

PRONÓSTICOS BAYESIANOS USANDO SIMULACIÓN ESTOCÁSTICA*

BAYESIAN FORECASTING USING STOCHASTIC SIMULATION

David F. Muñoz Negrón¹

RESUMEN

En este artículo se presenta un marco general para la construcción de pronósticos usando simulación. El marco permite la incorporación de datos disponibles en un modelo de pronóstico, con la finalidad de cuantificar la incertidumbre en los parámetros del modelo a través de una distribución posterior, la que es utilizada para estimar un pronóstico puntual en la forma de una esperanza condicional (dados los datos). La incertidumbre en el pronóstico puntual se mide a través de una varianza condicional y un intervalo de predicción. Se discute cómo construir intervalos de confianza asintóticos para evaluar la precisión de los estimadores obtenidos por simulación, y se presentan dos ejemplos para ilustrar cómo este enfoque es consistente con las técnicas bayesianas para pronóstico. Se discuten resultados experimentales que confirman la validez de las metodologías propuestas.

Palabras clave: pronósticos, análisis de experimentos por simulación, estimación bayesiana, estimación de cuantiles.

ABSTRACT

In this article, we present a general framework to construct forecasts using simulation. This framework allows us to incorporate available data into a forecasting model in order to assess parameter uncertainty through a posterior distribution, which is then used to estimate a point forecast in the form of a conditional (given the data) expectation. The uncertainty on the point forecast is assessed through the estimation of a conditional variance and a prediction interval. We discuss how to construct asymptotic confidence intervals to assess the estimation error for the estimators obtained using simulation. We illustrate how this approach is consistent with Bayesian forecasting by presenting two examples, and experimental results that confirm our analytical results are discussed.

Key words: Forecasting, simulation output analysis, Bayesian estimation, quantile estimation.

* Esta investigación ha sido patrocinada por la Asociación Mexicana de Cultura A. C.

1. Instituto Tecnológico Autónomo de México (México). Ph. D. en Investigación de Operaciones de Stanford University, California. <davidm@itam.mx>.

INTRODUCCIÓN

La creciente necesidad de coordinación entre los diferentes procesos de negocios, particularmente entre organizaciones de la misma cadena de suministro, permite apreciar el importante papel que juegan los pronósticos en la búsqueda de la excelencia operativa (véase Fisher et ál., 2000). Por otro lado, la búsqueda de pronósticos más precisos y confiables ha motivado el desarrollo de modelos que incorporan información detallada del sistema bajo estudio (véase Kalchschmidt et ál., 2006). La simulación ha ganado un amplio reconocimiento de su capacidad para modelar sistemas complejos, por lo que no es sorprendente que tenga un alto potencial como herramienta de pronóstico, cuando el modelo del sistema es complejo. Bajo estas consideraciones, el principal propósito de este artículo es el de proponer un marco teórico que combina estimación de parámetros (con base en información muestral) y técnicas de simulación para construir pronósticos para una variable de respuesta (desempeño) del sistema bajo estudio.

La principal dificultad para construir pronósticos con base en un modelo de simulación es la incorporación de la incertidumbre sobre el valor de ciertos parámetros del modelo, los que a menudo (en las técnicas para pronóstico) se estiman a partir de datos muestrales (o históricos) disponibles. Con la finalidad de incorporar la incertidumbre paramétrica en un modelo de simulación, se utilizará el marco de la estadística bayesiana, que a su vez es un marco apropiado para la construcción de pronósticos con base en información muestral, y ha sido exitosamente aplicada en mercadotecnia (véase Rossi et ál., 2005). Actualmente ya se han propuesto métodos para el análisis de experimentos por simulación incorporando la incertidumbre paramétrica, algunos de los cuales (véanse Cheng y Holland, 1998 y 2004) son consistentes con métodos de la estadística clásica («frecuentista»), ya que primero se calcula un estimador puntual para los parámetros, y luego se estima un valor esperado de la forma $E[W | \Theta = \hat{\theta}(x)]$, donde W es la salida de una simulación (véase más adelante la ecuación 3), Θ es un vector de parámetros, y $\hat{\theta}(x)$ es el estimador puntual (una función de los datos muestrales x). Otras propuestas (Chick, 2001 y 2004; Zouaoui y Wilson, 2003) estiman la esperanza usando métodos bayesianos (la incertidumbre paramétrica se incorpora a través de una distribución posterior). Sin embargo, para producir un buen pronóstico, se requiere medir la incertidumbre en el pronóstico puntual, por lo cual en este artículo se discute también la estimación de una varianza condicional y cuantiles, que permiten la construcción de intervalos de predicción (véase Bernardo y Smith, 2000) para la variable de respuesta.

En la sección 1 se introduce la notación y los conceptos matemáticos que permiten definir con precisión los problemas de estimación que deben resolverse para construir pronósticos basados en experimentos por simulación. Además, se ilustra esta notación incluyendo dos ejemplos de la literatura sobre pronósticos bayesianos. En la sección 2 se discuten las metodologías para resolver los problemas de estimación presentados en la sección 1, probando que dichas metodologías proporcionan estimadores consistentes, que permiten construir intervalos de confianza que son asintóticamente (a medida que el número de repeticiones crece) válidos para la estimación de la esperanza condicional, la varianza condicional, y cuantiles de la variable de respuesta. En la sección 3 se presentan resultados experimentales que confirman los resultados teóricos, aplicando las metodologías propuestas a problemas de estimación con solución conocida. Finalmente, en la sección 4 se presentan las conclusiones y se discuten direcciones para posibles investigaciones en el futuro.

DEFINICIÓN DEL PROBLEMA

El objetivo de esta sección es el de proporcionar un marco conceptual para construir pronósticos de una variable de respuesta a partir de las salidas de experimentos por simulación utilizando un modelo complejo (en el sentido

que no se han obtenido expresiones analíticas para el pronóstico). Se proponen dos pasos fundamentales (véase el gráfico 1); el primer paso consiste en la cuantificación de la incertidumbre paramétrica usando los datos (x) y una función de densidad (probabilidad en el caso discreto) a priori $p(\theta)$, de manera que la incertidumbre paramétrica se expresa a través de la correspondiente función de densidad posterior $p(\theta | x)$. En el segundo paso se utiliza el modelo de simulación y la densidad posterior para estimar los parámetros requeridos para pronosticar la variable de respuesta (W). A continuación se explican estos pasos en detalle.

Estimación bayesiana

Aunque la incorporación de la incertidumbre paramétrica ha sido discutida extensamente en la literatura sobre estadística bayesiana, a continuación se presenta una pequeña revisión de los conceptos y notación que se requieren para proponer métodos de pronóstico usando simulación. Como en la mayoría de modelos para pronóstico, la salida del modelo de simulación depende de un vector Θ de parámetros inciertos, donde algunos de ellos pueden caracterizar componentes aleatorios del modelo de simulación. Es conveniente remarcar que un componente aleatorio (también llamado entrada aleatoria) de una simulación estocástica es una secuencia U_1, U_2, \dots de cantidades aleatorias que se requieren como entrada del modelo. Cuando las U_i se asumen independientes e idénticamente distribuidas, un componente aleatorio se identifica con la correspondiente distribución de probabilidades, que típicamente es miembro de una familia paramétrica. Debido a que se desea incorporar la incertidumbre paramétrica, el vector de parámetros es una variable aleatoria (denotada Θ), y $\theta \in P$ denota un valor particular, donde P es el espacio de parámetros.

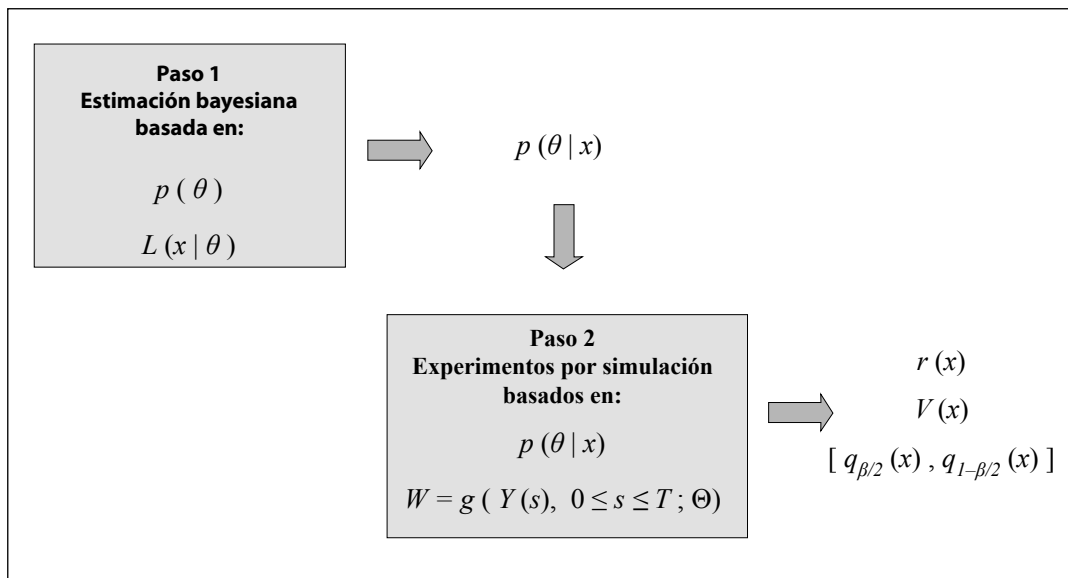


Gráfico 1
Pasos para construir un pronóstico usando simulación

Se asume que la información (datos) sobre los parámetros del modelo está disponible a través de un vector de observaciones $x \in \mathfrak{R}^n$ que satisface una función de verosimilitud $L(x | \theta)$. Si $p(\theta)$ es una densidad a priori para el parámetro Θ , la función de densidad posterior (dada x) para Θ resulta:

$$p(\theta | x) = \frac{p(\theta) L(x | \theta)}{\int_P p(\theta) L(x | \theta) d\theta}, \quad (1)$$

para $x \in \mathfrak{R}^n$ y $\theta \in P$. Nótese que la formulación discutida en Muñoz y Muñoz (2008) es un caso especial de (1), y en particular, cuando $x = (x_1, x_2, \dots, x_n)$ es un conjunto de observaciones de una muestra aleatoria $X = (X_1, X_2, \dots, X_n)$ de una función de densidad $f(y | \theta)$, la verosimilitud toma la forma:

$$L(x | \theta) = f(x_1 | \theta) f(x_2 | \theta) \dots f(x_n | \theta).$$

Nótese también que la existencia de $p(\theta)$ no es restrictiva en la práctica, ya que un prior para un parámetro puede obtenerse usando métodos «objetivos» (véase Berger et ál., 2009) y también a partir de pronósticos basados en opiniones de expertos (véase Kraan y Bedford, 2005).

Construcción de pronósticos usando simulación

Sea $Y = \{ Y(s), s \geq 0; \Theta \}$ el proceso estocástico (posiblemente multivariado) que representa la salida de un experimento por simulación. Nótese que procesos en tiempo discreto pueden incorporarse a este marco haciendo $Y(s) = Y(\lfloor s \rfloor)$, donde $\lfloor s \rfloor$ denota a la parte entera de s . Como se desea utilizar simulación transitoria, se denota por $W = g(Y(s), 0 \leq s \leq T; \Theta)$ a la variable de respuesta de interés, donde T es el tiempo de corrida, asumiendo que T es un tiempo de detención¹ con respecto al proceso Y .

Se asume que los datos (X) son independientes del experimento por simulación, lo que significa que, si $h(w | x, \theta)$ y $h(w | \theta)$ denotan a la función de densidad condicional (probabilidad en el caso discreto) de W (dado $[X = x, \Theta = \theta]$ y $[\Theta = \theta]$, respectivamente), entonces:

$$h(w | x, \theta) = h(w | \theta), \quad (2)$$

para $w \in \mathfrak{R}$, $x = (x_1, x_2, \dots, x_n) \in \mathfrak{R}^n$, y $\theta \in P$.

En general, un pronóstico para W está completamente definido por su función de distribución acumulativa (f.d.a.) $F(w | x) = P[W \leq w | X = x]$ (dado $[X = x]$). Sin embargo, desde un punto de vista práctico, un pronóstico se expresa a través de pronósticos puntuales y medidas de la incertidumbre en el pronóstico, por lo que en este artículo se considera la estimación de los siguientes parámetros de $F(w | x)$:

$$r(x) = E[W | X = x], \quad (3)$$

$$V(x) = E[W^2 | X = x] - E[W | X = x]^2 \quad \text{y} \quad (4)$$

$$q_\alpha(x) = \inf \{w: F(w | x) \geq \alpha\}, \quad (5)$$

donde $0 < \alpha < 1$ es un valor dado.

1. Del término inglés *stopping time*, véase la definición en Asmussen, 2003.

El valor esperado $r(x)$ es una medida tradicional de tendencia central para la variable de respuesta W , por lo que se considera el pronóstico puntual, mientras que la varianza $V(x)$ es una medida clásica de dispersión (incertidumbre), y el α -cuantil $q_\alpha(x)$ permite construir intervalos de predicción para W de la forma $[q_{\beta/2}(x), q_{1-\beta/2}(x)]$, el cual es llamado un intervalo de predicción del $(1 - \beta)100\%$ debido a que:

$$P [q_{\beta/2}(x) \leq W \leq q_{1-\beta/2}(x) | X = x] = 1 - \beta \quad ,$$

siempre y cuando $F(w | x)$ es continua en $q_{\beta/2}(x)$ y $q_{1-\beta/2}(x)$.

La siguiente proposición muestra como $r(x)$ depende explícitamente de la función de densidad posterior $p(\theta | x)$, y será requerida para establecer la validez de los procedimientos de estimación propuestos en la sección 2, sobre metodología.

Proposición 1. Si $p(\theta | x)$ es la función de densidad posterior definida en (1), entonces:

$$r(x) = \int_p r_1(\theta) p(\theta | x) d\theta \quad ,$$

donde $r_1(\theta) = E [W | \Theta = \theta]$ y $r(x)$ está definida en (3).

Corolario 1. Bajo la misma notación de la proposición 1:

$$V(x) = \int_p r_2(\theta) p(\theta | x) d\theta - r(x)^2 \quad ,$$

donde $r_2(\theta) = E [W^2 | \Theta = \theta]$, $r(x)$ está definida en (3) y $V(x)$ en (4).

La prueba de la proposición 1 requiere de (2) y es inmediata, por lo que se ha omitido. Los siguientes dos ejemplos de la literatura clásica de pronósticos pueden ser incorporados en el marco propuesto, e ilustran la notación presentada.

Ejemplo 1. Supóngase que W es una función de un vector $y = (y_1, y_2, \dots, y_p) \in \mathfrak{R}^p$ de variables de control, un vector de parámetros $\beta = (\beta_0, \beta_1, \dots, \beta_k) \in \mathfrak{R}^{k+1}$, y un error aleatorio ε , de acuerdo con:

$$W = w(y, \beta) + \varepsilon \quad , \tag{6}$$

donde ε sigue una distribución $N(0, \sigma^2)$ (normal con media 0 y varianza σ^2), y $w(y, \beta)$ es una función (posiblemente no-lineal). En particular, cuando $p = k$, y:

$$w(y, \beta) = \beta_0 + \beta_1 y_1 + \dots + \beta_k y_k \quad , \tag{7}$$

el modelo (6) corresponde al conocido modelo de regresión lineal. Cuando la variable de respuesta W satisface (6), puede incorporarse al marco propuesto haciendo $T = 0$, $Y(0) = (y_1, y_2, \dots, y_p, \varepsilon)$, y:

$$\theta = (\beta_0, \beta_1, \dots, \beta_k, \sigma^2) \quad . \tag{8}$$

En el análisis de regresión, los parámetros del modelo (6) se estiman a partir de un conjunto de observaciones $(w_1, z_1), \dots, (w_m, z_m)$ que satisfacen (6), es decir, $w_i = w(z_i, \beta) + \varepsilon_i$, donde $z_i = (y_{i1}, y_{i2}, \dots, y_{ip})$, $i = 1, 2, \dots, m$, los ε_i se asumen independientes e idénticamente distribuidos de acuerdo con una distribución $N(0, \sigma^2)$, y la función de verosimilitud toma la forma:

$$L(x | \theta) = (2\pi\sigma^2)^{-m/2} \exp \left[-\sum_{i=1}^m (w_i - w(z_i, \beta))^2 / 2\sigma^2 \right] ,$$

donde:

$$x = (w_1, \dots, w_m, z_1, \dots, z_m) , \tag{9}$$

y θ está definida en (8). Para el caso particular del modelo (7) de regresión lineal, cuando $p(\theta)$ es no-informativa, la densidad posterior $p(\theta | x)$ corresponde a la distribución normal-gama-inversa, denotada $NIG(a, b, \mu, V)$:

$$p(\theta | x) = \frac{(a/2)^{b/2} (\sigma^2)^{-(b+k+3)/2}}{(2\pi)^{(k+1)/2} |V|^{1/2} \Gamma(b/2)} \exp \left[-\{a + (\beta - \hat{\beta})^T V^{-1} (\beta - \hat{\beta})\} / 2\sigma^2 \right] , \tag{10}$$

donde θ y x están definidas en (8) y (9), $b = m - k - 3$, $V = (Z^T Z)^{-1}$, $a = Z^T [I - Z(Z^T Z)^{-1} Z^T] W$,

$$\hat{\beta} = (W^T W)^{-1} W^T Z, \quad W = \begin{pmatrix} w_1 \\ w_2 \\ \vdots \\ w_m \end{pmatrix}, \quad Z = \begin{pmatrix} 1 & z_1 \\ 1 & z_2 \\ \vdots & \vdots \\ 1 & z_m \end{pmatrix} .$$

Nótese que $\hat{\beta}$ es el estimador clásico de mínimos cuadrados para β en el modelo de regresión lineal (véase Montgomery, 1998). Más aun, cuando W corresponde a (6) y (7), los parámetros (3), (4) y (5) toman según O'Hagan y Forster (2004) la forma:

$$r(x) = \hat{\beta} x_0, \quad V(x) = \hat{\sigma}^2 x_0^T (Z^T Z)^{-1} x_0, \quad q_\alpha(x) = r(x) + t_{(\alpha, m-k-1)} \sqrt{\hat{\sigma}^2 (1 + x_0^T (Z^T Z)^{-1} x_0)} , \tag{11}$$

donde $x_0 = (1, y_1, y_2, \dots, y_k)^T$, $t_{(\alpha, m-k-1)}$ es el α -cuantil de una distribución t-Student con $(m - k - 1)$ grados de libertad, $\hat{\sigma}^2 = a / (m - k - 1)$, y $a, \hat{\beta}, Z$ son como en (10). Nótese que el intervalo de predicción definido en (11) coincide con el intervalo de confianza (IC) clásico usado para pronosticar bajo el modelo de regresión lineal (véase Montgomery, 1998). Sin embargo, debemos mencionar que un IC obtenido usando métodos «frecuentistas» no necesariamente coincide con el intervalo de predicción obtenido por métodos bayesianos.

Ejemplo 2. El siguiente ejemplo es un modelo markoviano reportado en Muñoz et ál. (2009), donde fue usado para pronosticar la demanda por material educacional en un programa de educación para adultos en México. Este modelo tiene una solución analítica que puede ser aprovechada para probar las metodologías a discutir en la sección 2. Sea $U = \{U_i : i = 0, 1, \dots\}$ una cadena de Markov (CM) con espacio de estados finito $E = \{0, 1, \dots, k\}$.

Supóngase que existen q clientes potenciales al tiempo t , y que el estado de cada cliente es conocido. Las transiciones de cada cliente ocurren independientemente, de acuerdo con la cadena de Markov U . Cuando el cliente potencial l está en el estado U_{t+1}^l al tiempo $t + 1$, el número de unidades que demanda de cierto producto es una función no-negativa $D(U_{t+1}^l)$ de su estado. En el artículo mencionado se asume que:

$$D(i) = \begin{cases} 1, & \text{si } i \in A, \\ 0, & \text{de otra forma,} \end{cases}$$

donde $A \subseteq E$ es el conjunto de estados donde el cliente demanda una unidad del producto. La variable de respuesta de interés es:

$$W = \sum_{l=1}^q D(U_{t+1}^l).$$

Como el estado de cada individuo al tiempo t es conocido, W puede simularse a partir del conocimiento de:

$$\theta_i = P[D(U_{t+1}^l) = 1 | U_t^l = i] = \sum_{j \in A} P[U_{t+1}^l = j | U_t^l = i],$$

$i \in E, t = 1, 2, \dots$

Este modelo puede incorporarse a nuestro marco tomando el proceso Y en tiempo discreto y $T = 1$, donde $Y(j) = (U_{t+j}^1, \dots, U_{t+j}^q), j = 0, 1$. Nótese que el conocimiento de $Y(0)$ y $\theta = (\theta_0, \theta_1, \dots, \theta_k)$ es suficiente para generar W usando simulación.

El estado 0 corresponde a los clientes nuevos, para los cuales no existe información disponible. Sin embargo, para $i = 1, \dots, k$, la información sobre θ_i proviene del registro del comportamiento de n_i clientes potenciales que estuvieron en el estado i en periodos pasados, y toma la forma de $x_i = (x_{i1}, x_{i2}, \dots, x_{in_i})$, donde x_{ij} es igual a 1 si el cliente potencial j demandó una unidad del producto, y 0 de otra forma. De acuerdo con (1) y (2), la función de verosimilitud resulta:

$$L(x | \theta) = \prod_{i=1}^k \theta_i^{r_i} (1 - \theta_i)^{n_i - r_i},$$

donde $r_i = \sum_{j=1}^{n_i} x_{ij}$, $\theta = (\theta_0, \theta_1, \dots, \theta_k)$ y $x = (x_1, x_2, \dots, x_k)$. Usando la distribución a priori de referencia (no-informativa), puede mostrarse (véanse los detalles en Muñoz et ál., 2009) que los parámetros de pronóstico puntual y de incertidumbre propuestos en (3) y (4) tienen la forma:

$$r(x) = \sum_{i=1}^k n_{it} \hat{p}_i, V(x) = \sum_{i=1}^k n_{it} \hat{p}_i (1 - \hat{p}_i) \left(1 + \frac{n_{it} - 1}{2 + n_{it}} \right), \tag{12}$$

donde $\hat{p}_i = (r_i + 1/2) / (n_i + 1)$. No se dispone de una expresión analítica sencilla para (5).

METODOLOGÍA

En esta sección se discuten procedimientos basados en simulación para la estimación de los parámetros definidos en (3), (4) y (5). No solo se discute la estimación puntual de los parámetros, sino también la construcción de un intervalo de confianza asintótico (véase Asmussen y Glynn, 2007) para medir el error de estimación de la simulación.

La metodología por discutir en el primer ítem de esta sección está basada en el procedimiento descrito en Chick (2001) con la adición de algunos cálculos para permitir la estimación de la varianza y de los cuantiles requeridos. Aunque esta metodología proporciona estimadores consistentes, se requiere de un generador de muestras de la distribución posterior $p(\theta | x)$, el que estará disponible cuando se ha reconocido la familia de distribuciones a la que pertenece $p(\theta | x)$ (como en los ejemplos de la sección anterior). Desafortunadamente, en muchas situaciones no es posible (o es muy difícil) encontrar una expresión analítica para $p(\theta | x)$, y es por esta razón que en el segundo y tercer ítem de esta sección se discuten dos metodologías de estimación que no requieren de la generación de muestras de la densidad $p(\theta | x)$.

Construcción de pronósticos por muestreo de la distribución posterior

En el gráfico 2 se presenta un primer algoritmo para la estimación puntual de los parámetros (3), (4) y (5). Este método es apropiado cuando se dispone de un algoritmo para generar muestras de $p(\theta | x)$. Como las W_i del gráfico 2 son independientes e idénticamente distribuidas, bajo débiles condiciones de regularidad², estos estimadores son consistentes y asintóticamente normales, de manera que los siguientes IC del $100(1 - \beta)\%$ de confianza son asintóticamente válidos para $r(x)$, $V(x)$ y $q_\alpha(x)$, respectivamente:

$$\left[\hat{r}(x) - z_{1-\beta/2} \frac{S_1(x)}{\sqrt{m}}, \hat{r}(x) + z_{1-\beta/2} \frac{S_1(x)}{\sqrt{m}} \right], \quad (13)$$

$$\left[\hat{V}(x) - z_{1-\beta/2} \frac{S_2(x)}{\sqrt{m}}, \hat{V}(x) + z_{1-\beta/2} \frac{S_2(x)}{\sqrt{m}} \right], \quad (14)$$

$$[Y_{n_1}, Y_{n_2}], \quad (15)$$

donde $\hat{r}(x)$, $\hat{V}(x)$ y $\hat{q}_\alpha(x)$ están definidas en el gráfico 2, $S_1(x) = \sqrt{\hat{V}(x)}$, z_α denota el α -cuantil de una distribución $N(0, 1)$, para $0 < \alpha < 1$, y:

$$S_2(x) = \sqrt{\frac{1}{m} \sum_{i=1}^m [W_i - \hat{r}(x)]^4 - \hat{V}^2(x)}, \quad n_1 = \lfloor m\alpha - z_{1-\beta/2} [m\alpha(1-\alpha)]^{1/2} \rfloor, \quad n_2 = \lceil m\alpha + z_{1-\beta/2} [m\alpha(1-\alpha)]^{1/2} \rceil.$$

Es conveniente mencionar que la validez asintótica del intervalo de confianza (15) se prueba en la sección 2.6.3 de Serfling (1980).

2. Véanse los detalles en las secciones 2.2.4 y 2.3.3 de Serfling, 1980.

Para $i = 1$ hasta el número de repeticiones m :

- a. Generar (independientemente) θ_i por muestreo de $p(\theta | x)$.
- b. Correr (independientemente) un experimento por simulación con $\Theta = \theta_i$ para obtener:

$$W_i = g (Y(s), 0 \leq s \leq T; \Theta)$$

Fin Loop

Calcular:

$$\hat{r}(x) = \frac{1}{m} \sum_{i=1}^m W_i, \quad \hat{V}(x) = \frac{1}{m} \sum_{i=1}^m [W_i - \hat{r}(x)]^2 .$$

Ordenar las W_i : $Y_1 \leq \dots \leq Y_m$ y hacer $\hat{q}_\alpha(x) = Y_{\lceil m\alpha \rceil}$.

*Gráfico 2
Estimación puntual de $r(x)$, $V(x)$ y $q_\alpha(x)$ por muestreo de la distribución posterior*

Construcción de pronósticos usando muestreo por importancia

El muestreo por importancia (MI) es una metodología muy difundida que permite estimar (vía simulación) esperanzas de la forma $E [g(\Theta)]$ y cuantiles de una función de distribución acumulativa $F(\theta)$, sin generar muestras de la variable aleatoria Θ (véanse Glynn y Iglehart, 1989; Glynn, 1996). El MI utiliza una muestra de otra variable aleatoria (con densidad conocida $q(\theta)$) para estimar el parámetro correspondiente. En esta sección se discute cómo el MI puede aplicarse para estimar los parámetros $r(x)$, $V(x)$ y $q_\alpha(x)$ sin necesidad de generar muestras de la distribución posterior $p(\theta | x)$.

Sea $q(\theta)$ una función de densidad que es estrictamente positiva para $\theta \in P$, y cero para $\theta \notin P$. Sigue de (1) y la proposición 1 que $r(x)$ puede verse como el cociente de esperanzas:

$$r(x) = \int_P r_1(\theta) p(\theta | x) d\theta = \frac{\int_P r_1(\theta) p(\theta) q(\theta)^{-1} L(x | \theta) q(\theta) d\theta}{\int_P p(\theta) q(\theta)^{-1} L(x | \theta) q(\theta) d\theta} = \frac{E_q[r_1(\theta) p(\theta) q(\theta)^{-1} L(x | \theta)]}{E_q[p(\theta) q(\theta)^{-1} L(x | \theta)]} , \quad (16)$$

y similarmente:

$$V(x) = \frac{E_q[r_2(\theta) p(\theta) q(\theta)^{-1} L(x | \theta)]}{E_q[p(\theta) q(\theta)^{-1} L(x | \theta)]} - r^2(x) , \quad (17)$$

donde $r_2(\theta)$ se define en el corolario 1.

Para $i = 1$ hasta el número de repeticiones m :

- a. Generar (independientemente) θ_i por muestreo de $q(\theta)$.
- b. Correr (independientemente) un experimento por simulación con $\Theta = \theta_i$ para obtener:

$$W_i = g(Y(s), 0 \leq s \leq T; \Theta).$$

- c. Calcular: $V_i = p(\theta_i) q(\theta_i)^{-1} L(x | \theta)$, $V_{2i} = W_i V_i$, $V_{3i} = W_i^2 V_i$.

Fin Loop

Calcular:

$$\hat{r}_{IS}(x) = \frac{\sum_{i=1}^m V_{2i}}{\sum_{i=1}^m V_i}, \quad \hat{V}_{IS}(x) = \frac{\sum_{i=1}^m V_{3i}}{\sum_{i=1}^m V_i} - \hat{r}_{IS}^2(x).$$

Ordenar las W_i : $Y_1 \leq \dots \leq Y_m$ y hacer:

$$\hat{q}_\alpha^{IS}(x) = Y_j, \quad j = \inf \left\{ l: \sum_{i=1}^l V_i \geq \alpha \sum_{i=1}^m V_i \right\}.$$

Gráfico 3

Estimación puntual de $r(x)$, $V(x)$ y $q_\alpha(x)$ usando el muestreo por importancia

A partir de (16) y (17) se aprecia que $r(x)$ y $V(x)$ son funciones no-lineales de esperanzas (las que pueden estimarse por simulación), de manera que el procedimiento de estimación puntual del gráfico 3 solo proporciona estimadores consistentes para los parámetros de pronóstico. Más aun, bajo débiles suposiciones (véase el anexo), los tres estimadores del gráfico 3 satisfacen un Teorema de Límite Central (TLC), donde las varianzas asintóticas para $r(x)$, $V(x)$ y $q_\alpha(x)$ son:

$$\sigma_r^2 = \frac{\mu_2^2}{\mu_1^4} \sigma_{11} + \frac{\sigma_{22}}{\mu_1^2} - \frac{2\mu_2}{\mu_1^3} \sigma_{12}, \tag{18}$$

$$\sigma_v^2 = \left[\frac{2\mu_2^2}{\mu_1^3} - \frac{\mu_3}{\mu_1^2} \right] \left[\left(\frac{2\mu_2^2}{\mu_1^3} - \frac{\mu_3}{\mu_1^2} \right) \sigma_{11} - \frac{4\mu_2}{\mu_1^2} \sigma_{12} + \frac{2}{\mu_1} \sigma_{13} \right] + \frac{4\mu_2^2}{\mu_1^4} \sigma_{22} + \frac{\sigma_{33}}{\mu_1^2} - \frac{4\mu_2}{\mu_1^3} \sigma_{23}, \tag{19}$$

$$\sigma_q^2 = \frac{C^{-2}(x) E_q[V_1^2 I(W_1 \leq q_\alpha(x))] - \alpha^2}{F'(q_\alpha(x) | x)}, \tag{20}$$

respectivamente, donde $C(x) = \int_p p(\theta) L(x | \theta) d\theta$, $\mu_j = E[V_{j1}]$, $\sigma_{kj} = E[(V_{k1} - \mu_k)(V_{j1} - \mu_j)]$, $V_{ji} = W_i^{j-1} V_i$, $j, k = 1, 2, 3$, $i = 1, 2, \dots, m$, y las V_i están definidas en el gráfico 3. Nótese que:

$$\hat{\mu}_j = \frac{1}{m} \sum_{i=1}^m V_{ji} \quad \text{y} \quad S_{kj} = \frac{1}{m} \left[\sum_{i=1}^m V_{ki} V_{ji} - \frac{1}{m} \left(\sum_{i=1}^m V_{ki} \right) \left(\sum_{i=1}^m V_{ji} \right) \right] \quad , \quad (21)$$

$j, k = 1, 2, 3$ son estimadores consistentes (cuando $m \rightarrow \infty$) para μ_j y σ_{kj} , respectivamente, y en consecuencia los siguientes IC del $100(1 - \beta)\%$ son asintóticamente válidos para $r(x)$ y $V(x)$, respectivamente:

$$\left[\hat{r}_{IS}(x) - z_{1-\beta/2} \frac{S_1^{IS}(x)}{\sqrt{m}} \quad , \quad \hat{r}_{IS}(x) + z_{1-\beta/2} \frac{S_1^{IS}(x)}{\sqrt{m}} \right] \quad , \quad (22)$$

$$\left[\hat{V}_{IS}(x) - z_{1-\beta/2} \frac{S_2^{IS}(x)}{\sqrt{m}} \quad , \quad \hat{V}_{IS}(x) + z_{1-\beta/2} \frac{S_2^{IS}(x)}{\sqrt{m}} \right] \quad , \quad (23)$$

donde:

$$S_1^{IS}(x) = \sqrt{\frac{\hat{\mu}_2^2}{\hat{\mu}_1^4} S_{11} + \frac{S_{22}}{\hat{\mu}_1^2} - \frac{2\hat{\mu}_2}{\hat{\mu}_1^3} S_{12}} \quad ,$$

$$S_2^{IS}(x) = \sqrt{\left[\frac{2\hat{\mu}_2^2}{\hat{\mu}_1^3} - \frac{\hat{\mu}_3}{\hat{\mu}_1^2} \right] \left[\left(\frac{2\hat{\mu}_2^2}{\hat{\mu}_1^3} - \frac{\hat{\mu}_3}{\hat{\mu}_1^2} \right) S_{11} - \frac{4\hat{\mu}_2}{\hat{\mu}_1^2} S_{12} + \frac{2}{\hat{\mu}_1} S_{13} \right] + \frac{4\hat{\mu}_2^2}{\hat{\mu}_1^4} S_{22} + \frac{S_{33}}{\hat{\mu}_1^2} - \frac{4\hat{\mu}_2}{\hat{\mu}_1^3} S_{23}} \quad ,$$

y $\hat{\mu}_j, S_{kj}, j, k = 1, 2, 3$ están definidos en (21). Como la varianza asintótica σ_q^2 depende de $F'(q_\alpha(x) | x)$ (típicamente desconocida), es preferible no tratar de estimar σ_q^2 , y proporcionar un IC asintótico del $100(1 - \beta)\%$ de confianza para $q_\alpha(x)$ en la forma de (15), donde las Y_i se definen ahora en el gráfico 3, y:

$$n_1 = \inf \left\{ l: \sum_{i=1}^l V_i \geq \alpha_1 \sum_{i=1}^m V_i \right\} \quad , \quad n_2 = \inf \left\{ l: \sum_{i=1}^l V_i \geq \alpha_2 \sum_{i=1}^m V_i \right\} \quad , \quad (24)$$

donde $\alpha_1 = \alpha - m^{-1/2} z_{1-\beta/2} S_q$, $\alpha_2 = \alpha + m^{-1/2} z_{1-\beta/2} S_q$ y $S_q = \left(\sum_{i=1}^m V_i \right)^{-1} \sqrt{\sum_{i=1}^m V_i^2 [I(W_i \leq \hat{r}_{IS}(x)) - \alpha]^2}$.

Construcción de pronósticos usando la cadena de Markov Monte Carlo

Dada una función no-negativa π sobre P y una función real r sobre P , la técnica cadena de Markov Monte Carlo (CMMC) ha sido usada para estimar parámetros de la forma:

$$E[r(\Theta)] = \int_P r(\theta) \pi^*(\theta) d\theta \quad , \quad (25)$$

utilizando el estimador puntual:

$$\hat{r}_m = \frac{1}{m} \sum_{i=1}^m r(Z_m) \quad , \quad (26)$$

donde Z_0, Z_1, \dots, Z_m es una muestra de una cadena de Markov con espacio de estados P , que tiene una distribución ergódica dada por $\pi^*(\theta) = \pi(\theta) / C$, donde $C = \int_P \pi(\theta) d\theta$. Como se explica a continuación, la CMMC puede aplicarse para estimar los parámetros de pronóstico (3), (4) y (5) tomando $\pi(\theta) = p(\theta) L(x | \theta)$.

El algoritmo de Metropolis-Hastings es una implementación particular de la CMMC que utiliza la técnica de aceptación-rechazo y una función de densidad de transición $q(z_1 | z_0)$ para generar la siguiente observación de la cadena de Markov. Cuando $Z_n = z_0$, la siguiente observación Z_{n+1} se genera de $q(z_1 | z_0)$ y el valor $Z_{n+1} = z_1$ es aceptado si $U < c(z_0, z_1)$, donde U es un número aleatorio independiente, y:

$$c(z_0, z_1) = \frac{\pi(z_1) q(z_0 | z_1)}{\pi(z_0) q(z_1 | z_0)} \quad ,$$

de otra forma se hace $Z_{n+1} = z_0$. De acuerdo con el teorema 3.5 de Asmussen y Glynn (2007), cuando la cadena de Markov definida por el algoritmo de Metropolis-Hastings es irreducible, el estimador puntual definido en (26) es un estimador consistente para la esperanza (25).

1. Generar z_0 de $q(\theta)$.

2. Para $i = 1$ hasta el número de grupos b y para $j = 1$ hasta $m_b = m/b$:

- a. Generar (independientemente) z_j de $q(\theta)$ y U uniforme en $(0, 1)$.
- b. Si $U < \frac{p(z_j) L(x | z_j) q(z_0)}{p(z_0) L(x | z_0) q(z_j)}$ hacer $z_0 \leftarrow z_j$.
- c. Correr (independientemente) un experimento por simulación con $\Theta = z_0$ para obtener:

$$W_{ij} = g(Y(s), 0 \leq s \leq T; \Theta).$$

Fin Loop

Calcular:

$$\hat{r}_i = \frac{1}{m_b} \sum_{j=1}^{m_b} W_{ij}, \quad \hat{s}_i = \frac{1}{m_b} \sum_{j=1}^{m_b} W_{ij}^2, \quad \hat{V}_i = \frac{1}{m_b} \sum_{j=1}^{m_b} (W_{ij} - \hat{r}_i)^2.$$

Ordenar las $W_{ij} : Y_1 \leq \dots \leq Y_{m_b}$ y hacer $\hat{q}_\alpha^i = Y_{\lceil m_b \alpha \rceil}$.

Fin Loop

3. Calcular:

$$\hat{r}_{MC}(x) = \frac{1}{b} \sum_{i=1}^b \hat{r}_i, \quad \hat{V}_{MC}(x) = \frac{1}{b} \sum_{i=1}^b \hat{s}_i - \hat{r}_{MC}^2(x).$$

Ordenar las $W_{ij} : Z_1 \leq \dots \leq Z_m$ y hacer $\hat{q}_\alpha^{MC}(x) = Z_{\lceil m \alpha \rceil}$.

Gráfico 4

Estimación puntual de $r(x)$, $V(x)$ y $q_\alpha(x)$ usando la cadena de Markov Monte Carlo

El muestreador independiente³ es un caso particular del algoritmo de Metropolis-Hastings que considera una función de densidad de transición de la forma $q(z_1 | z_0) = q(z_1)$, donde $q(z_1) > 0$ cuando $\pi(z_1) > 0$. El algoritmo del gráfico 4 es una implementación del muestreador independiente, y en este caso:

$$c(z_0, z_1) = \frac{\pi(z_1) q(z_0)}{\pi(z_0) q(z_1)} = \frac{p(z_1) L(x | z_1) q(z_0)}{p(z_0) L(x | z_0) q(z_1)}.$$

Nótese que —aunque no se requiere para calcular los estimadores puntuales— se está dividiendo el número de repeticiones m en b grupos de longitud m_b . Esta implementación se sugiere para aplicar el método de promedios por grupos⁴ para producir un IC asintótico para cada estimador puntual. Aunque se puede establecer un Teorema de Límite Central para cada estimador puntual (bajo suposiciones débiles), es preferible no tratar de

3. Del término inglés *independence sampler*.
 4. Del término inglés *batch means*.

estimar la varianza asintótica de cada estimador puntual, debido a que puede ser difícil, y el método de promedios por grupos podría también requerirse (véase Song y Chih, 2008). En cambio, se puede escoger un número de grupos b entre 5 y 20 (como se sugiere en Schmeiser, 1982), de manera que —bajo suposiciones razonables— los siguientes IC del $100(1 - \beta)\%$ de confianza son asintóticamente (cuando $m \rightarrow \infty$) válidos para $r(x)$, $V(x)$ y $q_\alpha(x)$, respectivamente:

$$\left[\hat{r}_{MC}(x) - t_{(b-1, 1-\beta/2)} \frac{S_1^{MC}(x)}{\sqrt{b}}, \hat{r}_{MC}(x) + t_{(b-1, 1-\beta/2)} \frac{S_1^{MC}(x)}{\sqrt{b}} \right], \quad (27)$$

$$\left[\hat{V}_{MC}(x) - t_{(b-1, 1-\beta/2)} \frac{S_2^{MC}(x)}{\sqrt{b}}, \hat{V}_{MC}(x) + t_{(b-1, 1-\beta/2)} \frac{S_2^{MC}(x)}{\sqrt{b}} \right], \quad (28)$$

$$\left[\hat{q}_\alpha^{MC}(x) - t_{(b-1, 1-\beta/2)} \frac{S_3^{MC}(x)}{\sqrt{b}}, \hat{q}_\alpha^{MC}(x) + t_{(b-1, 1-\beta/2)} \frac{S_3^{MC}(x)}{\sqrt{b}} \right], \quad (29)$$

donde $t_{(b-1, 1-\beta/2)}$ denota el $(1 - \beta / 2)$ -cuantil de una distribución t-Student con $(b - 1)$ grados de libertad, $\hat{r}_{MC}(x)$, $\hat{V}_{MC}(x)$ y $\hat{q}_\alpha^{MC}(x)$ se definen en el gráfico 4, y:

$$S_1^{MC}(x) = \sqrt{\frac{\sum_{i=1}^b [\hat{r}_i - \hat{r}_{MC}(x)]^2}{b-1}}, \quad S_2^{MC}(x) = \sqrt{\frac{\sum_{i=1}^b [\hat{V}_i - \bar{V}]^2}{b-1}}, \quad S_3^{MC}(x) = \sqrt{\frac{\sum_{i=1}^b [\hat{q}_\alpha^i - \bar{q}_\alpha]^2}{b-1}},$$

donde \hat{r}_i , \hat{V}_i y \hat{q}_α^i se definen en el gráfico 4, $\bar{V} = b^{-1} \sum_{i=1}^b \hat{V}_i$ y $\bar{q}_\alpha = b^{-1} \sum_{i=1}^b \hat{q}_\alpha^i$.

En Muñoz y Glynn (1997) se proporcionan condiciones suficientes para que el IC (27) y el IC (28) sean asintóticamente válidos. Similarmente, de acuerdo con la proposición XIII.4.3 de Asmussen y Glynn (2007), una condición para que la cadena de Markov del algoritmo del gráfico 4 sea geoméricamente ergódica es que $\sup_{\theta \in P} \{p(\theta) L(x | \theta) q(\theta)^{-1}\} < \infty$, que es suficiente para que el IC (29) sea asintóticamente válido⁵.

5. Véanse los detalles en Muñoz, 2009.

PRUEBAS EXPERIMENTALES

Aunque las metodologías propuestas se extienden fácilmente para el caso en que W es discreta (véase Iglehart, 1976), por simplicidad, en esta sección se considera el ejemplo 2 modificado para que W sea continua. Se desea pronosticar la demanda $W = \sum_{j=1}^q D_{t+1}^j$, donde D_{t+1}^j que es la demanda del cliente j al tiempo $t + 1$ satisface:

$$P[D_{t+1}^j \leq s \mid U_t^j = i] = 1 - e^{-\theta_i s},$$

para $s \geq 0$, $i = 1, 2, \dots, q$, donde $U_t^j \in \{1, 2, \dots, k\}$ es el estado del cliente j al tiempo t . Se asume que los datos sobre clientes pasados tienen la forma $x = (x_{11}, \dots, x_{1n_1}, \dots, x_{k1}, \dots, x_{kn_1})$, donde x_{ij} es la demanda del j -ésimo cliente que estuvo en el estado i . Bajo esta notación, la función de verosimilitud resulta $L(x \mid \theta) = \prod_{i=1}^k [\theta_i^{n_i} e^{-\theta_i t_i}]$, donde $t_i = \sum_{j=1}^{n_i} x_{ij}$, por lo que asumiendo un prior constante, sigue de (1) que $p(\theta \mid x)$ corresponde a la densidad conjunta de k variables aleatorias independientes, donde θ_i sigue una distribución Gama $(1 + n_i, 1 / t_i)$. Se puede verificar fácilmente que:

$$r(x) = \sum_{i=1}^k \frac{n_{it}}{\hat{\theta}_i} \quad \text{y} \quad V(x) = \sum_{i=1}^k \frac{n_{it} (n_{it} + n_i)}{\hat{\theta}_i^2 (n_i - 1)}, \quad (30)$$

donde $\hat{\theta}_i = n_i / t_i$, n_{it} y n_i son como en (12).

En todos los experimentos que se reportan en esta sección se han considerado $k = 5$ estados, con $n_i = n_{it} = 10$, para $i = 1, \dots, 5$. Como cada experimento por simulación depende solo de los valores t_i , se decidió no simular valores para x , asumiendo que $t_i = 10/i$, $i = 1, \dots, 5$. Bajo estas suposiciones se puede verificar de (30) que $r(x) \approx 22,8333$ y $V(x) \approx 32,5247$. Por otro lado, se obtuvo $q_{0,9}(x) \approx 30,1875$ —para un ancho medio del 90% de confianza de 0,0066— de una corrida con $m = 10^6$ repeticiones. Estos valores fueron considerados como los valores verdaderos de los parámetros, con la finalidad de calcular el cubrimiento empírico, error cuadrático promedio y sesgo para cada conjunto de experimentos.

En los cuadros 1, 2 y 3 se presentan los resultados de tres conjuntos de experimentos correspondientes a cada uno de los métodos propuestos en la sección precedente. Los resultados reportados en el cuadro 1 corresponden al algoritmo de muestreo posterior del gráfico 2, mientras que los resultados del cuadro 2 corresponden al algoritmo de muestreo por importancia del gráfico 3, donde $q(\theta)$ corresponde a la distribución límite para $p(\theta \mid x)$ (véase Bernardo y Smith, 2000), es decir, una distribución $N(\mu, V^{-1})$ en el cual fueron rechazados los valores negativos, donde μ es el estimador de máxima verosimilitud para Θ y V es la matriz de influencia. Puede verificarse que el i -ésimo elemento de μ es $\mu_i = n_i / \sum_{j=1}^{n_i} x_{ij}$ y V es una matriz diagonal con elemento diagonal $V_i = n_i / m_i^2$. Por otro lado, los resultados reportados en el cuadro 3 corresponden al algoritmo de la cadena de Markov Monte Carlo del gráfico 4 considerando el número de grupos $b = 10$. Cada conjunto de experimentos consistió de $M = 1000$ repeticiones independientes del procedimiento de estimación correspondiente, con diferentes números de repeticiones por experimento.

Cuadro 1. Desempeño del muestreo posterior para diferentes números de repeticiones m y $M = 1000$

Muestreo posterior	Parámetro estimado	Cubrimiento empírico	Ancho medio		Error cuadrático promedio	Sesgo
			Promedio	Desviación estándar		
$m = 100$ $M = 1000$	$r(x)$	0,901	0,9284	0,0956	0,3121	0,0207
	$V(x)$	0,823	9,9583	5,7213	47,6784	-0,3325
	$q_{0,9}(x)$	0,901	2,3029	0,7271	1,7081	-0,1719
$m = 400$ $M = 1000$	$r(x)$	0,907	0,4683	0,0248	0,0820	0,0002
	$V(x)$	0,871	5,4850	1,7765	12,0104	-0,0049
	$q_{0,9}(x)$	0,906	1,1173	0,2534	0,4490	-0,0446
$m = 1600$ $M = 1000$	$r(x)$	0,910	0,2347	0,0065	0,0191	0,0035
	$V(x)$	0,880	2,8316	0,5617	3,2480	0,0841
	$q_{0,9}(x)$	0,918	0,5622	0,0884	0,1083	0,0009

Cuadro 2. Desempeño de muestreo por importancia para diferentes números de repeticiones m y $M = 1000$

MI	Parámetro estimado	Cubrimiento empírico	Ancho medio		Error cuadrático promedio	Sesgo
			Promedio	Desviación estándar		
$m = 100$ $M = 1000$	$r(x)$	0,848	1,2143	0,4044	0,8666	-0,0063
	$V(x)$	0,817	9,3841	3,5467	48,2605	-1,0946
	$q_{0,9}(x)$	0,875	2,9426	8,2351	2,8352	-0,1874
$m = 400$ $M = 1000$	$r(x)$	0,876	0,7489	0,3980	0,4547	-0,0984
	$V(x)$	0,844	5,5520	2,7428	20,2123	-0,4378
	$q_{0,9}(x)$	0,878	1,3265	0,9694	0,9570	-0,2016
$m = 1600$ $M = 1000$	$r(x)$	0,881	0,4863	0,3721	0,2640	-0,1553
	$V(x)$	0,882	3,3241	1,8989	6,0338	-0,1961
	$q_{0,9}(x)$	0,866	0,7514	0,5240	0,3407	-0,2165

Cuadro 3. Desempeño de la cadena de Markov Monte Carlo para diferentes números de repeticiones m , $b = 10$ y $M = 1000$

CMMC	Parámetro estimado	Cubrimiento empírico	Ancho medio		Error cuadrático promedio	Sesgo
			Promedio	Desviación estándar		
$m = 100$ $b = 10$ $M = 1000$	$r(x)$	0,844	1,3690	0,3624	1,1433	0,0045
	$V(x)$	0,799	11,6931	6,2616	63,4274	-0,9842
	$q_{0,9}(x)$	0,804	2,2792	0,6621	3,2823	-0,2822
$m = 400$ $b = 10$ $M = 1000$	$r(x)$	0,860	0,8393	0,3009	0,4705	-0,1089
	$V(x)$	0,838	6,7660	2,7582	19,8071	-0,3002
	$q_{0,9}(x)$	0,862	1,4558	0,4762	1,0639	-0,2193
$m = 1600$ $b = 10$ $M = 1000$	$r(x)$	0,870	0,5266	0,2835	0,2520	-0,1561
	$V(x)$	0,871	3,7427	1,2736	6,2508	-0,1604
	$q_{0,9}(x)$	0,885	0,8463	0,4069	0,4159	-0,2229

Como puede observarse de los cuadros 1, 2 y 3, se obtuvieron cubrimientos razonables (cerca del nominal de 90%) con un número relativamente pequeño de repeticiones para cada uno de los 3 algoritmos, confirmando la validez asintótica de los intervalos de confianza (IC) propuestos. Los resultados muestran cómo el error cuadrático promedio y el ancho medio decrecen consistentemente cuando se incrementa el número de repeticiones, con un pequeño deterioro del sesgo para los métodos de muestreo por importancia (MI) y de la cadena de Markov Monte Carlo (CMMC).

CONCLUSIONES

En este artículo se muestra cómo pueden incorporarse datos que proporcionan una medición de la incertidumbre paramétrica en un modelo de simulación, para producir estimadores de los parámetros necesarios para construir pronósticos bajo un enfoque bayesiano, con la ventaja que bajo este enfoque se puede utilizar un modelo complejo que incorpore información detallada sobre el sistema bajo estudio.

Se ha presentado un marco bayesiano para producir intervalos de predicción, así como tres metodologías que permiten estimar el intervalo utilizando simulación. Para cada una de estas metodologías se muestra cómo se pueden construir intervalos de confianza asintóticos para medir el error de estimación, e incrementar el número de repeticiones del experimento si fuera necesario.

La primera metodología (muestreo posterior) requiere de un algoritmo válido para generar muestras de la distribución posterior $p(\theta | x)$, mientras que las dos últimas (muestreo por importancia y muestreo de la cadena de Markov Monte Carlo) no requieren someter una muestra de $p(\theta | x)$, por lo que son apropiadas cuando la familia de distribuciones a la que pertenece $p(\theta | x_0)$ no ha sido identificada. Pruebas analíticas, así como resultados experimentales muestran que las metodologías propuestas producen estimadores consistentes e intervalos de confianza asintóticamente válidos para los parámetros requeridos para pronosticar la variable de respuesta, y como es de esperar, el error de estimación (medido por el ancho medio del IC) tiende a cero a medida que el número de repeticiones del experimento por simulación crece.

ANEXO

Teoremas de límite central para los estimadores puntuales del algoritmo de muestreo por importancia

Supóngase que " \Rightarrow " denota convergencia débil (a medida que $m \rightarrow \infty$) y $N(0, 1)$ denota una distribución normal estándar. En este anexo se presentan pruebas de las siguientes proposiciones.

Proposición A.1. Supóngase que $\sigma_{kj} < \infty$, para $k, j = 1, 2$. Entonces:

$$\sqrt{m} [\hat{r}_{IS}(x) - r(x)] \Rightarrow \sigma_r N(0, 1) \quad ,$$

donde σ_r y las σ_{kj} están definidas en (18).

Proposición A.2. Supóngase que $\sigma_{kj} < \infty$, para $k, j = 1, 2, 3$. Entonces:

$$\sqrt{m} [\hat{V}_{IS}(x) - V(x)] \Rightarrow \sigma_v N(0, 1) \quad ,$$

donde σ_v y las σ_{kj} están definidas en (19).

Proposición A.3. Supóngase que $E[V_I^3] < \infty$, y que $F(w | x)$ tiene derivada positiva en $q_\alpha(x)$. Entonces:

$$\sqrt{m} [\hat{q}_\alpha^{IS}(x) - q_\alpha(x)] \Rightarrow \sigma_q N(0, 1) \quad ,$$

donde V_I se define en el gráfico 3 y $F(w | x)$ es como en (5).

La prueba de la proposición A.3 sigue directamente del teorema 1 de Glynn (1996), nótese que el estimador $\tilde{\alpha}_{3m}$ corresponde al estimador $\hat{q}_\alpha^{IS}(x)$ cuando $F(w) = F(w | \theta)$, $p = \alpha$ y $L_I = C(x)^{-1} V_I$.

Las proposiciones A.1 y A.2 se probarán usando el método Delta, que se enuncia en el lema A.1 (se puede encontrar una prueba en la proposición 2 de Muñoz y Glynn, 1997).

Hipótesis A.1. Los vectores aleatorios (p -variados) X_i , $i = 1, 2, \dots$ satisfacen el Teorema de Límite Central:

$$m^{1/2} [\bar{X}_m - r] \Rightarrow GN_p(0, I) \quad ,$$

donde r es un vector (p -variado), $\bar{X}_m = m^{-1} \sum_{i=1}^m X_i$, G es una matriz $p \times p$ y $N_p(0, I)$ denota una distribución normal (p -variada) con media 0 y varianza I (la matriz identidad).

Lema A.1. Supóngase que se cumple la hipótesis A.1 y que la función $f: \mathfrak{R}^p \rightarrow \mathfrak{R}$ es diferenciable en una vecindad de r . Entonces:

$$m^{1/2} [f(\bar{X}_m) - f(r)] \Rightarrow \sigma N(0, 1) \quad ,$$

donde $\sigma = \sqrt{\nabla f^T(r) G G^T \nabla f(r)}$.

Usando la misma notación que en (18), sean $p = 2$, $X_i = (V_{1i}, V_{2i})$, $r = (\mu_1, \mu_2)$, $f(r) = \mu_2 / \mu_1$ y

$$G G^T = \begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{12} & \sigma_{22} \end{bmatrix} .$$

Entonces $\nabla f(\mu) = (\mu_2 / \mu_1^2, -1 / \mu_1)^T$ y la proposición A.1 sigue del lema A.1.

Nótese que la hipótesis A.1 se cumple del teorema B de la sección 1.91 de Serfling (1980).

Similaramente, sean $p = 3$, $X_i = (V_{1i}, V_{2i}, V_{3i})$, $r = (\mu_1, \mu_2, \mu_3)$, $f(r) = (\mu_3 / \mu_1) - (\mu_2 / \mu_1)^2$ y

$$G G^T = \begin{bmatrix} \sigma_{11} & \sigma_{12} & \sigma_{13} \\ \sigma_{12} & \sigma_{22} & \sigma_{23} \\ \sigma_{13} & \sigma_{23} & \sigma_{33} \end{bmatrix} .$$

Entonces $\nabla f(\mu) = (2\mu_2^2 / \mu_1^3 - \mu_3 / \mu_1^2, -2\mu_2 / \mu_1^2, 1 / \mu_1)^T$ y la proposición A.2 sigue del lema A.1.

Referencias bibliográficas

- Asmussen, S. (2003). *Applied probability and queues* (2.^a ed.). New York: Springer.
- Asmussen, S. & Glynn, P.W. (2007). *Stochastic simulation algorithms and analysis*. New York: Springer.
- Berger, J. O., Bernardo, J. M. & Sun, D. (2009). The formal definition of reference priors. *Annals of Statistics*, 37 (2), 905-938.
- Bernardo, J. M. & Smith, A. F. M. (2000). *Bayesian theory*. Chichester: Wiley.
- Cheng, R. C. H. & Holland, W. (1998). Two points methods for assessing variability in simulation output. *Journal of Statistical Computation and Simulation*, 60 (3), 183-205.
- Cheng, R. C. H. & Holland, W. (2004). Calculation of confidence intervals for simulation output. *ACM Transactions on Modeling and Computer Simulation*, 14 (4), 344-362.
- Chick, S. E. (2001). Input distribution selection for simulation experiments: accounting for input uncertainty. *Operations Research*, 49 (5), 744-758.
- Chick, S. E. (2004). Bayesian methods for discrete event simulation. En R. G. Ingalls, M. D. Rossetti, J. S. Smith, B. A. Peters (eds.). *Proceedings of the 2004 winter simulation conference* (89-100). New Jersey: IEEE.
- Fisher, M. L., Raman, A., McClelland, A. S. (2000). Rocket science retailing is almost here. *Harvard Business Review*, 78 (4), 115-124.
- Glynn, P. W. (1996). Importance sampling for Monte Carlo estimation of quantiles. En *Proceedings of the second international workshop on mathematical methods in stochastic simulation and experimental design* (San Petersburgo), 180-185.
- Glynn, P. W. & Iglehart, D. L. (1989). Importance sampling for stochastic simulations. *Management Science*, 35 (11), 1367-1392.
- Glynn, P. W. & Iglehart, D. L. (1991). Simulation output analysis using standardized time series. *Mathematics of Operations Research*, 15 (1), 1-16.
- Iglehart, D. L. (1976). Simulating stable stochastic systems, VI: Quantile estimation. *Journal of the Association for Computing Machinery*, 23 (2), 347-360.
- Kalchschmidt, M., Verganti, R. & Zotteri, G. (2006). Forecasting demand from heterogeneous customers. *International Journal of Operations & Production Management*, 26 (6), 619-638.
- Kraan, B. & Bedford, T. (2005). Probabilistic inversion of expert judgments in the quantification of model uncertainty. *Management Science*, 51 (6), 995-1006.
- Montgomery, D. C. (1998). *Design and analysis of experiments*. New York: Wiley.
- 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. & Muñoz, D. G. (2008). A Bayesian framework for the incorporations of priors and sample data in simulation experiments. *The Open Operational Research Journal*, 2, 44-51.
- Muñoz, D. F. (2009). On the validity of the batch quantile method in Markov chains. Aceptado para su publicación en *Operations Research Letters*.
- Muñoz, D. F., Romero-Hernández, O., Detta-Silveira, J. E. & Muñoz, D. G. (2009). Forecasting demand for educational material for adult learners in Mexico. Aceptado para su publicación en la *International Transactions in Operational Research*.
- O'Hagan, A. & Forster, J. (2004). *J. Kendall's advanced theory of statistics volume 2B Bayesian inference* (2.^a ed.). London: Arnold.
- Rossi, P. E., Allenby, G. M. & McCulloch, R. (2005). *Bayesian statistics and marketing*. Chichester: Wiley.
- Schmeiser, B. (1982). Batch size effects in the analysis of simulation output. *Operations Research*, 30 (3), 556-568.
- Serfling, R. J. (1980). *Approximation theorems of mathematical statistics*. New York: John Wiley.
- Song, T. W. & Chih, M. (2008). Implementable mse-optimal dynamic partial-overlapping batch means estimators for steady-state simulations. En S. J. Mason, R. R. Hill, L. Mönch, O. Rose, T. Jefferson & J. W. Fowler (eds.). *Proceedings of the 2008 winter simulation conference* (426-435). New Jersey: IEEE.
- Zouaoui, F. & Wilson, J. R. (2003). Accounting for parameter uncertainty in simulation input modeling. *IIE Transactions*, 35 (9), 781-792.