Modelos de previsión para artículos con demanda intermitente

Mª Eugenia Babiloni Griñón, Manuel Cardós Carboneras,
José Miguel Albarracín Guillem, Marta E. Palmer Gato

Dpto. de Organización de Empresas. Universidad Politécnica de Valencia. Camino de Vera, s/n. 46022. Valencia.
mabagri@doctor.upv.es, mcardos@omp.upv.es, malbarr@omp.upv.es, marpalga@omp.upv.es.

Resumen

Dada la importancia de la previsión de demanda en la gestión de inventarios conocer qué métodos de previsión existen en la literatura permite escoger aquel que asegure el éxito del sistema y aun más si el patrón de demanda es intermitente. Dentro de los métodos de previsión tradicionales, el más extendido es el alisado exponencial simple y basado en él se desarrolla el método de Croston, el cual utiliza dos estimadores distintos e insesgados, uno para el intervalo entre demandas y otro para el tamaño de las mismas. Otras investigaciones se centran en deducir factores de corrección para el sesgo producido por los métodos tradicionales, como medias móviles y alisado exponencial, cuando se utilizan para prever demanda intermitente. Adicionalmente algunos modelos de previsión aplican el procedimiento de la estimación autosuficiente (bootstrap method) para ítems con patrones de demanda intermitente.

Palabras clave: previsión de la demanda, demanda intermitente

1. Introducción

La previsión de la demanda es un factor esencial en las decisiones a tomar dentro de un sistema de gestión de inventarios, e incluso en la determinación del mismo. Los errores cometidos en la previsión tienen un gran impacto sobre la determinación del tamaño óptimo de lote, sobre el nivel de servicio al cliente, sobre los costes totales del sistema de gestión de inventarios o sobre el tiempo medio de retraso esperado en servir los pedidos aplazados por falta de existencias. Si además la demanda se caracteriza como intermitente, es decir, si aparece variabilidad tanto en el tamaño de las órdenes de demanda como en el intervalo entre períodos con demanda no nula, encontrar el mejor procedimiento de previsión es básico para asegurar la eficacia del sistema de gestión de inventarios.

El modelo de previsión más utilizado para patrones de demanda intermitente es el alisado exponencial simple. Sin embargo, este método: (i) subestima sistemáticamente el valor de la demanda para cada periodo con demanda positiva; (ii) sobreestima el valor de la demanda media. Es lógico pensar que (i) y (ii) ocurran al utilizar un método de previsión que asigna un mayor peso a los datos recientes. Sin embargo, ante un patrón de demanda intermitente, donde no se produce demanda en los $n-1$ periodos anteriores, la previsión dada para el periodo n será siempre inferior a la demanda real, lo que explica (i). Por el mismo motivo, la media estimada de la demanda en todos los periodos considerados es superior a la real, puesto que en la estimación se está considerando demanda en todos los periodos.

En el presente artículo se realiza una revisión de los distintos modelos diseñados para realizar la previsión de ítems con patrón de demanda intermitente, clasificados como paramétricos (sección 2) y no paramétricos (sección 3). Por último se expondrán brevemente las conclusiones alcanzadas así como las posibles líneas futuras de investigación.

2. Modelos de previsión paramétricos

2.1. Método de Croston

Croston (1972) presenta un método de previsión para ítems con demanda intermitente basado en el alisado exponencial simple, pero introduciendo dos estimadores diferentes, insesgados, para calcular la previsión, por un lado, del tamaño de la demanda y por otro, del intervalo entre demandas no nulas. Ambos estimadores sólo se actualizan cuando un episodio de demanda tiene lugar, por lo tanto, el método de Croston es idéntico al alisado exponencial simple cuando se produce demanda en cada periodo.

Dado el modelo:

$$y_t = x_t(\bar{z}_{\eta-1} + e_\eta), \tag{1}$$

donde:

- t hace referencia al instante de revisión y η a los instantes de demanda distinta de cero,
- la ocurrencia de demanda x_t en un intervalo específico, sigue un proceso de Bernoulli, con una probabilidad de ocurrencia de demanda de *1/p*,
- las observaciones de demanda distintas de cero z_η, se distribuye siguiendo un proceso ARIMA (0,1,1)[Box and Jenkins (1970)]. Además z_η es *i.i.d* y sigue una distribución normal N(μ,σ^2),
- se considera que existen, de media, p periodos entre dos demandas consecutivas,
- se introduce una variable q_t para medir el lapso de tiempo desde la última observación no nula para estimar el intervalo entre dos demandas consecutivas. La variable q_t sigue una distribución geométrica de media p y varianza $(p-1)^2$, y es *i.i.d.,*
- el tamaño de las órdenes de demanda y el intervalo entre demandas no nulas no están correlacionados,
- La constante de alisado α toma valores comprendidos entre 0.1 y 0.2 y es común a ambos estimadores, y por último,
- se asume que el error de estimación e_η sigue una distribución normal N(0,σ^2), entonces:

$$\begin{aligned}
&\text{Si } y_t = 0,\\
&\quad \bar{z}_\eta = \bar{z}_{\eta-1}\\
&\quad \bar{p}_\eta = \bar{p}_{\eta-1}\\
&\quad q_t = q_{t-1} + 1\\
&\text{Si } y_t \neq 0,\\
&\quad \bar{z}_\eta = \bar{z}_{\eta-1} + \alpha\left(y_t - \bar{z}_{\eta-1}\right)\\
&\quad \bar{p}_\eta = \bar{p}_{\eta-1} + \alpha\left(q_{t-1} - \bar{p}_{\eta-1}\right)\\
&\quad q_t = 1
\end{aligned} \tag{2}$$

Combinando el resultado de las dos estimaciones, se obtiene la estimación de la demanda media por periodo como:

$$\bar{y}_t = \bar{z}_\eta \big/ \bar{p}_\eta \tag{3}$$

Según Croston el valor esperado de la estimación, inmediatamente después de producirse la demanda es insesgado

$$E\left(\bar{y}_t\right) = \mu / p, \tag{4}$$

con una varianza

$$\operatorname{var}\left(\bar{y}_t\right) = \frac{\alpha}{2-\alpha}\left[\frac{(p-1)^2}{p^4}\mu^2 + \frac{\sigma^2}{p^2}\right], \tag{5}$$

menor que la varianza producida por un método de alisado exponencial convencional.

En Willemain et al. (1994) se lleva a cabo una comparación entre el método de Croston y un alisado exponencial simple para distintos escenarios que violan las hipótesis iniciales hechas por Croston. Del estudio se deduce que, para obtener beneficios de la utilización de este método, se requiere cierto grado de intermitencia en el patrón de demanda. Johnston y Boylan (1996) demuestran que, cuando el intervalo entre demandas no nulas excede en 1.25 veces el intervalo de revisión utilizado para ítems con patrones de demanda no intermitente, empieza a ser aconsejable utilizar el método de Croston.

2.2. Correcciones y modificaciones sobre el método de Croston

De las experiencias llevadas a cabo por Willemain et al. (1994) y Sani y Kingsman (1997) se deduce que, a pesar de que el método de Croston es teóricamente superior a otros métodos de previsión convencionales, los resultados de su aplicación en datos reales no son muy positivos. Syntetos y Boylan (2001) en un intento por encontrar las causas de este inesperado comportamiento encuentran un error matemático en el procedimiento desarrollado por Croston, donde se calcula la demanda media estimada por periodo como en (4), es decir supone que

$$E\left(\bar{y}_t\right) = E\left(\frac{\bar{z}_\eta}{\bar{p}_\eta}\right) = \frac{E\left(\bar{z}_\eta\right)}{E\left(\bar{p}_\eta\right)} = \frac{\mu}{p}, \tag{6}$$

sin embargo $E\left(\frac{z'_t}{p'_t}\right) = E\left(z'_t\right)E\left(\frac{1}{p'_t}\right)$ pero $E\left(\frac{1}{p'_t}\right) \neq \frac{1}{E\left(p'_t\right)}$.

Este error introduce un sesgo que se cuantifica en Syntetos y Boylan (2005) como:

$$\frac{\alpha}{2-\alpha}\mu\frac{(p-1)}{p^2} \tag{7}$$

e introducen un nuevo estimador para la demanda media que cumple (4):

$$\bar{y}_t = \left(1-\frac{\alpha}{2}\right)\frac{\bar{z}_\eta}{\bar{p}_\eta} \tag{8}$$

El método de Croston modificado por Syntetos y Boylan (2005), conocido también como método aproximado, del inglés "Approximation method" [Eaves y Kingsman (2004)], es insesgado puesto que elimina el error sistemático detectado en el método de Croston.

Continuando con este razonamiento, Leven y Segerstedt (2004) proponen otra modificación del método de Croston para evitar el error sistemático encontrado por Syntetos y Boylan (2001). Además introducen la utilización de una distribución en Erlang para modelar el tamaño de las órdenes de demanda (en lugar de la normal que sugiere Croston). Leven y Segerstedt (2004) proponen el siguiente estimador para la demanda media:

$$\bar{y}_t = \alpha\left(\frac{z_\eta}{p_\eta}\right) + (1-\alpha)\bar{y}_{t-1} \tag{9}$$

Sin embargo en Boylan y Syntetos (2007) se demuestra que, para un proceso de Bernoulli estacionario, el valor esperado de (9) es:

$$E(\bar{y}) = -\frac{E(\bar{z})}{E(\bar{p})-1}\log\left(\frac{1}{E(\bar{p})}\right) \tag{10}$$

y no $E(\bar{y}) = \dfrac{E(\bar{z})}{E(\bar{p})}$,

luego el método propuesto por Leven y Segerstedt (2004) no sólo no elimina el sesgo del método de Croston, sino que según Boylan y Syntetos (2007) genera un error sistemático aun mayor.

Schultz (1987) adapta a su modelo de gestión de inventarios un método de previsión que es, en esencia, como el método de Croston, con la particularidad de que utiliza únicamente los dos estimadores propuestos por Croston, para el intervalo entre demandas y para el tamaño de las mismas, pero con dos constantes de alisado distintas. Al no estimar la demanda media por periodo, no se introduce el sesgo detectado en Syntetos y Boylan (2001).

Por último, según Snyder (2002) el método propuesto por Croston indica implícitamente que la varianza del error cometido σ^2 así como la proporción de periodos activos p no varían con el tiempo, y propone definir estos valores como:

$$\widehat{\sigma}^2 = \sum_{t=1}^{n} e_t^2 \Big/ \sum_{t=1}^{n} x_t \quad \text{y} \quad \widehat{p} = \sum_{t=1}^{n} x_t \Big/ n \tag{11}$$

recomendando estimar las distribuciones de (11) mediante un procedimiento de estimación autosuficiente.

2.3. Modelos basados en la corrección del sesgo

En Shale et al. (2006) se recupera un enfoque más tradicional lo que supone utilizar métodos de previsión convencionales como el de medias móviles o alisado exponencial simple pero introduciendo un factor de corrección para eliminar el sesgo que estos métodos producen ante patrones de demanda intermitente. Además se asume que la llegada de órdenes de demanda sigue un proceso de Poisson distribuida según una gamma. Por último, las series se asimilan como localmente estacionarias.

Sea,

x	duración media de los últimos intervalos de llegada de órdenes
$v = 1/x$	estimación del número medio de órdenes por periodo

z	tamaño medio de las últimas órdenes de demanda
$w=z/x$	estimación de la demanda

Para que la estimación realizada como w sea insesgada es necesario introducir un factor de corrección. _Se considera que la estimación de la duración media de los intervalo de llegada de órdenes x es la media aritmética de los últimos k valores (medias móviles de orden k). Siendo $t_1, t_2, ..., t_k$ los intervalos entre las k últimas demandas no nulas, la distribución resultante de la suma de esas k distribuciones (gamma) así como su media aritmética x, tienen una función de distribución de Erlang (equivalente a una distribución gamma con parámetro de forma k, entero positivo), como sigue:

$$f(x)=\frac{\lambda^k x^{k-1} e^{-\lambda x}}{(k-1)!}, \quad donde \quad \lambda = k/p \tag{12}$$

El número medio de órdenes por periodo, $v = 1/x$, tiene por tanto una distribución inversa a una de Erlang, que según Eaveset. al (2000) se puede expresar como:

$$f(v)=x^2 f(x)=\frac{\lambda^k x^{k+1} e^{-\lambda x}}{(k-1)!} \tag{13}$$

cuya media es:

$$Media(v)=\frac{\lambda}{k-1} \tag{14}$$

Si se usa la *Media(v)* como estimador de *1/p*, se introduce un sesgo que es corregido mediante un factor denominado ratio de la inversa de p y la *Media(v)*. Recordemos que, de la expresión (13) se deduce que $1/p = \lambda/k$, por lo tanto

$$Factor\ de\ correccion = (1/m_T)/Media(v) = \frac{k-1}{k} \tag{15}$$

Este factor puede usarse para eliminar el sesgo producido mediante la estimación realizada como $w=z/x$ cuando se utilizan medias móviles como método de previsión.

El alisado exponencial simple suele ser tratado de forma equivalente a las medias móviles cuando la media estimada por ambos es la misma Brown (1962). Esta relación se establece entre la constante de alisado α, y el orden de las medias móviles k:

$$k = \frac{2-\alpha}{\alpha} \tag{16}$$

Sustituyendo la expresión (16) en la (15) se obtiene el factor de corrección del sesgo para un alisado exponencial, en función de α:

$$Factor\ correccion = 1 - \frac{\alpha}{2-\alpha} \tag{17}$$

En Shale et al. (2006) se demuestra que, para patrones de demanda intermitente, según aumenta la constante de alisado, α, y disminuye el orden de las medias móviles, *k,* los métodos de previsión tradicionales son más sesgados y el factor de corrección más impreciso.

3. Métodos de previsión no paramétricos

3.1. La estimación autosuficiente (bootstraping method)

Este método, desarrollado por Efron (1979), consiste en calcular directamente la varianza del estimador considerando la muestra como la población total. El método procede como sigue:

1-Sean $x_1,....,x_n$ los valores muestrales. Se considera la muestra como la población de una variable que toma los valores muestrales con una probabilidad de *1/n*. Se toman muestras con reemplazamiento de tamaño *n* mediante el método de Montecarlo. Se denomina $y_1,....,y_n$ a los valores de estas muestras.

2-Se calcula para cada muestra el estimador θ ($y_1,....,y_n$) resultado de aplicar el proceso de cálculo de θ a los valores particulares resultado de la extracción con reemplazamiento.

3-Repitiendo el proceso un número *B* elevado de veces, obtendremos una distribución de valores de θ cuya media y varianza:

$$\theta_T = \frac{1}{B}\sum\theta_i \tag{18}$$

$$\text{Var}(\theta) = \frac{1}{B}\sum(\theta_i - \theta_T)^2 \tag{19}$$

Puede demostrarse que, en muchos casos, este método obtiene asintóticamente la varianza del estimador θ.

3.2. Aplicación del método de estimación autosuficiente en previsión

Algunos autores utilizan en el desarrollo de sus métodos de previsión la estimación autosuficiente. Por ejemplo, en Snyder (2002) se utiliza para calcular (11).

Willemain et al. (2004) desarrollan un nuevo método de previsión que utiliza la estimación autosuficiente en uno de sus pasos. El método, calcula la previsión de demanda para cada periodo incluido en el periodo de aprovisionamiento y sumando estas previsiones obtiene la previsión de demanda para el periodo de aprovisionamiento completo. Los autores asumen que existe cierta correlación entre los periodos de demanda no nula y aquellos con demanda nula. Para modelar esta correlación utilizan un proceso de Markov de primer orden.

Tabla 1. Resumen método de previsión de Willemain et al. (2004) para ítems con demanda intermitente.

PASO 1-Se obtienen la serie de datos histórica y se periodiza en la unidad de tiempo seleccionada (semanas, días, meses, etc.).
PASO 2-Se calculan las probabilidades de transición de demanda, es decir las probabilidades de que se produzca una demanda no nula cuando se ha producido una demanda nula, etc.

PASO 3-Se realiza la previsión de la secuencia de periodos de demanda nula y demanda no nula. Estas previsiones están condicionadas a si en el periodo en que se realiza la previsión, la demanda es o no nula. Por lo tanto, es necesario el cálculo de las probabilidades de transición (paso 2).
PASO 4-Se calcula, para aquellos instantes en los que se ha previsto demanda, la previsión del tamaño de la misma. Para ello se utiliza el procedimiento de estimación autosuficiente con algunas modificaciones (ver paso 5).
PASO 5-Para evitar la falta de variabilidad de la estimación de la población considerada, es decir, a los datos históricos de demanda, una vez seleccionados, se les aplica un operador para distorsionarlos. A este operador se le denomina JITTERED y opera como sigue: Sea X^* una de las muestras de demanda seleccionada aleatoriamente de los datos históricos de la demanda. Sea Z la desviación estándar normal de la muestra, el proceso se describe como: $$\text{JITTERED} = 1 + \text{INT}\left\{X^* + Z\sqrt{X^*}\right\}$$ $$\text{IF JITTERED} = 0,\ \text{THEN JITTERED} = X^*$$
PASO 6-Se suman las previsiones de demanda que se producen dentro del periodo de aprovisionamiento para calcular la demanda estimada total en dicho periodo, es decir la “Lead Time Demand” (LTD).
PASO 7-Se repiten los pasos 2-6 un número grande de veces.
PASO 8-Se clasifica y utiliza la distribución de LTD obtenida a través del procedimiento de estimación autosuficiente.

Los autores contrastan este método con el método de Croston y con el alisado exponencial simple, concluyendo que es el más preciso de los tres. Sin embargo, la precisión disminuye conforme aumenta el periodo de aprovisionamiento, lo cual no ocurre con los otros dos métodos debido a su hipótesis de normalidad y al teorema central del límite.

3.3. Nuevas líneas de investigación

Bao et al. (2004); Bao et al. (2006) desarrollan un método de previsión basado en un novedoso algoritmo de redes neuronales llamado “Support vector machines (SVMs)”. Los autores comparan los resultados obtenidos con éste método y con el método de Croston, y concluyen que la utilización de SVMs, aunque todavía en proceso de investigación, es prometedora.

4. Conclusiones

El método de Croston es el más extendido para calcular previsiones de demanda cuando ésta se califica como intermitente, tanto por su robustez, por su modo de operar y por su aplicación a cualquier tipo de patrón de demanda. Sin embargo, se aconseja su utilización si el intervalo medio entre demandas *p* es superior a 1.25 veces el periodo de revisión. A pesar de que se comete un error sistemático en la estimación de la demanda media por periodo, para valores pequeños de la constante de alisado dicho error es muy pequeño.

El alisado exponencial simple y el método de Croston asumen normalidad de la demanda lo cual: (i) para patrones de demanda intermitente no permite beneficiarse de los efectos del teorema central del límite; (ii) puede llevar a prever demanda negativa para coeficientes de variación de la serie grandes, muy común por otro lado para este tipo de patrones; (iii) es simétrica, mientras que los patrones de demanda intermitente suelen presentar asimetría. Además el método de Croston asume que la demanda es *i.i.d.,* lo cual no tiene por qué ser cierto en patrones de demanda intermitente. Con todo ello las líneas futuras de investigación se encaminan a investigar la aplicación de funciones de distribución más adecuadas al patrón de demanda, además de estudiar su aplicación al caso más general de demanda correlacionada.

Por último, Willemain et al. (2004) utilizan la estimación autosuficiente para calcular la distribución de la demanda durante el periodo de aprovisionamiento, además de asumir cierta correlación entre el intervalo de demandas no nulas. Sin embargo, cualquier procedimiento de estimación autosuficiente asume que la muestra de datos es estacionaria o puede ser transformada en estacionaria con cierta facilidad, lo cual debe ser comprobado a priori.

Referencias

Bao, Y. K.; Wang, W.; Zhang, J. (2004). Forecasting Intermittent Demand by SVMs Regression. 2004 IEE International Conference on Systems, Man and Cibernetics.

Bao, Y. K.; Zou, H.; Liu, Z. T. (2006). Forecasting intermittent demand by fuzzy support vector machines. Advances in Applied Artificial Intelligence, Proceedings, Vol.4031, No 1080-1089.

Box and Jenkins, G. M. (1970). Time Series Analysis Forecasting and Control. Holden-Day.

Boylan, J. E.; Syntetos, A. A. (2007). The accuracy of a Modified Croston procedure. International Journal of Production Economics, Vol.107, No 511-517.

Brown, R. G. (1962). Smoothing Forecasting and Prediction of Discrete Time Series. Prentice-Hall.

Croston, J. D. (1972). Forecasting and Stock Control for Intermittent Demands. Operational Research Quarterly, Vol.23, No 3, pp. 289-303.

Eaves, A. H. C.; Hastings, N.; Peacock, B. (2000). Statistical Distributions. 3rd edn. Wiley:New York.

Eaves, A. H. C.; Kingsman, B. G. (2004). Forecasting for the ordering and stock-holding of spare parts. Journal of the Operational Research Society, Vol.55, No 4, pp. 431-437.

Efron, B. (1979). Bootstrap Methods - Another Look at the Jackknife. Annals of Statistics, Vol.7, No 1, pp. 1-26.

Johnston, F. R.; Boylan, J. E. (1996). Forecasting for items with intermittent demand. Journal of the Operational Research Society, Vol.47, No 1, pp. 113-121.

Leven, E.; Segerstedt, A. (2004). Inventory control with a modified Croston procedure and Erlang distribution. International Journal of Production Economics, Vol.90, No 3, pp. 361-367.

Sani, B.; Kingsman, B. G. (1997). Selecting the best periodic inventory control and demand forecasting methods for low demand items. Journal of the Operational Research Society, Vol.48, No 7, pp. 700-713.

Schultz, C. R. (1987). Forecasting and Inventory Control for Sporadic Demand Under Periodic Review. Journal of the Operational Research Society, Vol.38, No 5, pp. 453-458.

Shale, E. A.; Boylan, J. E.; Johnston, F. R. (2006). Forecasting for intermittent demand: the estimation of an unbiased average. Journal of the Operational Research Society, Vol.57, No 5, pp. 588-592.

Snyder, R. (2002). Forecasting sales of slow and fast moving inventories. European Journal of Operational Research, Vol.140, No 3, pp. 684-699.

Syntetos, A. A.; Boylan, J. E. (2001). On the bias of intermittent demand estimates. International Journal of Production Economics, Vol.71, No 1-3, pp. 457-466.

Syntetos, A. A.; Boylan, J. E. (2005). The accuracy of intermittent demand estimates. International Journal of Forecasting, Vol.21, No 2, pp. 303-314.

Willemain, T. R.; Smart, C. N.; Schwarz, H. F. (2004). A new approach to forecasting intermittent demand for service parts inventories. International Journal of Forecasting, Vol.20, No 3, pp. 375-387.

Willemain, T. R.; Smart, C. N.; Shockor, J. H.; Desautels, P. A. (1994). Forecasting Intermittent Demand in Manufacturing - A Comparative-Evaluation of Croston Method. International Journal of Forecasting, Vol.10, No 4, pp. 529-538.