

Aplicación de modelación bayesiana y optimización para pronósticos de demanda

Bayesian modeling application and optimization to demand forecasting

Marisol Valencia Cárdenas*

Juan Carlos Correa Morales**

Francisco Javier Díaz Serna***

Universidad Nacional de Colombia, Sede Medellín (Colombia)

Sebastián Ramírez Agudelo****

Universidad Pontificia Bolivariana, Sede Medellín (Colombia)

* M.Sc. en Ciencias-Estadística. Universidad Nacional de Colombia, Sede Medellín. Ph(C) en Ingeniería-Industria y Organizaciones, Facultad de Minas, Universidad Nacional de Colombia (UNAL), Sede Medellín. *solmarival@gmail.com*

** Ph.D. En Estadística de la Universidad de Kentucky. Profesor asociado de la Universidad Nacional de Colombia (UNAL), Sede Medellín, Escuela de Estadística, *jccorrea@unal.edu.co*

*** Ph.D. en Sistemas e Informática de la Universidad Nacional de Colombia, sede Medellín, Facultad de Minas. Profesor asociado del Departamento de Ciencias de la Computación y la Decisión, Facultad de Minas, Universidad Nacional de Colombia (UNAL), Sede Medellín, *javidiaz@unal.edu.co*

**** Ingeniero Industrial de la Universidad Pontificia Bolivariana (UPB), Sede Medellín, *sebastian.ramirezag@gmail.com*

Correspondencia: Marisol Valencia Cárdenas, Cll 6 C Sur No 84C-45 Urb Jardines del Rodeo Apt 113, T4. e-mail: *solmarival@gmail.com*, *mvalencia@unal.edu.co*

Resumen

Las prácticas para el manejo óptimo de inventarios son una necesidad en las cadenas de abastecimiento, en especial para productos industriales terminados. Un aporte al mejoramiento de esta cadena logística consiste en encontrar modelos eficientes para el pronóstico de la demanda de estos productos y que a su vez permitan minimizar los costos del manejo de los inventarios; aspectos que se dificultan cuando hay presencia de pocos datos históricos. La propuesta de este trabajo consiste en aplicar varias técnicas bayesianas con un método de optimización, comparando su eficiencia mediante el indicador MAPE para el pronóstico de la demanda, en casos de pocos datos. Los resultados indican que la técnica de pronóstico del valor esperado con retardo de orden 1 en los parámetros, usando la metaheurística Tabú, es la que muestra mejor acierto en el pronóstico.

Palabras clave: estadística bayesiana, modelos de pronóstico, optimización.

Abstract

Practices for optimal inventory management are a need at supply chains, especially for finished industrial products. A contribution to this logistic chain consists in finding efficient forecast of products demand, which permits to minimize cost inventory management, aspects that are more difficult in the presence of few historical data. This work proposal consists in the application of various Bayesian techniques with an optimization method, comparing its efficiency with MAPE indicator for demand forecasting, with few historical data. Results indicate that the expected value technique with an order 1 delay in the parameters, using Tabu metaheuristic shows the best accuracy in the forecast.

Keywords: bayesian statistics, forecast models, optimization.

Fecha de recepción: 20 de junio de 2013
Fecha de aceptación: 10 de febrero de 2014

1. INTRODUCCIÓN

La planeación y seguimiento de un sistema de inventarios contempla muchas variables, como la demanda, tiempos de suministro, entre otras [1]. Este sistema, a su vez, orienta la planeación de producción y distribución [2], [3], lo que hace exigente una estimación adecuada de la demanda de producto terminado. De un buen pronóstico de la demanda de los productos terminados depende la planeación eficiente del abastecimiento, pues afecta la logística en general, así como las utilidades de la compañía, lo cual impacta en gran medida su funcionamiento. Por ello, el reconocimiento de la aleatoriedad de la demanda y su adecuada modelación para predecir sobre esta es objeto de numerosos estudios, como lo muestran las revisiones de los autores: [3], [4].

La demanda de producto terminado ha sido pronosticada en numerosos trabajos: [2], [3], [5], [6], con modelos como regresión, Modelos Integrados Autorregresivos y de Medias Móviles (ARIMA), o los estacionales SARIMA, suavización exponencial, entre otros, con el fin de lograr una planeación adecuada de la logística e inventarios de producto terminado. También se muestra en [7], quienes proponen un modelo de pronóstico basado en una distribución empírica de probabilidad para la demanda, logrando aparentemente mantener un servicio el 95 % en el cumplimiento de pedidos, afirmando que el requerimiento de normalidad de los modelos ARIMA y de regresión no se cumple en muchos casos, o incluso no se cuenta siempre con la cantidad de datos requerida por estos modelos. Sin embargo, con dicha distribución hallada queda la pregunta respecto a qué tanto hubiera mejorado el pronóstico si se tienen en cuenta otras características, como: tendencias, rezagos y estacionalidad, para capturar mejor la variación de la demanda.

Por todo esto, encontrar alternativas de pronósticos acertadas que no requieran demasiados datos para ello, son objeto de interés en la industria manufacturera y de varias investigaciones, en especial las que utilizan los métodos bayesianos como los que se presentan en este trabajo aplicado [8]-[11].

El proceso de la modelación bayesiana parte de la definición de una distribución de probabilidad a priori para el (los) parámetro (s) y de otra para

los datos, obteniendo con su multiplicación la distribución de probabilidad a posteriori [12], con la cual se estima la función predictiva. Este proceso es básico para estimar diversos modelos de predicción bayesianos [10], [13], como los utilizados en este trabajo: el método de inferencia predictiva bayesiana [13], el valor esperado bayesiano [14] y la propuesta de parámetros autorregresivos en el método de valor esperado, útiles cuando existen pocos datos históricos.

La propuesta de este trabajo consiste en describir cuatro métodos estadísticos, tres bayesianos y uno basado en generación de variable aleatoria con distribución Poisson para realizar pronósticos de demanda. Se agregó además el algoritmo metaheurístico Tabú para dos métodos bayesianos con el fin de optimizar su resultado. Se simula la eficiencia de los métodos bayesianos explicados y se aplican al caso de un producto terminado para una empresa manufacturera, con el fin de comparar cuál es el más eficiente para pronosticar usando el indicador de error de pronóstico o capacidad (MAPE).

2. MODELAMIENTO BAYESIANO

La estadística bayesiana utiliza distribuciones de probabilidad para modelar la incertidumbre de variables aleatorias, que en este caso son los parámetros distribucionales, los modelos de regresión o de series de tiempo. Las estimaciones de parámetros de los modelos propuestos se calculan con la función a posteriori, que es proporcional a la verosimilitud observada multiplicada por la distribución a priori.

En el modelamiento de pronóstico bayesiano se utiliza la distribución predictiva como base del valor que se va a pronosticar, que a su vez se basa en la distribución a posteriori para los parámetros involucrados en dicha estimación [15]. Dicho modelamiento requiere algunas veces la provisión de información de expertos o conocimiento del comportamiento distribucional, así como datos actualizados en el instante de tiempo previo a su estimación.

Para estimar los pronósticos con inferencia predictiva bayesiana es usual actualizar el valor del parámetro principal con un proceso que comienza por derivar el logaritmo natural de la función a posteriori, denominado estimador de Bayes de la función de pérdida escalonada. Las siguientes

son notaciones asociadas a términos del proceso de estimación bayesiana y pronósticos llevadas a cabo en este trabajo:

- X_t denota el t ésimo valor de la serie en el tiempo t .
- La serie empieza en $t=1$.
- Los parámetros como variables aleatorias inciertas pueden denotarse con caracteres griegos, considerados los parámetros desconocidos, y romanos para los conocidos.
- La distribución a priori para cada variable aleatoria, denotada $\xi(\lambda)$. La función de verosimilitud es $L(\lambda | \text{datos})$. La distribución conjunta a posteriori de dos cantidades aleatorias es $\xi(\lambda, \alpha | D)$. A partir de la conjunta anterior se determina la distribución a posteriori condicional para cada variable: $\xi(\alpha | \lambda, D)$.
- Los valores observados de la serie serán $X_1, X_2, X_3, \dots, X_t$; los valores futuros inciertos serán $X_{t+1}, X_{t+2}, X_{t+3}, \dots$.
- La distribución predictiva para pronosticar X_t se denota $P(Y_t | D_{t-1})$; ver su definición en [13].

La simulación de parámetros y valores pronosticados a partir de la distribución a posteriori y predictiva se realiza con técnicas como Monte Carlo por Cadenas de Markov, introducida en la siguiente sección.

Monte Carlo por Cadenas de Markov (MCMC)

El método de Monte Carlo genera valores independientes tomando como base una distribución de probabilidad deseada. Las cadenas de Markov se basan en una estructura de dependencia entre los valores simulados consecutivamente. Al unir ambas técnicas se crea el método Monte Carlo por Cadenas de Markov, que consiste en hacer un muestreo a partir de distribuciones de probabilidad basadas en la construcción de cadenas, donde cada valor simulado, de acuerdo a [12] y [16], tiene dependencia con el dato anterior, llegando a una convergencia a la distribución deseada. Después de gran cantidad de simulaciones se establecen resultados que se utilizan como una muestra incorrelacionada de la distribución deseada.

Los métodos MCMC son una estrategia para generar valores de la variable aleatoria buscada λ^i mientras explora el espacio usando cadenas de Markov.

La simulación para actualizar el valor del parámetro λ para este trabajo se llevó a cabo con el proceso de MCMC, que se basa en el muestreador de Gibbs, y este, a su vez, utiliza las distribuciones a posteriori condicionales conocidas de los parámetros, valor que permite la actualización de la función predictiva para el pronóstico.

Muestreador de Gibbs

Es una de las técnicas más usadas para las simulaciones de procesos de Monte Carlo por Cadenas de Markov (MCMC). El muestreador de Gibbs requiere conocimiento específico sobre la naturaleza condicional de la relación entre variables de interés. La idea básica es realizar un muestreo para el parámetro, por ejemplo, λ , a partir de su distribución a posteriori condicional. Cuando se involucran más parámetros, este muestreador utiliza cada una de las distribuciones a posteriori condicionales por parámetro.

Este método se puede definir como un kernel de transición [12], creado por una serie de distribuciones condicionales en un esquema markoviano, actualizado entre ellas mismas; estas se basan en la distribución a posteriori $\xi(\lambda)$, donde λ es el parámetro para producir una cadena de Markov cíclica alrededor de las condicionales y que converja a dicha distribución.

El siguiente es un resumen del proceso iterativo del muestreador de Gibbs, si únicamente se requiere hacerlo para el parámetro λ :

Elegir valor inicial: $\lambda^{[0]}$

En la j ésima iteración el ciclo se completa generando valores de cada parámetro condicionados a los valores previamente encontrados de las k distribuciones dadas por $\lambda^{[j]} \sim \xi[\lambda | \lambda^{[j-1]}]$.

Así, se incrementa j hasta la convergencia de la distribución a posteriori del parámetro [12]. Cuando existen más parámetros, cada uno se actualiza a partir del cambio de los anteriores, por ejemplo:

$$\lambda^{[j]} \sim \xi[\lambda | \alpha^{[j-1]}, \lambda^{[j-1]}] \text{ y } \alpha^{[j]} \sim \xi[\lambda | \alpha^{[j-1]}, \lambda^{[j]}]$$

3. METODOLOGÍA

En este artículo se estudian cuatro métodos estadísticos: tres modelos bayesianos y uno de generación de variable aleatoria. En las tres subsecciones siguientes se describe el proceso analítico de cada una de las tres técnicas: método de inferencia predictiva bayesiana, método de valor esperado bayesiano de la función predictiva y valor esperado bayesiano con parámetros autorregresivos, y en la siguiente, el algoritmo Tabú para optimizar las últimas dos; procesos necesarios para mostrar la metodología estadística que lleva a cada ecuación usada en los pronósticos, cuyo desempeño específico se muestra en los resultados para el caso de estudio al que fueron aplicados. Estas se comparan con otra técnica clásica: la generación de variable aleatoria Poisson, por medio del MAPE, el cual muestra mejor desempeño cuando su valor es menor.

El caso de estudio se trata de un producto de alto volumen de ventas en la empresa, que constantemente busca mejores técnicas para optimizar sus pronósticos, y con ello, la planeación de sus inventarios. La empresa proporciona 24 datos de demanda mensual real. A continuación se explica el proceso analítico de los métodos bayesianos propuestos.

Pronósticos con la función predictiva bayesiana

Los datos tienen una naturaleza discreta que de acuerdo con una prueba de bondad de ajuste siguen una distribución de Poisson, cuyo parámetro λ se asume con distribución a priori Gamma, que a su vez tiene parámetros α y β que se asumirán con distribución a priori uniforme. La distribución Gamma es recomendada en [8]. La función predictiva será explicada a continuación.

La media de los datos con distribución Poisson es 100.17, que se establece como valor inicial para la función de verosimilitud dada por $L(\lambda | \text{datos})$, donde λ representa una tasa de ventas mensuales de naturaleza continua. A partir de esta se obtuvo la distribución a posteriori $\xi(\lambda | \alpha, \beta, \text{datos})$ condicional, con la cual se realiza la función predictiva, como se expresa de forma analítica a continuación.

Función de Verosimilitud: Distribución Poisson para los datos

$$L(\lambda|x_1, x_2, \dots, x_n) = \prod_{i=1}^n \frac{\lambda^{x_i} e^{-\lambda}}{x_i!} = \frac{\lambda^{\sum_{i=1}^n x_i} e^{-n\lambda}}{\prod_{i=1}^n x_i!} \quad (1)$$

Función de distribución a priori para Lambda: dicho parámetro seguirá la distribución Gamma (α, β). Para simplificar las ecuaciones se usará la palabra: Datos para representar la muestra: X_1, X_2, \dots, X_n .

$$\xi(\lambda) = \frac{\beta^\alpha}{\Gamma(\alpha)} \lambda^{\alpha-1} e^{-\lambda\beta} \text{ con } \lambda > 0 \quad (2)$$

Distribución a posteriori para λ :

$$\xi(\lambda, \alpha, \beta | \text{Datos}) \propto \frac{\beta^\alpha}{\Gamma(\alpha)} \lambda^{\alpha + \sum_{i=1}^n x_i - 1} e^{-\lambda\beta - n\lambda} \quad (3)$$

Quedando la siguiente:

$$\xi(\lambda | \alpha, \beta, \text{Datos}) \propto \lambda^{\alpha + \sum_{i=1}^n x_i - 1} e^{-\lambda(\beta+n)}, \lambda > 0 \quad (4)$$

Distribución a posteriori para α :

$$\xi(\alpha | \lambda, \beta, \text{Datos}) = \frac{\beta^\alpha}{\Gamma(\alpha)} \lambda^{\alpha + \sum_{i=1}^n x_i - 1} \quad (5)$$

Distribución a posteriori para β :

$$\xi(\beta | \lambda, \alpha, \text{Datos}) = \beta^\alpha e^{-\lambda(\beta+n)} \quad (6)$$

Para las ecuaciones (5) y (6), el parámetro λ se actualiza usando su estimador, que es el valor esperado de la distribución a posteriori $\text{Gamma}(\alpha + \sum_{i=1}^n x_i, \beta + n)$. A partir de la distribución dada en (4), actualizándola con los parámetros vía MCMC por las distribuciones dadas por (5) y (6), se estima la función predictiva para pronosticar X_{t+1} :

$$f_{X_{t+1}|X}(X_{t+1}|X) = \int_0^\infty f(x_{t+1}|\lambda) \xi(\lambda | \alpha, \beta, \text{Datos}) d\lambda \quad (7)$$

$$f_{X_{t+1}|X}(X_{t+1}|X) = \int_0^\infty \frac{\lambda^{x_{t+1}} e^{-\lambda}}{x_{t+1}!} \frac{\beta^\alpha}{\Gamma(\alpha)} \lambda^{\alpha + \sum_{i=1}^n x_i - 1} e^{-\lambda(\beta+n)} d\lambda \quad (8)$$

Tomando $\alpha_1 = \alpha + \sum_{i=1}^n x_i$ y $\beta_1 = \beta + n$, y agregando términos en la integral y externamente para completar la forma de la distribución gamma cuya integral vale 1, la expresión anterior se puede simplificar así:

$$f_{X_{t+1}|X}(X_{t+1}|X_t) = \frac{\Gamma(\alpha_1 + x_{t+1})}{x_{t+1}! (1 + \beta_1)^{\alpha_1 + x_{t+1}}} \frac{\beta^\alpha}{\Gamma(\alpha)} \int_0^\infty \frac{(1 + \beta_1)^{\alpha_1 + x_{t+1}}}{\Gamma(\alpha_1 + x_{t+1})} \lambda^{x_{t+1} + \alpha_1 - 1} e^{-\lambda(\beta_1 + 1)} d\lambda \quad (9)$$

Lo anterior lleva a

$$f_{X_{t+1}|X}(X_{t+1}|X_t) = \frac{\Gamma(\alpha_1 + x_{t+1})}{x_{t+1}! (1 + \beta_1)^{\alpha_1 + x_{t+1}}} \frac{\beta^\alpha}{\Gamma(\alpha)} \quad (10)$$

A partir de la función predictiva dada en (10) se simulará el valor futuro X_{t+1} , o pronóstico de las ventas para este caso, usando los siguientes pasos:

- Calcular la sumatoria y los valores del primer año ($n = 12$ valores).
- Actualizar la demanda modificando la sumatoria con la ampliación del conjunto de datos (incremento en n) y guardar estos para el siguiente paso hasta abarcar los 24 meses.
- Actualizar los parámetros $\alpha^{(i)}$ y $\beta^{(i)}$, con la cadena MCMC dada por el muestreador de Gibbs, usando las a posteriori respectivas.
- Realizar predicción con la función predictiva para encontrar el pronóstico X_{t+1} en cada mes (t) y medir el error relativo absoluto.
- Estimar cada error relativo y promediar los de todo el año (12) para obtener el MAPE de pronósticos de 2011, indicador usado para comparar la eficiencia de todos los métodos.

Pronóstico con el valor esperado

El valor esperado para la predicción bayesiana se deduce a partir de la a posteriori obtenida y es condicional a los datos pasados, así $E[X_{t+1} | x]$, donde X_{t+1} es el valor que se va a pronosticar en el periodo *expost* ($t+1$) y

x representa los datos pasados, lo que por definición se ilustra de manera analítica a continuación:

Distribución a priori para λ : Gamma (α, β) , haciendo un cambio en la forma de escritura: $1/\beta$ en lugar de β .

La distribución a posteriori para λ , asumiendo los parámetros α y β fijos, es:

$$\xi(\lambda|Datos) = \frac{\frac{1}{\Gamma(\alpha)\beta^\alpha} \lambda^{\alpha-1} e^{-\lambda/\beta} \frac{\lambda^{\sum_{i=1}^n x_i} e^{-n\lambda}}{\prod_{i=1}^n x_i!}}{\frac{1}{\prod_{i=1}^n x_i!} \Gamma(\alpha)\beta^\alpha \lambda^{\alpha+\sum_{i=1}^n x_i-1} e^{-\lambda/\beta-n\lambda}} \quad (11)$$

Donde

$$f_x(x) = \int_0^\infty \frac{1}{\prod_{i=1}^n x_i! \Gamma(\alpha)\beta^\alpha} \lambda^{\alpha+\sum_{i=1}^n x_i-1} e^{-\lambda(\frac{1}{\beta}+n)} d\lambda \quad (12)$$

$$f_x(x) = \frac{1}{\prod_{i=1}^n x_i! \Gamma(\alpha)\beta^\alpha} \int_0^\infty \lambda^{\alpha+\sum_{i=1}^n x_i-1} e^{-\lambda(\frac{1+\beta n}{\beta})} d\lambda \quad (13)$$

La integral de la ecuación (13) corresponde a la forma de una distribución Gamma, como se hizo en la sección anterior, agregando términos tanto en la integral como externamente; la expresión (13) se puede simplificar así:

$$f_x(x) = \frac{1}{\prod_{i=1}^n x_i! \Gamma(\alpha)\beta^\alpha} \Gamma\left(\alpha + \sum_{i=1}^n x_i\right) \left(\frac{\beta}{1 + \beta n}\right)^{\alpha+\sum_{i=1}^n x_i} \quad (14)$$

Quedando:

$$\xi(\lambda|Datos) = \frac{\frac{1}{\prod_{i=1}^n x_i! \Gamma(\alpha)\beta^\alpha} \lambda^{\alpha+\sum_{i=1}^n x_i-1} e^{-\lambda/\beta-n\lambda}}{\frac{1}{\prod_{i=1}^n x_i! \Gamma(\alpha)\beta^\alpha} \Gamma(\alpha + \sum_{i=1}^n x_i) \left(\frac{\beta}{1 + \beta n}\right)^{\alpha+\sum_{i=1}^n x_i}} \quad (15)$$

Así:

$$\xi(\lambda|Datos) = \frac{\lambda^{\alpha+\sum_{i=1}^n x_i-1} e^{-\lambda(\frac{1+\beta n}{\beta})}}{\Gamma(\alpha + \sum_{i=1}^n x_i) \left(\frac{\beta}{1 + \beta n}\right)^{\alpha+\sum_{i=1}^n x_i}} \quad (16)$$

La ecuación (16) tiene la forma de una distribución $\text{Gamma}\left(\alpha + \sum_{i=1}^n x_i, \frac{\beta}{1 + \beta n}\right)$. El valor esperado para la predicción bayesiana se basa en la distribución a posteriori para λ , como se muestra a continuación:

Valor esperado:

$$E(X_{t+1}|x) = \int_0^{\infty} E(x_{t+1}|\lambda) \xi(\lambda|\text{Datos})d\lambda \tag{17}$$

$$E(X_{t+1}|x) = \int_0^{\infty} \lambda \frac{\lambda^{\alpha + \sum_{i=1}^n x_i - 1} e^{-\lambda\left(\frac{1+\beta n}{\beta}\right)}}{\Gamma(\alpha + \sum_{i=1}^n x_i) \left(\frac{\beta}{1 + \beta n}\right)^{\alpha + \sum_{i=1}^n x_i}} d\lambda \tag{18}$$

$$E(X_{t+1}|x) = \frac{1}{\Gamma(\alpha + \sum_{i=1}^n x_i) \left(\frac{\beta}{1 + \beta n}\right)^{\alpha + \sum_{i=1}^n x_i}} \int_0^{\infty} \lambda^{1+\alpha + \sum_{i=1}^n x_i - 1} e^{-\lambda\left(\frac{1+\beta n}{\beta}\right)} d\lambda \tag{19}$$

En la ecuación (20) la integral tiene la forma de distribución Gamma; agregando términos de forma similar a la expresión (16) se tiene:

$$E(X_{t+1}|x) = \frac{1}{\Gamma(\alpha + \sum_{i=1}^n x_i) \left(\frac{\beta}{1 + \beta n}\right)^{\alpha + \sum_{i=1}^n x_i}} \Gamma\left(\alpha + \sum_{i=1}^n x_i\right) \left(\frac{\beta}{1 + \beta n}\right) \tag{20}$$

$$E(X_{t+1}|x) = \left(\alpha + \sum_{i=1}^n x_i\right) \left(\frac{\beta}{1 + \beta n}\right) \tag{21}$$

El valor esperado de la ecuación (21) constituye la base para elaborar el pronóstico de la demanda de cada mes. Con base en esta se realiza un método de optimización Tabú, que ha mostrado tener óptimos resultados en [17], [19]. Para esta técnica, los parámetros α_t y β_t se convierten en variables de decisión y se optimizan 24 valores, 12 para cada uno, correspondientes a los 12 meses del año 2011.

Pronóstico con parámetros autorregresivos del valor esperado

Los pronósticos incorporan valores actualizados hasta $t-1$ de los datos; por lo tanto, en esta sección se propone incorporar una modificación al valor esperado (21), agregando retardos de sus parámetros, los cuales se adicionan multiplicándolos por el coeficiente 0.1, el cual sale del análisis de autocorrelación de orden 1 que tienen los datos originales, por tanto, se toma como un efecto que puede causar dicho retardo en la estimación siguiente.

La ecuación (22) es el valor esperado modificado con dichos retardos:

$$E(X_{t+1}|x) = \left(\alpha_t + 0.1\alpha_{t-1} + \sum_{i=1}^{n_t} x_i \right) \left(\frac{\beta_t + 0.1\beta_{t-1}}{1 + (\beta_t n_t + 0.1\beta_{t-1})} \right) \quad (22)$$

Donde

X_{t+1} : valor que se va a pronosticar en el tiempo $t+1$.

n_t : cantidad de datos disponibles para pronosticar.

α_t : parámetro de forma para el período t y α_{t-1} es su retardo de orden 1.

β_t : parámetro de escala para el periodo t y β_{t-1} es su retardo de orden 1.

A partir de la ecuación (22), para el valor esperado bayesiano se aplica la técnica Tabú para encontrar los parámetros respectivos que minimizan el MAPE.

Metaheurística Tabú

Este es un método que busca optimizar el valor de una función objetivo $f(x)$ con n variables de decisión, de manera que se genere una lista con los valores de los puntos en el espacio $n+1$ dimensional (n variables y el valor de la función objetivo $f(x)$), llamada lista Tabú. Dicha lista representa la memoria del recorrido de la búsqueda, para evitar repetir los puntos evaluados previamente, hasta encontrar el mejor valor entre los puntos explorados. En este trabajo se propone una variación al presentado en [20], quienes se basan en una explotación de puntos dentro de un vecindario y una exploración para buscar la mejor solución dentro del espacio de posibles soluciones, que no garantiza encontrar el óptimo global.

La variación propuesta en este trabajo consiste en la exploración de puntos nuevos usando una combinación de una distribución uniforme y una normal, que no se propone en [20], de manera que se encuentren parámetros mensuales para minimizar el error de pronóstico de un producto terminado, aplicado a un caso empresarial. La lista Tabú se actualiza cada que se encuentra un punto que genera un valor de la función objetivo menor al mínimo de todos los demás que están en la lista Tabú.

Dicho algoritmo se aplica al método del valor esperado bayesiano de las ecuaciones (21) y (22), buscando minimizar el MAPE (ecuación 23).

$$MAPE = \frac{1}{N} \sum_{t=1}^N \left| \frac{e_t}{X_t} \right| \quad (23)$$

Donde $e_t = X_t - X_{t+1}$ es el error entre el valor real (X_t) y el pronóstico (X_{t+1}), es decir, la función objetivo es el promedio del error relativo absoluto de 12 pronósticos con respecto a los valores reales respectivos de 2011.

Variables de decisión: 24 parámetros: 12 para α_t y 12 para β_t ($t=1,2,\dots,12$).

El proceso de estimación usando el valor esperado en cualquiera de los dos casos mostrados en las ecuaciones (21) o (22) es el siguiente:

- Estimar un valor esperado inicial con la sumatoria y valor de n de los primeros 12 meses del año 2010 y asignar un vector inicial α_t y β_t .
- Actualizar la demanda del siguiente año (2011), cambiando la sumatoria y el valor de n en cada periodo de tiempo t o mes.
- Realizar la búsqueda estadística del vecindario de valores de la función objetivo, variando los 24 parámetros α_t y β_t , como variables de decisión, y almacenar en la lista del algoritmo Tabú los mejores puntos que minimizan el MAPE y posteriormente la depuran, tratando de encontrar un valor muy cercano al óptimo
- Estimar cada error relativo y promediar los de todo el año (12) para obtener el MAPE de pronósticos de 2011.

4. RESULTADOS Y DISCUSIÓN

En esta sección se detalla la prueba de bondad de ajuste para los datos, la generación de variable aleatoria y la aplicación de las ecuaciones y métodos bayesianos para encontrar los 12 pronósticos del año 2011, así como la medición del indicador MAPE respectivo a cada método, comparando los resultados para el caso de estudio.

Se estimó una prueba de bondad de ajuste a 12 valores de las ventas mensuales reales del producto elegido, del año 2010 (nombre reservado por confidencialidad). Al realizar la prueba Chi cuadrado, el Valor-P resultante es $0.3154 > 0.05$; lo anterior prueba que los datos se ajustan a una distribución Poisson con el parámetro lambda estimado λ . De acuerdo con este análisis, la distribución tiene una media de 100,17 y una desviación de 10,01.

Pronóstico con variable aleatoria

La distribución Poisson con $\lambda=100.17$ es una alternativa de pronóstico cuando no se cuenta con muchos datos históricos, mientras un modelo de regresión o un ARIMA no lo serían debido a la exigencia de una cantidad grande de datos.

Con fines de ilustración, al usar la distribución Poisson, luego de 1000 simulaciones, se encuentra un error relativo absoluto medio (MAPE para 12 meses) del 23.3 % con un rango de variación entre 18.7 y 27.5 %, como se muestra en la tabla 1, lo cual representa un alto error para efectos de una planeación de producción adecuada para este producto.

Tabla 1. Pronóstico con distribución Poisson (100.17)

Estadístico	Valor del MAPE
Min.	0,1865
1st Qu	0,2204
Median	0,2318
Mean	0,2328
3rd Qu	0,2455
Max.	0,2754

Fuente: Construcción propia.

Pronósticos con la función predictiva bayesiana

El proceso analítico de la técnica de inferencia predictiva bayesiana explicado previamente se aplica a los datos usando una programación en el lenguaje R, usando la ecuación (10). Proceso que finaliza con el pronóstico de 12 meses de demanda del producto para el año 2011, con los cuales se calculan los errores relativos absolutos, llevando al promedio: MAPE. Este indicador fue simulado 1000 veces, y se encontró un rango de variación entre 27,9 y 58,9 %.

Tabla 2. Pronóstico con función predictiva

Estadístico	Valor del MAPE
Min.	0,2788
1st Qu	0,3926
Median	0,4214
Mean	0,4244
3rd Qu	0,4573
Max.	0,5891

Fuente: Construcción propia.

La simulación basada en la ecuación (10) y el muestreador de Gibbs mostró que los pronósticos encontraron incorrelación en los valores simulados, lo cual muestra que hay convergencia; a pesar de esto y de acuerdo con su MAPE, no genera óptimos resultados para el caso de estudio. Se elaboró una prueba de bondad de ajuste en relación con el parámetro simulado λ , y se encontró la forma de la distribución Gamma, como originalmente se propuso, lo cual mostró eficiencia del algoritmo de muestreo diseñado.

Simulación de escenarios

Se simularon 12 valores de ventas, a partir de tres escenarios: procesos AR1 con correlaciones 0.1, 0.5 y 0.8, con la media de 100.17. El proceso se repitió 500 veces, y los resultados son muy similares al caso de estudio, acorde con la tabla 3.

Tabla 3. Simulaciones de Inferencia Predictiva en tres escenarios

Estadística	AR1(0.1)	AR1(0.5)	AR1(0.8)
Min.	29,32 %	27,22 %	28,26 %
1st Qu	40,34 %	40,62 %	40,51 %
Median	43,26 %	43,51 %	43,62 %
Mean	43,41 %	43,65 %	43,52 %
3rd Qu	46,28 %	46,83 %	46,67 %
Max.	60,26 %	58,44 %	56,70 %

Fuente: Construcción propia.

La media de errores absolutos relativos MAPE es muy similar en los tres casos.

Pronóstico del Valor esperado con optimización Tabú

En esta sección se utiliza el valor esperado bayesiano ilustrado en la ecuación (22). Es de anotar que sin emplear la metaheurística de optimización Tabú, el indicador MAPE de pronóstico oscila entre 17 y 20 %, pero cuando se utiliza los resultados mejoran significativamente, reduciendo el MAPE a 13.21 %, como se observa en la tabla 4, que además ilustra los resultados obtenidos para los parámetros α , β que mejor ajustan el pronóstico mensual.

Tabla 4. Pronóstico con Prima bayesiana

t	1	2	3	4	5	6	7	8	9	10	11	12	MAPE
α	183,4	320,5	30,91	264,1	294,4	263,8	280,2	248,1	95,84	284,5	259,5	257,8	13,21%
β	287,5	294,3	292,7	268,5	280,5	301,6	405,5	394,4	396	386,9	420,9	389,1	

Fuente: Construcción propia.

Simulación de escenarios

Se utilizaron los tres escenarios del caso anterior. Los resultados se presentan en la tabla 5. Se aprecia la mejora en el error medio al utilizar el método de optimización Tabú.

Tabla 5. Simulación de valor esperado original con método Tabú y sin este

Escenarios con valor esperado original, sin Tabú				Escenarios de valor esperado original con metaheurística Tabú		
Estadísticas	AR1(0.1)	AR1(0.5)	AR1(0,8)	AR1(0.1)	AR1(0.5)	AR1(0.8)
Min.	10,2 %	10,3 %	7,1 %	6,63 %	7,36 %	6,94 %
1st Qu	10,6 %	10,6 %	9,7 %	6,80 %	7,59 %	7,09 %
Median	10,7 %	10,8 %	10,0 %	6,97 %	7,65 %	7,23 %
Mean	10,8 %	10,9 %	10,0 %	6,99 %	7,67 %	7,19 %
3rd Qu	11,0 %	11,2 %	10,4 %	7,13 %	7,74 %	7,31 %
Max.	11,3 %	11,6 %	12,8 %	7,54 %	7,91 %	7,38 %

Fuente: Construcción propia.

En la siguiente subsección se muestra que al incorporar los retardos de cada parámetro el MAPE mejora aun más.

Parámetro autorregresivo del valor esperado con optimización Tabú

Al incorporar el retardo de orden 1 de cada parámetro, ecuación (22), el MAPE se reduce a 8,86 %, lo cual muestra una mejora considerable en los pronósticos con respecto a la metodología anterior.

Aunque este sea una aplicación específica, sí muestra que el algoritmo de búsqueda metaheurístico Tabú mejora mucho el desempeño del método bayesiano por sí solo, aun cuando se incorporan dependencias pasadas, lo cual da ventajas para el uso de esta técnica de pronósticos propuesta.

La tabla 6 ilustra los mejores resultados de los parámetros α y β para cada tiempo t encontrados con el algoritmo Tabú diseñado e incorporando el retardado de orden 1 en los parámetros, para el caso real.

Tabla 6. MAPE del valor esperado bayesiano con parámetros autorregresivos

t	1	2	3	4	5	6	7	8	9	10	11	12	MAPE
α	46,19	248,7	33,69	273,1	270,6	236,5	199,1	86,43	115,5	277,9	284,8	259,9	8,86%
β	320,9	150,5	371,6	289	248,6	245,6	347,5	385,8	479,1	370,8	389,6	364,3	

Fuente: Construcción propia.

Simulación de escenarios

Utilizando los mismos escenarios descritos previamente, los resultados de MAPE se aprecian en la tabla 7. De todos los resultados mostrados para pronosticar, estos son los mejores.

Tabla 7. Simulación usando valor esperado con parámetros autorregresivos y método Tabú

Estadísticas	AR1(0.1)	AR1(0.5)	AR1(0,8)
Min.	0,181 %	0,135 %	0,817 %
1st Qu	0,217 %	0,258 %	0,837 %
Median	0,234 %	0,311 %	0,857 %
Mean	0,234 %	0,280 %	0,913 %
3rd Qu	0,250 %	0,333 %	0,960 %
Max.	0,266 %	0,365 %	1,064 %

Fuente: Construcción propia.

Síntesis de resultados

El indicador que mide el error porcentual medio en el pronóstico (MAPE) para cada uno de los modelos aplicados se muestra en la tabla 8, tanto para el caso real de ventas del producto como para los casos simulados.

Tabla 8. Comparaciones del indicador MAPE, caso real y simulaciones

Escenarios	Pronóstico generación de variable aleatoria Poisson	Inferencia predictiva bayesiana	Valor esperado	Valor esperado con Tabú	Valor esperado con Parámetros autorregresivos y Tabú
Caso real	23.3 %	42.44 %	18 %	13.21 %	8.86 %
Sim.AR1(0.1)	20.01 %	43,41 %	10, 8%	6,99 %	0,234 %
Sim.AR1(0.5)	20.54 %	43,65 %	10, 9%	7,67 %	0,280 %
Sim.AR1(0.8)	21.18 %	43,52 %	10, 0%	7,19 %	0,913 %

Fuente: Construcción propia.

En la tabla 8 claramente se observa que el mejor resultado, tanto para el caso real estudiado como para los escenarios de simulación, es el valor esperado con retardos en los parámetros y uso de optimización Tabú, por obtener el menor error medio MAPE de 12 pronósticos.

5. DISCUSIÓN

Es de anotar que el método bayesiano debe actualizarse en cada tiempo t , así como el valor de la distribución a priori, de manera que se renueva su pronóstico a partir de información reciente; esto, sumado al uso del algoritmo Tabú, muestra para este caso un mejor acierto en el pronóstico de este producto terminado.

Extensiones de este trabajo pueden llevar a buscar híbridos de heurísticas de optimización en busca de un mejor acierto en la búsqueda de los parámetros, así como permite proponer otros modelos, empleando otras distribuciones, como la normal para el parámetro de la Poisson, o usar métodos de elicitación o, incluso, modelos de regresión bayesiana, o modelos lineales dinámicos en este campo.

CONCLUSIONES

Cuando existen pocos datos no puede usarse una técnica estadística clásica como un modelo ARIMA; si bien para algunos casos se usan solo distribuciones de probabilidad para generar variable aleatoria, esto no considera dependencias a corto o largo plazo o los impactos de una variación temporal, que lo hacen poco acertado, como lo muestra este caso de estudio.

Las técnicas propuestas en este trabajo alrededor de la teoría bayesiana proporcionan alternativas para pronosticar a corto plazo; pero al ser combinadas con el algoritmo Tabú muestran mejor desempeño, tanto para el caso real sobre las ventas del producto como para los escenarios simulados. Lo anterior sugiere que sí es posible proponer diferentes técnicas de pronóstico con relación a los modelos clásicos cuando existen pocos datos, con una precisión adecuada.

La metaheurística Tabú usada es una alternativa de optimización adecuada para una función objetivo que requiera alta precisión. Sin embargo, esta no garantiza un óptimo global.

Por otro lado, la programación de las técnicas bayesianas presentadas en este trabajo no requiere sistemas de cómputo muy complejos o largos periodos de simulación; estas pueden hacerse con el software libre R.

REFERENCIAS

- [1] D. Simchi-Levi, P. Kaminski, and E. Simchi-Levi, *Designing and Managing the Supply Chain*, 3rd ed., 2008, p. 498.
- [2] B. Sani and B. Kingsman, "Selecting the best periodic inventory control and demand forecasting methods for low demand items", *J. Oper. Res.*, vol. 48, n° 7, pp. 700-713, 1997.
- [3] V. Gutiérrez and C. Vidal, "Modelos de Gestión de Inventarios en Cadenas de Abastecimiento: Revisión de la Literatura", *Rev. Fac. Ing. la Univ.*, pp. 134-149, 2008.
- [4] H. Sarimveis, P. Patrinos, C. D. Tarantilis, and C. T. Kiranoudis, "Dynamic modeling and control of supply chain systems: A review", *Comput. Oper. Res.*, vol. 35, n° 11, pp. 3530-3561, Nov. 2008.
- [5] R. Watson, "The effects of demand-forecast fluctuations on customer service and inventory cost when demand is lumpy", *J. Oper. Res. Soc.*, vol. 38, n° 1, pp. 75-82, 1987.
- [6] F. Diebold, *Elementos de pronósticos*. México: International Thomson editores, 1999.
- [7] R. Cohen and F. Dunford, "Forecasting for Inventory Control: An Example of When 'Simple' Means 'Better'", *Interfaces (Providence)*, vol. 16, n° 6, pp. 95-99, 1986.
- [8] R. Neelamegham and P. Chintagunta, "A Bayesian model to forecast new product performance in domestic and international markets", *Mark. Sci.*, vol. 18, n° 2, pp. 115-136, 1999.
- [9] C. Pedroza, "A Bayesian forecasting model: predicting U.S. male mortality", *Biostatistics*, vol. 7, n° 4, pp. 530-550, Oct. 2006.
- [10] Q. Flora Lu, "Bayesian Forecasting of Stock Prices Via the Ohlson Model," WORCESTER POLYTECHNIC INSTITUTE, 2005. Thesis-Degree of Master of Science in Applied Statistics. Retrieved from https://www.wpi.edu/Pubs/ETD/Available/etd-050605-155155/unrestricted/Flora_Thesis_May_2005.pdf
- [11] P. M. Yelland, "Bayesian forecasting of parts demand", *Int. J. Forecast.*, vol. 26, n° 2, pp. 374-396, Apr. 2010.
- [12] J. Gill, *Bayesian Methods-A social and Behavioral Sciences Approach*, 2002, p. 459.
- [13] C. J. Barrera and J. C. Correa, "Distribución predictiva bayesiana para modelos de pruebas de vida vía MCMC"/"The Bayesian Predictive Distribution in Life Testing Models via", *Rev. Colomb. Estadística*, vol. 31, n° 2, pp. 145-155,

2008.

- [14] E. Flores, "Teoría de Credibilidad en el Cálculo de Reservas", in *Memorias Coloquio e Estadística*, 2012.
- [15] J. Harrison and C. Stevens, "Bayesian Forecasting", *J. R. Stat. Soc.*, vol. 38, n° 3, pp. 205-247, 1976.
- [16] A. Martin, K. Quinn, and J. H. Park, "MCMC pack: Markov Chain Monte Carlo in R", *J. Stat. Softw.*, vol. 42, n° 9, pp. 1-21, 2011.
- [17] A. Urrea and F. Torres, "Optimización de una política de inventarios por medio de búsqueda Tabú", in *III Congreso colombiano y I Conferencia Andina internacional*, 2006, p. 8.
- [18] E. A. Silver, "An overview of heuristic solution methods", *J. Oper. Res. Soc.*, vol. 55, n° 9, pp. 936-956, May 2004.
- [19] J. Van den Bergh, J. Beliën, P. De Bruecker, E. Demeulemeester, and L. De Boeck, "Personnel scheduling: A literature review", *Eur. J. Oper. Res.*, vol. 226, n° 3, pp. 367-385, May 2013.
- [20] R. Chelouah and P. Siarry, "A hybrid method combining continuous tabu search and Nelder-Mead simplex algorithms for the global optimization of multimimima functions", *Eur. J. Oper. Res.*, vol. 161, n° 3, pp. 636-654, March 2005.