Dirichlet process does not smooth beliefs in this manner, but leaves the relative probabilities of unobserved outcomes unchanged. In this regard the Dirichlet approach is likely to be most useful when observations occur frequently, or when the absence of smoothing is of minor concern when compared to the approximation errors that may results from the use of a finite-parameter conjugate family. Further, for finite parameter models, the true distribution belongs to a specific parametric class and the agent's beliefs are a member of the conjugate family for that class (Degroot (1970)). In contrast, the nonparametric nature of the Dirichlet process provides a consistent model of learning about any demand distribution whose support is a subset of the support of α. This is particularly useful when, a priori, very little is known about the true distribution.

III. OPTIMAL INVENTORY POLICIES UNDER LEARNING

In the single commodity inventory control problem a firm must decide each period whether to order additional inventory, and if so, how much. In doing so the firm faces a tradeoff between the costs of holding inventory (if the inventory level is too high) and the costs of not being able to satisfy consumer demand (if the inventory level is too low). Each order for additional inventory entails a variable cost depending on the quantity ordered and a fixed cost which is independent of the order size. The objective of the firm is to minimize the

---

[4] This assumes that demand is accurately observed each period. As Lovejoy [1993] notes, this may not occur if demand is not backlogged and lost sales go unobserved.

expected discounted sum of all of these costs over time. In the standard formulation, the per unit cost of holding inventory, h, and the per unit penalty cost associated with unmet demand, p, are assumed to be linear and are based on the end of period inventory level. The fixed order cost is denoted by K and the per-unit order cost by c, where it is assumed that $c < p$. The ordering cost function C is given by

$$C(u) = \begin{cases} K + c \cdot u & \text{if } u > 0 \\ 0 & \text{if } u = 0. \end{cases}$$

All cost parameters are assumed to be non-negative and the discount factor β satisfies $0 < \beta < 1$. The presence of a strictly positive fixed ordering cost makes the ordering cost function non-convex. It is assumed that excess demand is backlogged and there is no lag between ordering and delivery. The standard formulation assumes that the demand for the good in each period is random with a known distribution. In this paper we assume that the demand distribution is not known and that the prior beliefs of the firm on the space of demand distributions can be characterized by the Dirichlet process.

Let the initial inventory stock be given by x and let u denote the order quantity. Having observed n previous demand realizations, the expected holding and shortage cost for a one-period planning horizon is given by

$$\mathscr{E}\left\{ \int L(x+u,\omega)\, dF(\omega) \mid \omega_1,\ldots,\omega_n \right\} = \int L(x+u,\omega)\, d\mathscr{E}\left\{ F(\omega) \mid \omega_1,\ldots,\omega_n \right\} = \int L(x+u,\omega)\, d\hat{F}_n(\omega)$$

where $L(x+u,\omega) = h \cdot Max[x+u-\omega,0] + p \cdot Max[\omega-(x+u),0]$ and $\hat{F}_n(\omega)$ is the expectation of the demand distribution under α_n, the updated Dirichlet measure given a history of n demand realizations. Let $V_T(x,\alpha_n)$ be the minimum expected sum of discounted costs with T-periods to go until the end of the planning horizon when x is the current inventory level, α_n is the measure characterizing the firm's beliefs about the demand distribution, and an optimal ordering policy is followed in the future. Standard dynamic programming arguments (e.g., Schal [1975]) imply that there exists an optimal policy that satisfies the following functional equation for $T = 1,2,\ldots,\infty$:

$$V_T(x,\alpha_n) = \inf_u \left[C(u) + \int \{ L(x+u,\omega) + \beta\, V_{T-1}(x+u-\omega,\alpha_n+\delta_\omega) \}\, d\hat{F}_n(\omega) \right]$$

where $V_0 = 0$. Furthermore, the function V_T is lower-semicontinuous.

Define the post-order inventory level by $z = x+u$. In characterizing the optimal solution it is useful to define the cost functions

$$G_1(z,\alpha_n) = c \cdot z + \int L(z,\omega) \, d\hat{F}_n(\omega) \qquad (3.1)$$

$$G_T(z,\alpha_n) = c \cdot z + \int \{L(z,\omega) + \beta \, V_{T-1}(z-\omega,\alpha_n + \delta_\omega)\} \, d\hat{F}_n(\omega) \qquad (3.2)$$

Note that G_1 is convex in z and $\lim_{|z|\to\infty} G_1(z,\alpha_n) = \infty$. Let $S_1(\alpha_n)$ minimize $G_1(z,\alpha_n)$ in z and let $s_1(\alpha_n)$ be the smallest value of z such that $G_1(z,\alpha_n) = K + G_1(S_1(\alpha_n),\alpha_n)$. Using arguments developed by Scarf [1960b], the following facts can be established:

(i) $G_T(z,\alpha_n)$ is K-convex in z,[5]

(ii) $\lim_{|z|\to\infty} G_T(z,\alpha_n) = \infty$,

(iii) $V_T(x,\alpha_n)$ is K-convex and continuous in x for $T = 1,2,\ldots,\infty$,

(iv)   There exist scalars $(s_T(\alpha_n), S_T(\alpha_n))$ such that $S_T(\alpha_n)$ minimizes $G_T(z,\alpha_n)$ in z, and $s_T(\alpha_n)$ is the smallest value of z for which $G_T(z,\alpha_n) = K + G_T(S_T(\alpha_n),\alpha_n)$.

Define $u_T(x,\alpha_n)$ to be the optimal order quantity when the current inventory level is x and beliefs about the demand distribution are characterized by α_n. Given (i)-(iv), the classic arguments of Scarf [1960b] and Iglehart [1963] can be used to show that the optimal inventory policy is an (s,S) inventory rule that varies as expectations change in response to the observed history of demand observations.

LEMMA 3.1. *The optimal inventory policy is*

$$u_T(x,\alpha_n) = \begin{cases} S_T(\alpha_n) - x & \text{if } x < s_T(\alpha_n) \\ 0 & \text{if } x \geq s_T(\alpha_n) \end{cases}$$

for $T = 1,2,\ldots,\infty$.

      One important issue that arises is how does the optimal stocking rule behave as the firm accumulates information about the true demand distribution. Do optimal policies converge as the number of demand observations increases, and if so, what are the limit policies? The answers to these questions are complicated by the fact that beliefs about demand enter the firm's optimization problem in two ways. First, the firm's expectations over future discounted costs are determined by current beliefs. Second, beliefs are a state

---

[5] A real valued function $f(\cdot)$ is K-convex $(K \geq 0)$ if for all $x \in \Re$, $y \geq 0$, and $z > 0$ we have $K + f(x+y) \geq (y/z)[f(x) - f(x-z)]$.

variable in the firm's dynamic programming problem.[6] We analyze convergence by first proving that the solution to the firm's dynamic optimization yields an equicontinuous family of value functions. This then enables us to develop the main result of this section which characterizes the limiting behavior of optimal inventory policies under learning.

Consider the family of T-period planning horizon value functions $\{V_T(x,\alpha_n), n = 1,...\}$, where n is the number of previously observed demands.

LEMMA 3.2. *The family of value functions* $\{V_T(x,\alpha_n), n = 1,...\}$ *is equicontinuous at* x *for all finite* T.

Using Lemma 3.2 we now prove that as the number of demand observations increases, the history dependent optimal inventory policies under learning converge to policies that are optimal when the true demand distribution is known.

THEOREM 3.3. *For all* $T = 1, 2, ..., \infty$

(i) $\lim_{n \to \infty} V_T(x,\alpha_n) = V_T^*(x)$ *a.s.*

for all x *in any finite interval, where* V_T^* *is the value function for the inventory problem with known demand distribution,* F^*; *and*

(ii) for all $\varepsilon > 0$ *there exists an* $N(T,\varepsilon)$ *such that for all* $n > N(T,\varepsilon)$ *we have*

$$0 \le C\big(u_T(x,\alpha_n)\big) + \int\{L\big(x + u_T(x,\alpha_n),\omega\big) + \beta V_{T-1}^*\big(x + u_T(x,\alpha_n) - \omega\big)\}\, dF^* - V_T^*(x) < \varepsilon.$$

Further, if $\lim_{n \to \infty} u_T(x,\alpha_n)$ *exists a.s., then*

$$\lim_{n \to \infty}\big(s_T(\alpha_n), S_T(\alpha_n)\big) = \big(s_T^*, S_T^*\big) \quad a.s.,$$

where $\big(s_T^*, S_T^*\big)$ *are optimal for the inventory problem with known demand distribution,* F^*.

---

[6] The latter consideration precludes use of the Portmanteau theorem (Billingsley [1968, Theorem 2.1]) to analyze the convergence of policies. Further, since the convergence of beliefs is weaker than setwise convergence, the convergence result of Royden [1968, Proposition 17] does not apply. Finally, the problem satisfies neither the joint continuity nor the monotonicity properties required by Dutta, Majumdar and Sundaram [1994].

The first part of Theorem 3.3 states that once firms observe a sufficient number of demand realizations, the value of optimal policies under learning will be close to the optimal value when the true demand distribution is known.[7] The second part of the theorem gives a related result on the value of following u_T^n in the problem where the demand distribution is known. We are not able to make a stronger statement regarding the convergence of the optimal policies themselves because it has not been shown that such policies are unique. Theorem 3.3 is particularly relevant for situations where information accumulates rapidly, as is the case in many retail and wholesale operations where inventory levels are monitored on a daily or weekly basis. With the advent of computer bar-code technology there is the potential for very rapid updating of information since inventory levels can be monitored almost continuously.

IV. COMPARATIVE DYNAMICS

In this Section we assume that there is no fixed ordering cost, i.e., K = 0. We prove two results. The first compares optimal inventory policies under beliefs represented by two Dirichlet processes, $\mathfrak{D}(\alpha^1)$ and $\mathfrak{D}(\alpha^2)$ with $\alpha^1(\Omega) = \alpha^2(\Omega)$. The second result shows that if demand in any period is less than the minimum of past demand realizations, the optimal desired inventory level decreases.

When it exists, let V_T' denote the derivative of V_T with respect to x. For any two Dirichlet measures, α^1 and α^2, let $\hat{F}^1(\omega) = \left[\alpha^1(\omega)/\alpha^1(\Omega)\right]$ and $\hat{F}^2(\omega) = \left[\alpha^2(\omega)/\alpha^2(\Omega)\right]$ be the expected distributions before any observations are taken. The following lemma leads to the main result.

LEMMA 4.1. *Assume K = 0 and consider two Dirichlet measures α^1 and α^2, such that $\alpha^1(\Omega) = \alpha^2(\Omega)$. If $\hat{F}^1$ first-order stochastically dominates $\hat{F}^2$ then $V_T'(x, \alpha_n^1) \leq V_T'(x, \alpha_n^2)$. (If the derivative is not well-defined then the inequality holds for the right hand and left hand derivatives.)*

The main results of this section are stated below.

THEOREM 4.2. *Assume K = 0. Consider two Dirichlet measures α^1 and α^2, such that $\alpha^1(\Omega) = \alpha^2(\Omega)$. For simplicity assume that α^1 and α^2 are continuous with continuous first*

---

[7] While similar, Boylan's [1969] stability result for solutions to the optimal inventory equation does not apply under (Bayesian) learning.

derivatives.[8] *If* $\hat{F}^1(\omega)$ *first-order stochastically dominates* $\hat{F}^2(\omega)$, *then*

$$S_T(\alpha_n^1) \geq S_T(\alpha_n^2) \text{ for all } n, T \text{ and any common history of demand realizations.}$$

Theorem 4.2 implies that if the expected distribution under α^1 first order stochastically dominates the expected distribution under α^2 then the optimal inventory level under the updated Dirichlet process $\mathfrak{D}(\alpha_n^1)$ is always larger than that under the updated Dirichlet process $\mathfrak{D}(\alpha_n^2)$ for any common history of demand realizations. This result complements those obtained by Karlin [1960, Theorems 2 and 2'] for the case of known, but time-varying, demand distributions.

THEOREM 4.3. *If $K = 0$ then for all n and T the optimal inventory levels satisfy*

$$S_T(\alpha_n) \geq S_T(\alpha_n + \delta_{\omega \leq \text{Min}[\omega_1, \ldots, \omega_n]}).$$ *Furthermore, if* $S_T(\alpha_n) > S_T(\alpha_n + \delta_{\omega = 0})$ *then there exists a critical demand level* ω^* *such that if* $0 < \omega_{n+1} \leq \omega^*$, *the optimal policy is not to order additional inventory immediately after observing* ω_{n+1}.

The theorem says that a historically low demand realization leads to a pessimistic reassessment of the underlying true demand distribution and a downward adjustment in the optimal inventory level. This downward adjustment may be such that no additional inventory is desired, even after a positive demand realization. This is contrary to the behavior of optimal inventory policies with a known demand distribution and no fixed costs. Under the latter, if the inventory is at its desired level then any positive demand leads to a restocking of inventory to its desired level. Scarf [1959, Theorem 3] proves a related result in a model where demand has a density in the exponential class.

V. CONCLUDING REMARKS

This paper analyzes the stochastic inventory control problem when demand is identically and independently distributed according to a distribution that is not known to the firm. In principle, it should be possible to extend the analysis to the case where the demand distribution is characterized by a finite Markov chain with stationary, but unknown, transition probabilities. In addition, with recent and continuing gains in computing power, it is becoming increasingly possible to calculate optimal inventory rules for models that incorporate non-parametric learning. Further progress could be made if results on the reduction of state space dimensionality, similar to those of Scarf [1960a] and Azoury [1985], are obtained for non-parametric Bayesian inventory models. If this proves difficult or impossible in these models, it would be of interest to extend Lovejoy [1993] and develop bounds on the loss from using non-optimal, but simple and readily implementable policies.

---

[8] These continuity assumptions simplify the proof but are not essential to the result.

VI. APPENDIX: PROOFS

Proof of Lemma 3.1. The proof follows the classic arguments of Scarf [1960b] and Iglehart [1963].

Proof of Lemma 3.2. The proof proceeds by induction. Assume $\{V_{T-1}(x,\alpha_n), n = 1,...\}$ is an equicontinuous family at x. Since $\{L(z,\omega), \omega \in \Omega\}$ at z and $\{V_{T-1}(x,\alpha_n), n = 1,...\}$ at x are equicontinuous families, it is straightforward to show that $\{G_T(z,\alpha_n), n = 1,...\}$ is also equicontinuous at z, where $G_T(\cdot)$ is defined in (3.2).

For the sake of brevity let (s_T^n, S_T^n) denote the optimal policy when beliefs are characterized by α_n and there are T periods remaining, with the dependence on α_n being understood. Let z and z' be the optimal orders from x and x', respectively. It follows that

$$z = \begin{cases} x & \text{if } x > s_T^n \\ S_T^n & \text{if } x \leq s_T^n \end{cases} \quad \text{and} \quad z' = \begin{cases} x' & \text{if } x' > s_T^n \\ S_T^n & \text{if } x' \leq s_T^n. \end{cases}$$

Without loss of generality assume $x' \leq x$. We want to show that for any $\varepsilon > 0$, there exists a δ_ε such that $|x - x'| < \delta_\varepsilon$ implies $|V_T(x,\cdot) - V_T(x',\cdot)| < \varepsilon$. Three cases need to be considered.

Case 1: $z = z' = S_T^n$.

In this case, $|V_T(x,\cdot) - V_T(x',\cdot)| = |-c \cdot x + G_T(S_T^n, \alpha_n) - [c \cdot x + G_T(S_T^n, \alpha_n)]| = c(x - x')$. Given ε choose $\delta_\varepsilon = (\varepsilon/c)$. Equicontinuity of the family $\{V_T(x,\alpha_n), n = 1,...\}$ follows directly.

Case 2: $z = x, \quad z' = S_T^n$.

From $x' \leq s_T^n \leq S_T^n$ it follows that $|z - z'| = |x - S_T^n| \leq |x - s_T^n| \leq |x - x'|$. Note that

$$V_T(x,\alpha_n) = \begin{cases} K - c \cdot x + G_T(S_T^n, \alpha_n) & \text{if } x < s_T^n \\ -c \cdot x + G_T(x,\alpha_n) & \text{if } x \geq s_T^n. \end{cases}$$

This gives

$$\left| V_T(x,\cdot) - V_T(x',\cdot) \right| = \left| -c \cdot x + G_T(x,\cdot) - \left[K - c \cdot x' + G_T(S_T^n,\cdot) \right] \right| = \left| -c(x-x') + G_T(x,\cdot) - G_T(s_T^n,\cdot) \right|$$

The last equality follows from the fact that $G_T(s_T^n,\cdot) = K + G_T(S_T^n,\cdot)$ by the K-convexity of G and the definitions of s_T^n and S_T^n. Thus,

$$\left| V_T(x,\cdot) - V_T(x',\cdot) \right| = \left| -c(x-x') + G_T(x,\cdot) - G_T(s_T^n,\cdot) \right| \le -c(x-x') + \left| G_T(x,\cdot) - G_T(s_T^n,\cdot) \right|$$

By the equicontinuity of the family $\{ G_T(x,\cdot), n = 1,\ldots \}$, there exists a $\delta_{\varepsilon/2} > 0$ such that $|x - x'| < \delta_{\varepsilon/2}$ implies $\left| G_T(x,\cdot) - G_T(x',\cdot) \right| < \varepsilon/2$. Then, $|x - s_T^n| \le |x - x'| < \delta_{\varepsilon/2}$ implies $\left| G_T(x,\cdot) - G_T(s_T^n,\cdot) \right| < \varepsilon/2$.

Define $\delta'_{\varepsilon/2} = \varepsilon/2c$. From this we get $c(x-x') < \varepsilon/2$ whenever $|x - x'| < \delta'_{\varepsilon/2}$. Choose $\delta_\varepsilon = \mathrm{Min}(\delta_{\varepsilon/2}, \delta'_{\varepsilon/2})$. For this δ_ε it can be seen that $|x - x'| < \delta_\varepsilon$ implies $\left| V_T(x,\cdot) - V_T(x',\cdot) \right| < \varepsilon$.

Case 3: $z = x, \quad z' = x'$.

In this case, $|z - z'| = |x - x'|$ and

$$\left| V_T(x,\cdot) - V_T(x',\cdot) \right| = \left| -c \cdot x + G_T(x,\cdot) + c \cdot x' - G_T(x',\cdot) \right| \le c(x-x') + \left| G_T(x,\cdot) - G_T(x',\cdot) \right|.$$

The equicontinuity of the family $\{ V_T(x,\alpha_n), n = 1,\ldots \}$ at x follows directly from the fact that $\{ G_T(x,\alpha_n), n = 1,\ldots \}$ is an equicontinuous family at x. The proof is completed by noting that $\{ V_0(x,\alpha_n) \equiv 0, n = 1,\ldots \}$ is an equicontinuous family at x. //

Proof of Theorem 3.3. (i) First consider the case of finite T. Let x be in some finite interval $[0, \bar{x}]$. Consider the sequence $\{ V_T(x,\alpha_n) \}_{n=1}^\infty$. Since Ω is compact, $V_T(x,\alpha_n)$ is uniformly bounded on $[0, \bar{x}]$ (i.e., there exists an M independent of n such that $V_T(x,\alpha_n) \le M$). The Ascoli–Arzela theorem (Royden [1988], Theorem 7.40) implies that there exists a subsequence α_{nk} such that $V_T(x,\alpha_{nk})$ converges to a continuous function $V_T(x, F^*)$ where the convergence is uniform on each compact subset of $[0, \bar{x}]$. Since α_{nk} converges to F^* for all possible subsequences, nk can be taken to be the entire sequence, n. It remains to show

that $V_T(x, F^*) = V_T(x)$. We use the following preliminary lemma adapted from Hinderer [1970, Lemma 3.3].

Lemma A.1. *If $u(y)$ and $v(y)$ are continuous functions bounded below then*

$$\left| \inf_y u(y) - \inf_y v(y) \right| \le \sup_y \left| u(y) - v(y) \right|.$$

The proof of theorem 3.3 proceeds by induction. Suppose the theorem holds for $T-1$ and consider $\lim_{n \to \infty} V_T(x, \alpha_n)$. By Lemma A.1,

$$\left| V_T^*(x) - V_T(x, \alpha_n) \right| \le \sup_u \left| \begin{array}{c} C(u) + \int \left[L(x+u, \omega) + \beta V_{T-1}^*(x+u-\omega) \, dF^*(\omega) \right] \\ - \left\{ C(u) + \int \left[L(x+u, \omega) + \beta V_{T-1}(x+u-\omega, \alpha_n + \delta_\omega) \, d\hat{F}_n(\omega) \right] \right\} \end{array} \right|$$

$$= \sup_z \left| \begin{array}{c} \left[\int L(z, \omega) \, dF^*(\omega) - \int L(z, \omega) \, d\hat{F}_n(\omega) \right] + \left[\beta \int V_{T-1}^*(z-\omega) \, dF^*(\omega) - \beta \int V_{T-1}^*(z-\omega) \, d\hat{F}_n(\omega) \right] \\ + \left[\beta \int V_{T-1}^*(z-\omega) \, d\hat{F}_n(\omega) - \beta \int V_{T-1}^*(z-\omega, \alpha_n + \delta_\omega) \, d\hat{F}_n(\omega) \right] \end{array} \right|$$

$$\le \sup_z \left| \int \left[L(z, \omega) + \beta V_{T-1}^*(z-\omega) \right] dF^*(\omega) - \int \left[L(z, \omega) + \beta V_{T-1}^*(z-\omega) \right] d\hat{F}_n(\omega) \right|$$

$$+ \sup_z \beta \left| \int \left[V_{T-1}^*(z-\omega) - V_{T-1}^*(z-\omega, \alpha_n + \delta_\omega) \right] d\hat{F}_n(\omega) \right|$$

$$\le \sup_z \left| \int \left[L(z, \omega) + \beta V_{T-1}^*(z-\omega) \right] dF^*(\omega) - \int \left[L(z, \omega) + \beta V_{T-1}^*(z-\omega) \right] d\hat{F}_n(\omega) \right|$$

$$+ \sup_{0 \le j \le \infty} \sup_z \beta \left| \int \left[V_{T-1}^*(z-\omega) - V_{T-1}^*(z-\omega, \alpha_n + \delta_\omega) \right] d\hat{F}_j(\omega) \right|$$

$$\text{(A.2.1)}$$

where $z = x + u$.

Since $L(z, \omega) + \beta V_{T-1}^*(z-\omega)$ is continuous in ω and $\hat{F}_n$ converges weakly to F^*, we have

$$\lim_{n \to \infty} \int \left[L(z, \omega) + \beta V_{T-1}^*(z-\omega) \right] d\hat{F}_n(\omega) = \int \left[L(z, \omega) + \beta V_{T-1}^*(z-\omega) \right] dF^*(\omega)$$

where the convergence is uniform on finite intervals. Thus, the lim and sup operators can be exchanged (Hinderer [1970, remark following Lemma 3.4]) to obtain

$$\lim_{n\to\infty}\; \sup_{z}\; \left| \int\left[L(z,\omega)+\beta\, V_{T-1}^{*}(z-\omega)\right]dF^{*}(\omega) - \int\left[L(z,\omega)+\beta\, V_{T-1}^{*}(z-\omega)\right]d\hat{F}_{n}(\omega)\right|$$

$$= \sup_{z}\; \lim_{n\to\infty}\; \left| \int\left[L(z,\omega)+\beta\, V_{T-1}^{*}(z-\omega)\right]dF^{*}(\omega) - \int\left[L(z,\omega)+\beta\, V_{T-1}^{*}(z-\omega)\right]d\hat{F}_{n}(\omega)\right|$$

$$= 0.$$

$$(A.2.2)$$

Since the cost function with $T-1$ periods to go is uniformly bounded for all n and $V_{T-1}(z-\omega,\alpha_{n}+\delta_{\omega}) \to V_{T-1}^{*}(z-\omega)$ by the induction hypothesis, the dominated convergence theorem implies $\int V_{T-1}(z-\omega,\alpha_{n}+\delta_{\omega})\, d\hat{F}_{j}(\omega) \to \int V_{T-1}^{*}(z-\omega)\, d\hat{F}_{j}(\omega)$. It then follows that

$$\lim_{n\to\infty}\; \sup_{0\le j\le\infty}\; \sup_{z}\; \beta\left| \int\left[V_{T-1}^{*}(z-\omega)-V_{T-1}^{*}(z-\omega,\alpha_{n}+\delta_{\omega})\right]d\hat{F}_{j}(\omega)\right|$$

$$= \sup_{0\le j\le\infty}\; \sup_{z}\; \lim_{n\to\infty}\; \beta\left| \int\left[V_{T-1}^{*}(z-\omega)-V_{T-1}^{*}(z-\omega,\alpha_{n}+\delta_{\omega})\right]d\hat{F}_{j}(\omega)\right| = 0.$$

Combining this with (A.2.2) shows that (A.2.1) converges to zero as $n\to\infty$ which proves that $\lim_{n\to\infty} V_{T}(x,\alpha_{n}) = V_{T}^{*}(x)$. Since the result holds trivially for $T=0$, the induction argument is complete and the theorem holds for finite T.

Now consider the infinite horizon case. Standard arguments imply:
(1) $V_{\infty}(x,\alpha_{n})$ and $V_{\infty}^{*}(x)$ are continuous in x, and
(2) $\lim_{T\to\infty} V_{T}(x,\alpha_{n}) = V_{\infty}(x,\alpha_{n})$ and $\lim_{T\to\infty} V_{T}^{*}(x) = V_{\infty}^{*}(x)$, where the convergence is uniform in each case.

This implies that for all x and $\varepsilon>0$ there exists a T_{ε} such that $\left| V_{T}(x,\alpha_{n})-V_{\infty}(x,\alpha_{n})\right|< \varepsilon$ and $\left| V_{T}^{*}(x)-V_{\infty}^{*}(x)\right|< \varepsilon$ for all $T>T_{\varepsilon}$. Combining these two inequalities gives $\left| V_{T}(x,\alpha_{n})-V_{\infty}(x,\alpha_{n})-V_{T}^{*}(x)+V_{\infty}^{*}(x)\right|< 2\varepsilon$ for all $T>T_{\varepsilon}$.

Letting $n\to\infty$ implies $\lim_{n\to\infty}\left| V_{\infty}^{*}(x)-V_{\infty}(x,\alpha_{n})\right|< 2\varepsilon$, since $\lim_{n\to\infty} V_{T}(x,\alpha_{n}) = V_{T}^{*}(x)$. The proof follows immediately.

(ii) The proof is similar to the proof of Theorem 3.8 in Stokey, Lucas and Prescott [1989]. Let

$$W_{T}^{*}(x,u) = C(u) + \int\left\{L(x+u,\omega)+\beta\, V_{T}^{*}(x+u-\omega)\right\}dF^{*}(\omega),\; \text{and}$$

$$W_{T}^{n}(x,u) = C(u) + \int\left\{L(x+u,\omega)+\beta\, V_{T-1}(x+u-\omega,\alpha_{n}+\delta_{\omega})\right\}d\hat{F}_{n}(\omega).$$

Note that $V_T^*(x) = \underset{u}{Max}\, W_T^*(x,u)$ and $V_T(x,\alpha_n) = \underset{u}{Max}\, W_T^n(x,u)$. Part (i) of the theorem implies that $W_T^n(x,u)$ converges to $W_T^*(x,u)$ where the convergence is uniform on each compact subset $[0,\bar{x}] \times [0,\bar{u}]$. Let $u_T^*(x)$ be the optimal order from x for the inventory problem with known demand distribution, F^*. By the principle of optimality, we have

$$
\begin{aligned}
0 &\le W_T^*\big(x,u_T(x,\alpha_n)\big) - V_T^*(x) \\
&\le W_T^*\big(x,u_T(x,\alpha_n)\big) - W_T^*\big(x,u_T^*(x)\big) + W_T^n\big(x,u_T^*(x)\big) - W_T^n\big(x,u_T(x,\alpha_n)\big) \\
&\le 2 \sup\left\{ \left| W_T^*\big(x,u_T(x,\alpha_n)\big) - W_T^n\big(x,u_T(x,\alpha_n)\big) \right|, \left| W_T^*\big(x,u_T^*(x)\big) - W_T^n\big(x,u_T^*(x)\big) \right| \right\}
\end{aligned}
$$

for all x. Since W_T^n converges to W_T^* uniformly, it follows that for all $\varepsilon > 0$ there exists an $N(t,\varepsilon)$ such that $0 \le W_T^*\big(x,u_T(x,\alpha_n)\big) - V_T^*(x) < \varepsilon$ for all $n > N(T,\varepsilon)$. //

Proof of Lemma 4.1. The one period-to-go value function is given by

$$
V_1(x,\alpha_n) = \inf_{u\ge0} \left[c\cdot u + \int L(x+u,\omega)\,d\hat{F}_n(\omega) \right] = \inf_{u\ge0} \left[G_1(x+u,\alpha_n) - c\cdot x \right]
$$

where the function G_1 is defined in (3.1). Under the Dirichlet process, G_1 can be rewritten as

$$
\begin{aligned}
G_1(z,\alpha_n) &= c\cdot z + \lambda_n \left[h\cdot \int_0^z (z-\omega)\,d\hat{F}(\omega) + p\cdot \int_z^\infty (\omega-z)\,d\hat{F}(\omega) \right] \\
&\quad + (1-\lambda_n)\left[h\cdot\frac{1}{n}\sum_{i=1}^n Max[z-\omega_i,0] + p\cdot\frac{1}{n}\sum_{i=1}^n Max[\omega_i-z,0] \right] \\
&= c\cdot z + \lambda_n \left[h\cdot \int_0^z (z-\omega)\,d\hat{F}(\omega) + p\cdot \int_z^\infty (\omega-z)\,d\hat{F}(\omega) \right] \\
&\quad + (1-\lambda_n)\left[h\cdot\frac{1}{n}\sum_{i=1}^n (z-\omega_i)1_{(z\ge\omega_i)} + p\cdot\frac{1}{n}\sum_{i=1}^n (\omega_i-z)1_{(z<\omega_i)} \right].
\end{aligned}
$$

As noted earlier, G_1 is continuous and convex in z. The derivative of G_1 with respect to z is

$$
\begin{aligned}
G_1'(z,\alpha_n) &= c + \lambda_n\left[h\cdot\hat{F}(z) - p\cdot(1-\hat{F}(z)) \right] + (1-\lambda_n)\left[h\cdot\frac{1}{n}\sum_{i=1}^n 1_{(z\ge\omega_i)} + p\cdot\frac{1}{n}\sum_{i=1}^n 1_{(z<\omega_i)} \right] \\
&= c + h\cdot\hat{F}_n(z) - p\cdot(1-\hat{F}_n(z)) = c - p + (h+p)\cdot\hat{F}_n(z)
\end{aligned}
$$

This derivative exists except possibly at $\omega_1,\ldots,\omega_n$ and points of discontinuity of the prior; however, at these points the left-hand and right-hand derivatives exist and are bounded. (For simplicity we assume in the sequel that the prior measure has a density.)

It is clear that $G_1'(z,\alpha_n)$ is increasing in z, and given $p > c$ we have $\lim_{|z|\to\infty} G_1(z,\alpha_n) = \infty$. Hence, a minimum of $G_1(z,\alpha_n)$ exists. Let S_1^n be the smallest value minimizing $G_1(z,\alpha_n)$. The value function and its derivative are given by

$$V_1(x,\alpha_n) = \begin{cases} -c\cdot x + G_1(S_1^n,\alpha_n) & x < S_1^n \\ -c\cdot x + G_1(x,\alpha_n) & x \geq S_1^n \end{cases}$$

and

$$V_1'(x,\alpha_n) = \begin{cases} -c & x < S_1^n \\ -c + G_1'(x,\alpha_n) & x \geq S_1^n \end{cases}$$

with the derivative being well-defined except possibly at $\omega_1,\ldots,\omega_n$ and S_1^n. The continuity and convexity of $G_1(x,\alpha_n)$ imply that $V_1(x,\alpha_n)$ is continuous and convex in x. Furthermore, $V_1'(x,\alpha_n)$ is non-decreasing in x (because if $G_1'(x,\alpha_n)$ is well-defined then it is non-negative for $x > S_1^n$ by the convexity of G_1).

The T-period planning horizon value function is defined by

$$V_T(x,\alpha_n) = \inf_{u\geq 0}\left[c\cdot u + \int\{L(x+u,\omega) + \beta V_{T-1}(x+u-\omega,\alpha_n+\delta_\omega)\}\,d\hat{F}_n(\omega)\right]$$
$$= \inf_{u\geq 0}\left[-c\cdot x + G_T(x+u,\alpha_n)\right]$$

where $G_T(z,\alpha_n)$ is defined in (3.2). Using standard arguments, it can be established that: (i) $G_T(z,\alpha_n)$ is continuous and convex in z, (ii) $\lim_{|z|\to\infty} G_T(z,\alpha_n) = \infty$, (iii) $G_1'(z,\alpha_n)$ is increasing in z, and hence that

$$V_T(x,\alpha_n) = \begin{cases} -c\cdot x + G_T(S_T^n,\alpha_n) & x < S_T^n \\ -c\cdot x + G_T(x,\alpha_n) & x \geq S_T^n \end{cases}$$

where S_T^n is the smallest value minimizing $G_T(z,\alpha_n)$. Next,

$$V_T'(x,\alpha_n) = \begin{cases} -c, & x < S_T^n \\ -c + G_T'(x,\alpha_n), & x \geq S_T^n \end{cases}$$

is well defined except possibly at finite points of discontinuity. Clearly $V_T'(x,\alpha_n)$ is non-decreasing in x. Finally, $V_T(x,\alpha_n)$ is continuous and convex in x.

It is clear from the expression for $V_1'(x,\alpha_n)$ that the result is true for $T = 1$. We now assume it is true for $T-1$ and show it is true for T. Consider

$$\begin{aligned}
G_T'(x,\alpha_n^2) &= c - p + (h+p)\hat{F}_n^2(x) + \beta \int V_{T-1}'(x-\omega, \alpha_n^2 + \delta_\omega)\,d\hat{F}_n^2(\omega) \\
&= c - p + \beta \int \left\{ V_{T-1}'(x-\omega, \alpha_n^2 + \delta_\omega) + (h+p)1_{(\omega \leq x)} \right\}\,d\hat{F}_n^2(\omega) \\
&\geq c - p + \beta \int \left\{ V_{T-1}'(x-\omega, \alpha_n^2 + \delta_\omega) + (h+p)1_{(\omega \leq x)} \right\}\,d\hat{F}_n^1(\omega) \\
&\geq c - p + \beta \int \left\{ V_{T-1}'(x-\omega, \alpha_n^1 + \delta_\omega) + (h+p)1_{(\omega \leq x)} \right\}\,d\hat{F}_n^1(\omega) \\
&= G_T'(x,\alpha_n^1)
\end{aligned}$$

where the first inequality follows from the fact that $\hat{F}_n^1$ stochastically dominates $\hat{F}_n^2$ and the integrand is decreasing in ω. The integrand decreases in ω because
(i) $V_{T-1}'(z,\cdot)$ increases in z, and
(ii) $V_{T-1}'(z, \alpha_n^2 + \delta_\omega) \leq V_{T-1}'(z, \alpha_n^2 + \delta_{\omega'})$ for $\omega < \omega'$, by the induction assumption and the fact that $\alpha_n^2 + \delta_{\omega'}$ stochastically dominates $\alpha_n^2 + \delta_\omega$ for $\omega < \omega'$.
The second inequality follows from the induction assumption since $\alpha_n^1 + \delta_\omega$ stochastically dominates $\alpha_n^2 + \delta_\omega$. The critical number $S_T(\alpha_n^1)$ satisfies $G_T'(x,\alpha_n^1) = 0$ if the derivative exists, or it is the smallest ω for which the right hand derivative of $G_T(x,\alpha_n^1)$ is greater than or equal to zero. In either case, the fact that $G_T'(x,\alpha_n^1) \leq G_T'(x,\alpha_n^2)$ enables us to conclude that $S_T(\alpha_n^1) \geq S_T(\alpha_n^2)$ (since G_T' is increasing in its first argument).

To complete the proof of the lemma, consider the following three possible cases:

(i) $x \leq S_T(\alpha_n^2)$ which implies $V_{T-1}'(x,\alpha_n^2) - V_{T-1}'(x,\alpha_n^1) = -c + c = 0$,

(ii) $S_T(\alpha_n^2) \leq x \leq S_T(\alpha_n^1)$ which implies
$$V_{T-1}'(x,\alpha_n^2) - V_{T-1}'(x,\alpha_n^1) = -c + G_T'(x,\alpha_n^2) + c = G_T'(x,\alpha_n^2) \geq 0,$$

(iii) $x \geq S_T\left(\alpha_n^1\right)$ which implies

$$V'_{T-1}\left(x,\alpha_n^2\right) - V'_{T-1}\left(x,\alpha_n^1\right) = G'_T\left(x,\alpha_n^2\right) - G'_T\left(x,\alpha_n^1\right) \geq 0.$$

If the derivative is not well-defined at a point x, then the above inequalities can be established for the right hand and left hand derivatives. Hence, the lemma is proved. //

Proof of Theorem 4.2. The proof follows immediately from Lemma 4.1. //

Proof of Theorem 4.3. Note that $S_T\left(\alpha_n\right)$ satisfies $G'_T\left(z,\alpha_n\right) = 0$, or if the derivative is not well defined it is the smallest z such that the right hand derivative of $G_T\left(z,\alpha_n\right)$ is greater than or equal to zero. A similar relation characterizes $S_T\left(\alpha_n + \delta_{\omega \leq Min[\omega_1,...,\omega_n]}\right)$. Using arguments similar to those in the proof of Lemma 4.1, $G'_T\left(z, \alpha_n\right) \leq G'_T\left(z, \alpha_n + \delta_{\omega \leq Min[\omega_1,...,\omega_n]}\right)$ because $\hat{F}_n$ first-order stochastically dominates $\hat{F}_{n+1} \mid \omega_{n+1} \leq Min[\omega_1,...,\omega_n]$. The result follows because $G'_T\left(z,\cdot\right)$ is non-decreasing in z.

Remark. Theorems 4.2 and 4.3 are not true when there is a fixed ordering cost.

VII. REFERENCES

Arrow, K. J., T. Harris and J. Marschak, 1951, "Optimal Inventory Policy," *Econometrica*, 19, pp. 250-272.

Azoury, K. S., 1985, "Bayes Solution to Dynamic Inventory Models Under Unknown Demand Distribution," *Management Science*, 31, pp. 1150-1160.

Azoury, K. S. and B. L. Miller, 1984, "A Comparison of the Optimal Ordering Levels of Bayesian and Non-Bayesian Inventory Models," *Management Science*, 30, pp. 993-1003.

Bikhchandani, S. and S. Sharma, 1996, "Optimal Search with Learning," *Journal of Economic Dynamics and Control*, 20, pp. 333-359.

Billingsley, P., 1968, *Convergence of Probability Measures*, (New York: John Wiley).

Blackwell, D. and J. B. MacQueen, 1973, "Ferguson Distributions via Pólya Urn Schemes," *Annals of Statistics*, 1, pp. 353-355.

Boylan, E.S., 1969, "Stability Theorems for Solutions to the Optimal Inventory Equation," *Journal of Applied Probability*, 6, pp. 211-217.

Christensen, R., 1986, "Finite Stopping in Sequential Sampling Without Recall from a Dirichlet Process," *Annals of Statistics*, 14, pp. 275-282.

Clayton, M. K. and D. A. Berry, 1985, "Bayesian Nonparametric Bandits," *Annals of Statistics*, 13, pp. 1523-1534.

Degroot, M. H., 1970, *Optimal Statistical Decisions*, (New York: McGraw-Hill).

Dutta, P., M. Majumdar and R. Sundaram, 1994, "On the Parametric Continuity of Dynamic Programming Problems," *Journal of Economic Dynamics and Control*, 18, pp. 1069-1092.

Dvoretzky, A., J. Kiefer and J. Wolfowitz, 1952, "The Inventory Problem: II. Case of Unknown Distributions of Demand," *Econometrica*, 20, pp. 450-466.

Fabius, J., 1964, "Asymptotic Behavior of Bayes' Estimates, *Annals of Mathematical Statistics*, 35, pp. 846-856.

Ferguson, T. S., 1973, "A Bayesian Analysis of Some Nonparametric Problems," *Annals of Statistics*, 1, pp. 209-230.

Ferguson, T. S., 1974, "Prior Distributions on Spaces of Probability Measure," *Annals of Statistics*, 2, pp. 615-629.

Hinderer, K., 1970, *Foundations of Nonstationary Dynamic Programming with Discrete Time Parameter*, (Berlin: Springer-Verlag).

Iglehart, D. L., 1963, "Optimality of (s,S) Policies in the Infinite Horizon Inventory Problem," *Management Science*, 9, pp. 259-267.

Iglehart, D. L., 1964, "The Dynamic Inventory Problem with Unknown Demand Distribution," *Management Science*, 10, pp. 429-440.

Karlin, S., 1960, "Dynamic Inventory Policy with Varying Stochastic Demands," *Management Science*, 6, pp. 231-258.

Lovejoy, W. S., 1990, "Myopic Policies for Some Inventory Models with Uncertain Demand Distributions," *Management Science*, 36, pp. 724-738.

Lovejoy, W. S., 1993, "Suboptimal Policies, with Bounds, for Parameter Adaptive Decision Processes," *Operations Research*, 41, pp. 583-599.

Porteus, E. L., 1990, "Stochastic Inventory Theory," in *Handbooks in Operations Research and Management Science, vol 2*, ed. by D. P. Heyman and M. J. Sobel, (Amsterdam: Elsevier).

Rothschild, M., 1974, "Searching for the Lowest Price When the Distribution of Prices is Unknown," *Journal of Political Economy*, 82, pp. 689-711.

Royden, H. L., 1988, *Real Analysis* (3rd edition), (New York: Macmillan).

Scarf, H., 1959a, "Bayes Solutions to the Statistical Inventory Problem," *Annals of Mathematical Statistics*, 30, pp. 490-508.

Scarf, H., 1960a, "Some Remarks on Bayes Solutions to the Inventory Problem," *Naval Research Logistic Quarterly*, 7, pp. 591-596.

Scarf, H., 1960b, "On the Optimality of (s,S) Policies in the Dynamic Inventory Problem," in *Mathematical Methods in the Social Sciences,1959. Proceedings of the First Stanford Symposium. Stanford Mathematical Studies in the Social Sciences.* ed. by K. J. Arrow, S. Karlin, and P. Suppes, (Stanford: Stanford University Press).

Schal, M., 1975, "Conditions for Optimality in Dynamic Programming and for the Limit of n-Stage Optimal Policies to Be Optimal," *Z. Wahrschein-lichkeitstheorie und Verw. gebiete*, 32, pp. 179-196.

Stokey, N., R. Lucas and E. Prescott, 1989, *Recursive Methods in Economic Dynamics,* (Cambridge: Harvard University Press).

Talman, G., 1992, "Search from an Unknown Distribution: An Explicit Solution," *Journal of Economic Theory*, 57, pp. 141-157.