ON THE INCORPORATION OF PARAMETER UNCERTAINTY FOR INVENTORY MANAGEMENT

David F. Muñoz[(a)], David G. Muñoz[(b)]

[(a)] Departamento de Ingeniería Industrial y Operaciones, Instituto Tecnológico Autónomo de México, Río Hondo #1, Mexico City, Mexico
[(b)] Research and Development, AOL Advertising, 395 Page Mill Road, Palo Alto, California

[(a)]davidm@itam.mx, [(b)]davidmm82@gmail.com

ABSTRACT
The main purpose of this paper is to discuss how a Bayesian framework is appropriate to incorporate the uncertainty on the parameters of the model that is used for demand forecasting. We first present a general Bayesian framework that allows us to consider a complex model for forecasting. Using this framework we specialize (for simplicity) in the continuous-review (Q,R) system to illustrate how the main performance measures that are required for inventory management can be estimated from the output of simulation experiments. We discuss the use of sampling from the posterior distribution (SPD) and show that, under suitable regularity conditions, the estimators obtained from SPD satisfy a corresponding Central Limit Theorem, so that they are consistent, and the accuracy of each estimator can be assessed by computing an asymptotically valid halfwidth from the output of the simulation experiments.

Keywords: inventory simulation, reorder points, parameter uncertainty, output analysis

1. INTRODUCTION
Most of the proposed techniques to compute service levels and reorder points assume that the parameters of the model that is used for demand forecasting are known with certainty (see, e.g., Nahmias 2008). However in practice, parameters are estimated from available information (data and/or expert judgment), and there exists a certain degree of uncertainty in the value of these parameters. In this article, we use a Bayesian framework that allows us to incorporate parameter uncertainty that is induced from the estimation procedure. This framework is particularly useful when using a complex model for demand forecasting, in the sense that analytical expressions to obtain service levels and/or reorder points may not be available, so that the application of estimation procedures based on stochastic simulation is recommended.

Although it is not our intention to review the abundant literature on inventory simulation (see Jahangirian et al. 2010 for a recent survey), we mention that simulation has been extensively used to analyze inventory policies in a supply chain (see Tako and Robinson 2012 for a recent survey). Simulation has also been used as a tool to compare different forecasting procedures for inventory management (see, e.g., Syntetos et al. 2009, Bartezzaghi et al. 1999). Furthermore, simulation-based methodologies such as importance sampling (see, e.g., Glasserman and Liu 1996) and simulation optimization (see, e.g., Fu 2002) have been applied to inventory models. However, to the best of our knowledge, parameter uncertainty has not been considered in the related bibliography.

The Bayesian approach described in this article allows us to incorporate parameter uncertainty through a probability distribution, and it is in the spirit of the approach described, e.g., in Chick (2001) and Bermúdez et al. (2010), where the authors discuss the incorporation of parameter uncertainty in a forecasting model using stochastic simulation. In this article, we show how this approach can be extended to the estimation of Bayesian reorder points, and our main contribution is the development of a Central Limit Theorem (CLT) for each of the proposed point estimators. As is well known in the simulation literature, a CLT for an estimator allows us to construct an asymptotic confidence interval (ACI) to assess the accuracy of the point estimator obtained using simulation. We extend the results of Muñoz and Galindo (2010) to include the estimation of three measures of service level and corresponding reorder points.

The remainder of this article is organized as follows. In Section 2 we present a Bayesian framework under which performance measures for inventory management can be precisely defined, we illustrate the derivation of performance measures under this framework by considering the continuous-review (Q,R) system. In Section 3 we discuss how SPD can be used to estimate the performance measures defined in Section 2 as well as to assess the accuracy of the point estimators. Finally, in Section 4 we provide a simple example to illustrate our Bayesian Framework.

2. THEORETICAL FRAMEWORK

2.1. Notation and Model Assumptions
We assume that $W = g\left(Y(s), 0 \le s \le T; \Theta\right)$ is the demand for an item, where $Y = \{Y(s), s \ge 0; \Theta\}$ is a

stochastic process (possibly multivariate), T is a stopping time that represents the planning horizon, and Θ is a vector of parameters. This notation is general enough to include most forecasting models that are used in practice (e.g., ARIMA and regressions), and in particular, a discrete-time stochastic process can be incorporated into this framework by setting $Y(s) = Y(\lfloor s \rfloor)$ (where $\lfloor s \rfloor$ denotes the integer part of s).

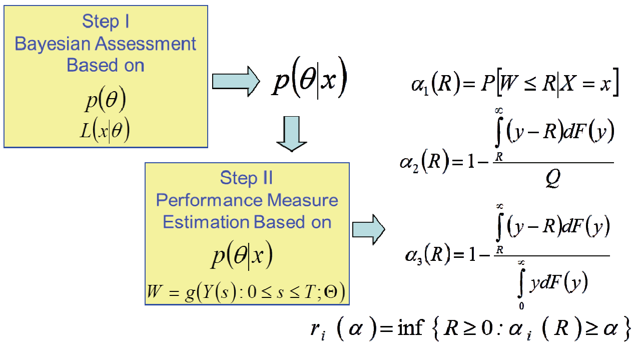

Figure 1: Main Steps to Compute Performance Measures for Inventory Management

A Bayesian approach for inventory management is summarized in two steps, as illustrated in Figure 1. In the first step parameter uncertainty is assessed by constructing the posterior density $p(\theta|x)$ from the available data x and a prior density $p(\theta)$. In the second step $W = g(Y(s), 0 \le s \le T; \Theta)$ and the posterior density $p(\theta|x)$ are used to compute the performance measures that are required for inventory management.
Input data on the model parameters is available through a vector of observations $x \in \Re^d$ that satisfies a likelihood function $L(x|\theta)$. If $p(\theta)$ is a prior density function for the vector of parameters Θ, then the posterior (given the data x) density function of Θ becomes

$$p(\theta|x) = \frac{p(\theta)L(x|\theta)}{\int_{S_0} p(\theta)L(x|\theta)d\theta}, \qquad (1)$$

for $x \in \Re^d$ and $\theta \in S_0$.

Note that we can consider an input $x \in \Re^d$ with correlated data, and for the special case where $x = (x_1, \ldots, x_n)$ is a set of observations of a random sample $X = (X_1, \ldots, X_n)$ from a density function $f(y|\theta)$, the likelihood function takes the form of

$$L(x|\theta) = f(x_1|\theta) f(x_2|\theta) \ldots f(x_n|\theta). \qquad (2)$$

The prior density $p(\theta)$ reflects the initial uncertainty on the vector of parameters Θ, and there are essentially two points of view to proposing a prior density $p(\theta)$. The first approach consists of using a non-informative prior, which is appropriate when we wish to consider a prior density that does not "favor" any possible value of Θ over others. This can be considered as an "objective" point of view (for a discussion on this subject see, e.g., Berger et al. 2008). The second approach is a "subjective" point of view, and consists in the establishment of a prior density based on expert judgment.

2.2. Performance Measures for Inventory Management

Although our previous notation can be useful to consider different inventory control policies, multi-item and/or multi-period systems, a definition of the appropriate performance measures for inventory management may be problem-dependent, and this is why we restrict our discussion to a single-item, single-period inventory system subject to a continuous-review (Q, R) policy. We set $T = T^* + L^*$, where T^* is the time at which the inventory level reaches the reorder point R, and L^* is the (possibly random) lead time for an order. A suitable definition for the output process is $Y(s) =$ cumulative demand of the single item on the interval $[0, s]$, so that $W = Y(T) + Y(T^*)$ is the demand during the lead time.

In general, a forecast for demand W is completely defined by its cumulative distribution function (c.d.f.) $F(w) = P[W \le w | X = x]$, which allows us to define the following important performance measures for inventory management. The expected demand

$$\mu \overset{def}{=} E[W | X = x] = \int_0^\infty ydF(y) \qquad (3)$$

is usually regarded as the point forecast. We assume that demand W is non-negative (i.e., $F(w) = 0$ for $w < 0$). Note that, although F or μ might depend on x, in order to simplify the notation, we denote (on purpose) F or μ not depending on x, and remark that the results of Section 3 can also be applied to the case where parameter uncertainty is not considered.

A performance measure that is of practical importance in inventory management is the probability of no stock-out

$$\alpha_1(R) \overset{def}{=} P[W \le R | X = x] = F(R), \qquad (4)$$

where $R \ge 0$. Thus, we say that the type-1 service level (T1SL) corresponding to a reorder point $R \ge 0$ is $100\alpha_1(R)\%$, and given $0 < \alpha < 1$, a reorder point for a $100\alpha\%$ T1SL is defined as

$$r_1(\alpha) \overset{def}{=} \inf\{R \ge 0 : \alpha_1(R) \ge \alpha\}, \qquad (5)$$

where $\alpha_1(R)$ is defined in (4).

Another measure of service is the proportion of demands that are met from stock,

$$\alpha_2(R) \overset{def}{=} 1 - \frac{\int_R^\infty (y-R)dF(y)}{Q}, \tag{6}$$

where $R \geq 0$, and $Q > 0$. We say that the type-2 service level (T2SL) corresponding to a reorder point $R \geq 0$ is $100\alpha_2(R)\%$, and given $0 < \alpha < 1$, a reorder point for a $100\alpha\%$ T2SL is defined as

$$r_2(\alpha) \overset{def}{=} \inf\{R \geq 0 : \alpha_2(R) \geq \alpha\}, \tag{7}$$

where $\alpha_2(R)$ is defined in (6). According to Nahmias (2008), the term "fill rate" is often used to describe T2SL, and is generally what most managers mean by service.

Finally, note that T1SL considers the probability of stock-out only during the lead time, so that we might also consider a measure of fill rate during the lead time,

$$\alpha_3(R) \overset{def}{=} 1 - \frac{\int_R^\infty (y-R)dF(y)}{\mu}, \tag{8}$$

where $\mu > 0$ is defined in (3). Similarly, we say that the type-3 service level (T3SL) corresponding to a reorder point $R \geq 0$ is $100\alpha_3(R)\%$, and given $0 < \alpha < 1$, a reorder point for a $100\alpha\%$ T3SL is defined as

$$r_3(\alpha) \overset{def}{=} \inf\{R \geq 0 : \alpha_3(R) \geq \alpha\}, \tag{9}$$

where $\alpha_2(R)$ is defined in (8).

When analytical expressions for the performance measures defined in this section cannot be obtained (or they are too complicated), simulation can be applied to estimate these parameters, as we explain in the next Section.

3. ESTIMATION USING SIMULATION

As illustrated in Figure 2, under the SPD algorithm we first sample from the posterior density $p(\theta|x)$ to obtain independent and identically distributed (i.i.d.) observations of the uncertain parameter Θ (given the data x), and then we simulate demand W to estimate μ, $\alpha_i(R)$ and $r_i(\alpha)$, by $\hat{\mu}$, $\hat{\alpha}_i(R)$ and $\hat{r}_i(\alpha)$, $i = 1,2,3$, respectively. The algorithm of Figure 2 is based on the algorithm proposed in Chick (2001) for the estimation of μ, and we also show how to produce consistent estimators for $\alpha_i(R)$ and $r_i(\alpha)$, $i = 1,2,3$. Furthermore, each of the point estimators defined in Figure 2 satisfies a corresponding CLT, as we explain below.

For the sake of completeness, we first show, using the case of the estimation of μ, how an ACI is obtained from a CLT. As is well known from the standard CLT, when $E\left[W_1^2 | X = x\right] < \infty$ we have

$$\frac{m^{1/2}(\hat{\mu} - \mu)}{\sigma_W} \Rightarrow N(0,1), \tag{10}$$

where "$\Rightarrow$" denotes weak converge (as $m \to \infty$), $N(0,1)$ is a standard normal distribution, and σ_W^2 is the variance of W_1 (given the data x, i.e., $\sigma_W^2 = E\left[\sigma_1^2 | X = x\right] - \mu^2$. Also, from a Weak Law of Large Numbers we also know that $S_W / \sigma_W \Rightarrow 1$, where

$$S_W = (m-1)^{-1/2} \sqrt{\sum_{j=1}^m W_j^2 - m^{-1}\left(\sum_{j=1}^m W_j\right)^2}$$

is the sample standard deviation, so that it follows from (10) and a converging together argument that $\frac{m^{1/2}(\hat{\mu} - \mu)}{S_W} \Rightarrow N(0,1)$, and the interval

$$I_m = \left[\hat{\mu} - z_\beta m^{-1/2} S_W, \hat{\mu} + z_\beta m^{-1/2} S_W\right]$$

tends (as $m \to \infty$) to cover the parameter μ with probability $(1-\beta)$, where $0 < \beta < 1$ is a given constant and $P\left[N(0,1) < z_\beta\right] = 1 - \beta/2$. The interval I_m is called a $100(1-\beta)\%$ ACI for μ, and the corresponding halfwidth

$$H_{\hat{\mu}} = z_\beta m^{-1/2} S_W, \tag{11}$$

is used in the simulation literature to assess the accuracy of $\hat{\mu}$ as an estimator of parameter μ. Similarly, $\alpha_1(R) = E[V_{11}|X = x]$, and $\alpha_2(R) = E[V_{21}|X = x]$, where $V_{11} = I[W_1 \leq R]$ and $V_{21} = 1 - Q^{-1}(W_1 - R)I[W_1 > R]$, so that CLT's for $\hat{\alpha}_1(R)$ and $\hat{\alpha}_2(R)$ are easily obtained from the standard CLT, and the halfwidths corresponding to a $100(1-\beta)\%$ ACI for $\alpha_i(R)$, $i = 1,2$ are given by

$$H_{\hat{\alpha}_i(R)} = z_\beta m^{-1/2} S_i, \tag{12}$$

where $S_i = (m-1)^{-1/2}\sqrt{\sum_{j=1}^m V_{ij}^2 - m^{-1}\left(\sum_{j=1}^m V_{ij}\right)^2}$, $V_{1j} = I\left[W_j \leq R\right]$, $V_{2j} = 1 - Q^{-1}\left(W_j - R\right)I\left[W_j > R\right]$, for $i = 1,2$, $j = 1,\ldots,m$, and the W_j's are defined in Figure 2.

Finally, note that $\alpha_3(R) = 1 - \mu_R / \mu$, where $\mu_R = E\left[\max\{0, W_1 - R\} | X = x\right]$, so that a CLT for

$\hat{\alpha}_3(R)$ can be obtained by applying the Delta method (see, e.g., Lemma 1 of Muñoz and Glynn 1997), and the halfwidth corresponding to a $100(1-\beta)\%$ ACI for $\alpha_3(R)$ is given by

$$H_{\hat{\alpha}_3(R)} = z_\beta m^{-1/2} S_3 , \tag{13}$$

where $S_3 = \hat{\mu}^{-1}\sqrt{S_3^2 - 2(\hat{\mu}_3 - \hat{\mu})S_{WV_3} + (\hat{\mu}_3 - \hat{\mu})^2 S_W^2}$, $\hat{\mu}_3 = m^{-1}\sum_{j=1}^m V_{3j}$, $V_{3j} = \max\{0, W_j - R\}$, $j = 1,\ldots,m$, $S_3^2 = (m-1)^{-1}\left(\sum_{j=1}^m V_{3j}^2 - m\hat{\mu}_3^2\right)$, and

$$S_{WV_3} = (m-1)^{-1}\left(\sum_{j=1}^m W_j V_{3j} - m\hat{\mu}\hat{\mu}_3\right).$$

When W is discrete, the value of a reorder point can be investigated from the estimation of the corresponding service levels $\alpha_i(R)$ for different values of R , and thus we assume that W is continuous when discussing how to compute a halfwidth for a reorder point. In particular, for the estimation of $r_1(\alpha)$, we assume that F is differentiable at $r_1(\alpha)$ with $F'(r_1(\alpha)) > 0$, so that it follows from Bahadur's representation for quantiles (Bahadur 1966) that $\hat{r}_1(\alpha)$ satisfies a CLT, and a halfwidth corresponding to a $100(1-\beta)\%$ ACI for $r_1(\alpha)$ is given by

$$H_{\hat{r}_1(\alpha)} = \left(Z_{m_{11}} + Z_{m_{12}}\right)/2 , \tag{14}$$

where $m_{11} = \left\lfloor m\alpha - z_\beta [m\alpha(1-\alpha)]^{1/2} \right\rfloor$, $m_{12} = \left\lceil m\alpha + z_\beta [m\alpha(1-\alpha)]^{1/2} \right\rceil$, and the Z_i 's are defined in Figure 2. The validity of the ACI corresponding to (14) relies on a CLT for $\hat{r}_1(\alpha)$ and is established in Section 2.6.3 of [29]. We remark that the asymptotic variance of the CLT for $\hat{r}_1(\alpha)$ depends on the density $F'(r_1(\alpha))$, and to avoid a density estimation we are using a halfwidth in the form of (14).

In order to establish a CLT for $\hat{r}_2(\alpha)$ and $\hat{r}_3(\alpha)$, we need to introduce some notation. For $R \geq 0$, set

$$\lambda_{21}(R) = (1-\alpha)Q - \int_R^\infty (y-R)dF(y),$$

$$\begin{aligned}\lambda_{22}(R) &= (1-\alpha)^2 Q^2 - 2Q(1-\alpha)\int_R^\infty (y-R)dF(y) \\ &\quad + \int_R^\infty (y-R)^2 dF(y),\end{aligned} \tag{15}$$

and

$$\lambda_{31}(R) = (1-\alpha)\mu - \int_R^\infty (y-R)dF(y),$$

$$\lambda_{32}(R) = (1-\alpha)^2 \int_0^R y^2 dF(y) + \int_R^\infty (R-\alpha y)^2 dF(y). \tag{16}$$

Proposition 1. Let us suppose that $\int_0^\infty y^2 dF(y) < \infty$ and $0 < \alpha < 1$. Then

(i) If $\int_0^\infty y^2 dF(y) > (1-\alpha)^2 Q^2$, $\lambda_{21}(R)$ is differentiable at $R = r_2(\alpha)$ and $\lambda_{22}(R)$ is continuous at $R = r_2(\alpha)$, we have
$m^{1/2}[\hat{r}_2(\alpha) - r_2(\alpha)] \Rightarrow \sigma_2 N(0,1)$,
where $\lambda_{21}(R)$ and $\lambda_{22}(R)$ are defined in (15), and
$$\sigma_2^2 = \frac{\int_{r_2(\alpha)}^\infty (y - r_2(\alpha))^2 dF(y) - (1-\alpha)^2 Q^2}{(1 - F(r_2(\alpha)))^2}.$$

(ii) If $\mu > 0$, $\lambda_{31}(R)$ is differentiable at $R = r_3(\alpha)$ and $\lambda_{33}(R)$ is continuous at $R = r_3(\alpha)$, we have
$m^{1/2}[\hat{r}_3(\alpha) - r_3(\alpha)] \Rightarrow \sigma_3 N(0,1)$,
where $\lambda_{31}(R)$ and $\lambda_{32}(R)$ are defined in (16), and
$$\sigma_3^2 = \frac{(1-\alpha)^2 \int_0^{r_3(\alpha)} y^2 dF(y) + \int_{r_3(\alpha)}^\infty (r_3(\alpha) - \alpha y)^2 dF(y)}{(1 - F(r_3(\alpha)))^2}.$$

For $j = 1$ to the number of replications m:

a. Generate (independently) a value θ_j by sampling from $p(\theta|x)$.

b. Run a simulation experiment with $\Theta = \theta_j$ to obtain an independent replication W_j of $W = g(Y(s), 0 \leq s \leq T; \Theta)$

End Loop

Compute $\hat{\mu} = \frac{1}{m}\sum_{j=1}^m W_j, \hat{\alpha}_1(R) = \frac{1}{m}\sum_{j=1}^m I[W_j \leq R]$,

$$\hat{\alpha}_2(R) = 1 - \frac{\sum_{j=1}^m (W_j - R)I[W_j > R]}{mQ}, \hat{\alpha}_3(R) = 1 - \frac{\sum_{j=1}^m (W_j - R)I[W_j > R]}{\sum_{j=1}^m W_j}.$$

Sort the W_j 's: $Z_1 \leq \ldots \leq Z_m$ and set $\hat{r}_1(\alpha) = Z_{k_1}, \hat{r}_2(\alpha) = Z_{k_2}, \hat{r}_3(\alpha) = Z_{k_3}$,

where $Z_0 = 0$, and

$$k_1 = \lceil m\alpha \rceil, k_2 = \min\left\{0 \leq j \leq m : \sum_{i=j}^m (Z_i - Z_j) \leq Qm(1-\alpha)\right\},$$

$$k_3 - \min\left\{0 \leq j \leq m : \sum_{i=j}^m (Z_i - Z_j) \leq (1-\alpha)\sum_{i=1}^m Z_i\right\}.$$

Figure 2: Estimation of Performance Measures Using SPD

Using Proposition 1 we can establish the following halfwidths, corresponding to a $100(1-\beta)\%$ ACI for $r_2(\alpha)$ and $r_3(\alpha)$, respectively,

$$H_{\hat{r}_2(\alpha)} = z_\beta m^{-1/2}\hat{\sigma}_2 \text{ and } H_{\hat{r}_3(\alpha)} = z_\beta m^{-1/2}\hat{\sigma}_3 , \tag{17}$$

where

$$\hat{\sigma}_2^2 = m(m-k_2)^{-2}\left[\sum_{j=k_2}^m (W_j - \hat{r}_2(\alpha))^2 - mQ^2(1-\alpha)^2\right],$$

$$\hat{\sigma}_3^2 = m(m-k_3)^{-2}\left[(1-\alpha)^2 \sum_{j=1}^{k_3-1} W_j^2 + \sum_{j=k_3}^m (\hat{r}_3(\alpha)W_j)^2\right],$$

k_2, k_3 and the W_j 's are defined in the algorithm of Figure 2.

4. AN ILLUSTRATIVE EXAMPLE

In order to illustrate our notation, we present a model that is inspired in the ideas of Silver (1965) and Croston (1972) to forecast intermittent demand. Suppose that the arrival of clients at a retailer occurs according to a Poisson process, however there is uncertainty on the arrival rate Θ_0, so that given $[\Theta_0 = \theta_0]$, the time between customers arrivals are i.i.d. according to the exponential density function

$$f(y|\theta_0) = \begin{cases} \theta_0 e^{-\theta_0 y}, & y > 0, \\ 0, & \text{otherwise,} \end{cases}$$

where $\theta_0 \in S_{00} = (0, \infty)$. In addition, every customer orders j items (independently of each other) with probability Θ_{1j}, $j = 1, \ldots, q$, $q \geq 2$.

Set $\Theta_1 = (\Theta_{11}, \ldots, \Theta_{1(q-1)})$ and $\Theta_{1q} = 1 - \sum_{j=1}^{q-1} \Theta_{1j}$, then $\Theta = (\Theta_0, \Theta_1)$ denotes the parameter vector, and the parameter space is $S_0 = S_{00} \otimes S_{01}$, where

$$E_{01} = \left\{ (\theta_{11}, \ldots, \theta_{1(q-1)}) : \sum_{j=1}^{q-1} \theta_{1j} < 1; \theta_{1j} \ldots, q-1 \right\}.$$

We are interested in the total demand (W) during a lead time of length L^*,

$$W = \begin{cases} \sum_{i=1}^{N(T^*+L^*)-N(T^*)} U_i, & N\left(T^* + L^*\right) - N\left(T^*\right) > 0, \\ 0, & \text{otherwise,} \end{cases} \tag{18}$$

where T^* is as in (3), $L^* > 0$ is a constant, $N(s)$ is the number of clients that arrived on $[0, s]$, $s \geq 0$, and $U_1, U_2 \ldots$ are the individual demands (assumed conditionally independent relative to Θ). Information on Θ is available from (i.i.d.) observations of past clients $v = (v_1, \ldots, v_n)$ and $u = (u_1, \ldots, u_n)$ where v_i is the interarrival time of client i, and u_i is the number of items ordered by client i. Note that, according to (2), the likelihood functions for v and u take the form of

$$L(v|\theta_0) = \theta_0^n e^{-\theta_0 \sum_{i=1}^{n} v_i}, \text{ and } L(u|\theta_1) = \left(1 - \sum_{j=1}^{q-1} \theta_{1j}\right)^{c_q} \prod_{j=1}^{q-1} \theta_{1j}^{c_j},$$

respectively, where $\theta_1 = (\theta_{11}, \ldots, \theta_{1(q-1)})$, and $c_j = \prod_{i=1}^{n} I[u_i = j]$ is the number of past clients that ordered j items.

If we adopt an objective point of view, we may wish to consider a non-informative prior density for Θ, and using Jeffrey's prior may be appropriate. As is well known, Jeffrey's prior density for the exponential model is $p(\theta_0) = \theta_0^{-1}$, $\theta_0 \in S_{00}$, so that it follows from (1) that

$$p(\theta_0|v) = \frac{\theta_0^{n-1} \left(\sum_{i=1}^{n} v_i \right)^n e^{-\theta_0 \sum_{i=1}^{n} v_i}}{(n-1)!}, \tag{19}$$

which corresponds to the $\text{Gamma}\left(n, \sum_{i=1}^{n} v_i\right)$ distribution, where, for $\beta_1, \beta_2 > 0$, $Gamma(\beta_1, \beta_2)$ denotes a Gamma distribution with expectation $\beta_1 \beta_2^{-1}$. Similarly, Jeffrey's prior density for the multinomial model (see, e.g., Berger and Bernardo 1992) is

$$p(\theta_1) = \frac{\left(1 - \sum_{j=1}^{q-1} \theta_{1j}\right)^{-1/2} \prod_{j=1}^{q-1} \theta_{1j}^{-1/2}}{B(1/2, \ldots, 1/2)},$$

where $B(1/2, \ldots, 1/2) = \Gamma\left(\sum_{j=1}^{q} a_j\right)^{-1} \prod_{j=1}^{q} \Gamma(a_j)$, for $a_1, \ldots, a_n > 0$, so that it follows from (1) and (19) that

$$p(\theta_1|u) = \frac{\left(1 - \sum_{j=1}^{q-1} \theta_{1j}\right)^{c_q - 1/2} \prod_{j=1}^{q-1} \theta_{1j}^{c_j - 1/2}}{B\left(c_1 + 1/2, \ldots, c_q + 1/2\right)}, \tag{20}$$

which corresponds to a $\text{Dirichlet}\left(c_1 + 1/2, \ldots, c_q + 1/2\right)$ distribution. Thus, if we set $x_i = (z_i, u_i)$, $i = 1, \ldots, n$, $x = (x_1, \ldots, x_n)$, and $\theta = (\theta_0, \theta_1)$, under an appropriate independence assumption, the posterior density becomes $p(\theta|x) = p(\theta_0|z) p(\theta_1|u)$, where $p(\theta_0|z)$ and $p(\theta_1|u)$ are defined in (19) and (20), respectively.

Note that in this example we can obtain an analytical expression for the point forecast $\mu = E[W|X = x]$, since from (19) and (20) we have

$E[\Theta_0|V = v] = n\left(\sum_{i=1}^{n} v_i\right)^{-1}$, and

$E[\Theta_{1j}|U = u] = c^{-1}\left(c_j + 1/2\right)$

(where $c = \sum_{j=1}^{n}\left(c_j + 1/2\right) = n + q/2$), so that from (18) we have

$$\begin{aligned} \mu &= E[E[W|\Theta, X = x]|X = x] \\ &= E\left[E\left[N\left(T^* + L^*\right) - N\left(T^*\right)\middle|\Theta\right] E[U_1|\Theta] \middle| X = x\right] \\ &= L^* E\left[\Theta_0 \sum_{j=1}^{q} j\Theta_{1j} \middle| X = x\right] \\ &= L^* E[\Theta_0|V = v] \sum_{j=1}^{q} jE\left[\Theta_{1j}|U = u\right] \\ &= L^* n \left(\sum_{j=1}^{n} v_i\right)^{-1} (n + q/2)^{-1} \sum_{j=1}^{q} j\left(c_j + 1/2\right), \end{aligned}$$

which allows us to compute the point forecast μ from the available data x. However in this case, analytic expressions for a service level or a reorder point may not be easy to obtain, and the SPD algorithm described in Section 3 may be useful to compute, via simulation, the other performance measures defined in Section 2. It is worth mentioning that Muñoz and Muñoz (2011) applied a simplified version of this model to the estimation of reorder points for a T1SL using data from a car dealer.

5. CONCLUSIONS

We discussed how performance measures for inventory management (service levels and reorder points) can be suitably defined under a Bayesian framework, and how these performance measures can be estimated from the output of simulation experiments.

In the case where the sample data has the form of $x = (x_1, \ldots, x_n)$ and the likelihood has the form of (2), this approach is particularly relevant when the sample size n is small, since in that case parameter uncertainty should be relatively large. It is worth mentioning that, as $n \to \infty$, this approach is consistent with the classical approach of ignoring parameter uncertainty and fixing the value of the parameter at the maximum likelihood estimator, since under regularity conditions (see e.g., Theorem 5.14 of Bernardo 2000), $p(\theta|x)$ has an asymptotically (as $n \to \infty$) normal distribution, with mean equal to the estimator $\hat{\theta}_n$ that maximizes $p(\theta|x)$.

Finally, note that SPD can be applied when a valid algorithm to generate samples from $p(\theta|x)$ is available. If this is not the case, methodologies based on Markov Chain Monte Carlo (see, e.g., Robert 2007) can be applied, and, under regularity conditions, a valid ACI for any of the performance measures defined in Section 2 can still be obtained (see, e.g., Muñoz and Glynn 1997 and Muñoz 2010 for regularity conditions of ACI's based on the batch means method).

ACKNOWLEDGMENTS

This research has been supported by the Asociación Mexicana de Cultura AC.

REFERENCES

Bahadur, R.R., 1966. A note on quantiles in large samples. *The Annals of Mathematical Statistics* 37 (3), 577-580.

Bartezzaghi, E., Verganti, R., Zotteri, G., 1999. A simulation framework for forecasting uncertain lumpy demand. *International Journal of Production Economics*, 59 (3), 499-510.

Berger, J.O., Bernardo, J.M., 1992. Ordered group reference priors with application to the multinomial problem. *Biometrika*, 79 (1), 25-37.

Berger, J.O., Bernardo, J.M., Sun, D., 2008. The formal definition of reference priors. *The Annals of Statistics*, 37 (2), 905-938.

Bernardo, J.M., Smith, A.F.M., 2000. *Bayesian Theory*, England: John Wiley.

Bermúdez, J.D., Segura, J.V., Vercher, E., 2010. Bayesian forecasting with the Holt-Winters model. *Journal of the Operational Research Society*, 61 (1), 164-171.

Chick, S.E., 2001. Input distribution selection for simulation experiments: accounting for input uncertainty. *Operations Research*, 49 (5), 744-758.

Croston, J.D., 1972. Forecasting and stock control for intermittent demands. *Operations Research Quarterly,* 23 (3), 289-303.

Fu, M.C., 2002. Optimization for simulation: Theory vs practice. *Informs Journal on Computing*, 14 (3), 192-215.

Glasserman, P., Liu T.W., 1996. Rare-event simulation for multistage production-inventory systems. *Management Science*, 42 (9), 1292-1307.

Jahangirian, M., Eldavi, T., Naser, A., Stergioulas, L.K., Young, T., 2010. Simulation in manufacturing and business: A review, *European Journal of Operational Research*, 203 (1), 1-13.

Muñoz, D.F., Glynn, P.W., 1997. A batch means methodology for estimation of a nonlinear function of a steady-state mean. *Management Science*, 43 (8), 1121-1135.

Muñoz, D.F., Galindo, R., 2010. Incorporating parameter uncertainty for inventory management. *Proceedings of the 2010 IIE Annual Conference*, A. Johnson & J. Miller (eds.), Institute of Industrial Engineers.

Muñoz, D.F., 2010. On the validity of the batch quantile method in Markov chains. *Operations Research Letters*, 38 (3), 223-226.

Muñoz, D.F., Muñoz, D.F., 2011. Bayesian forecasting of spare parts using simulation. In: Altay, N. and Litteral, L. A. (Eds.), *Service Parts Management: Demand Forecasting and Inventory Control*, NY: Springer, 105-124.

Nahmias, S., 2008. Production and Operations Analysis, 6th ed. NY: McGraw-Hill.

Robert, C.P., 2007. *The Bayesian Choice: From Decision-Theoretic Foundations to Computational Implementations*, NY: Springer.

Silver, E.A., 1965. Bayesian determination of the reorder point of a slow moving item, *Operations Research*, 13 (6), 989-997.

Syntetos, A.A., Boylan, J.E., Disney S.M., 2009. Forecasting for inventory planning: A 50-year review. *Journal of the Operational Research Society*, 60 (1), 149-160.

Tako, A.A., Robinson, S., 2012. The application of discrete event simulation and system dynamics in the logistics and supply chain context. *Decision Support Systems,* 52 (4), 802-815.

AUTHORS BIOGRAPHY

David Fernando Muñoz is Head of the Department of Industrial & Operations Engineering at the Instituto Tecnológico Autónomo de México. He received his PhD in Operations Research from Stanford University, and is an Edelman Laureate for participating in the project that received the 2010 Franz Edelman Award. His research interests include the statistical analysis of simulation output.

David Gonzalo Muñoz is an Analyst at the R&D Department of AOL Advertising. He received his MSc degree in Management Science and Engineering from Stanford University. His research interests include stochastic modeling and yield management.