

Comparación Empírica de Varios Métodos para Estimar los Parámetros de la Distribución Lognormal con Traslado

David F. Muñoz⁽¹⁾, César Ruiz⁽²⁾ y Sven Guzman⁽²⁾

(1) Dpto. de Ingeniería Industrial y de Operaciones, Instituto Tecnológico Autónomo de México, Río Hondo 1, 01080 México, Distrito Federal, México. (e-mail: davidm@itam.mx)

(2) Dpto. de Ingeniería de Negocios, Escuela Superior de Economía y Negocios, Santa Tecla, La Libertad, El Salvador (e-mail: cesarruiztorres@gmail.com, sguzman@esen.edu.sv)

Recibido Oct. 30, 2015; Aceptado Ene. 5, 2016; Versión final Ene. 19, 2016, Publicado Jun. 2016

Resumen

En este artículo se presentan resultados experimentales sobre la comparación de cinco metodologías para estimar los parámetros de la distribución lognormal con traslado. Los métodos considerados son el método de momentos, el método de máxima verosimilitud local, un método propuesto por Nagatsuka y Balakrishnan, una modificación de este último, y un nuevo método propuesto en este artículo. El método de máxima verosimilitud local proporcionó estimadores que tienen buenas propiedades cuando el tamaño de la muestra no es muy pequeño y el método no falla. Por otro lado, el método de Nagatsuka y Balakrishnan modificado tuvo el mejor desempeño desde el punto de vista de los errores cuadráticos promedio de los estimadores. El nuevo método propuesto en este trabajo mostró el mejor desempeño desde el punto de vista de una medida de bondad de ajuste.

Palabras clave: estimación de parámetros; ajuste de distribuciones; distribución lognormal; análisis de la entrada

Empirical Comparison of Several Methods for Parameter Estimation of a Shifted Lognormal Distribution

Abstract

This article presents the results of the comparison done using five methodologies for the fitting of parameters of a shifted lognormal distribution. The methods considered in this research are: the method of moments, the method of local maximum likelihood, a method proposed by Nagatsuka and Balakrishnan, a modification of this latter method, and a new method proposed in this article. The method of local maximum likelihood provided estimators with good properties when the sample size is not too small and the method did not fail. On the other hand, the modified method of Nagatsuka and Balakrishnan exhibited the best performance from the point of view of the mean squared error. The new method proposed in this work exhibited the best performance from the point of view of a measure of goodness of fit.

Keywords: parameter estimation; distribution fitting; lognormal distribution; input analysis

INTRODUCCIÓN

El ajuste de distribuciones es una de las herramientas clásicas para modelar un componente aleatorio de la entrada de una simulación estocástica y, debido a su gran flexibilidad, la distribución lognormal es una de las distribuciones más usadas para ajustar un conjunto de datos univariados. La inclusión de un parámetro de traslado (a) es muy conveniente para permitir un mejor ajuste de los datos a la distribución propuesta, habiéndose reportado aplicaciones en diversas áreas; por ejemplo, para modelar el diámetro de árboles de un rodal (Vallejos y Aedo, 2010) o para modelar caudales máximos anuales (Raynal y García, 2005), y es particularmente útil en el análisis de la entrada de experimentos por simulación (ver, e.g., Muñoz et al., 2015). Debido a estas razones, la distribución lognormal (con parámetro de traslado) tiene particular importancia para el análisis de la entrada de experimentos por simulación. La función de densidad de la distribución lognormal con traslado depende tres parámetros (a, μ, σ^2) y está dada por:

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma(x-a)} \exp\left(-(\ln(x-a)-\mu)^2 / 2\sigma^2\right), \quad (1)$$

para $x > a$, donde $\sigma^2 > 0$. En lo sucesivo denotaremos a la distribución que corresponde a esta función de densidad como $\text{logN}(a, \mu, \sigma^2)$.

Debido a que si una variable aleatoria X se distribuye como lognormal con parámetros a, μ, σ^2 , entonces $\ln(X-a)$ se distribuye normalmente con media μ y varianza σ^2 . Cuando a es conocido los estimadores máximo verosímiles de μ y σ^2 se estiman fácilmente a partir de datos muestrales x_1, \dots, x_n , tomando los correspondientes logaritmos. Sin embargo, cuando a es desconocido, Hill (1963) mostró que los estimadores máximo verosímiles pueden no existir, debido a que la función de verosimilitud no tiene un máximo global (tiende a ∞ para una secuencia apropiada de valores de los parámetros), y es por dicha razón que el problema de estimación de parámetros de la distribución lognormal con traslado ha sido tratado extensamente en la literatura sobre ajuste de distribuciones (ver, por ejemplo, Nagatsuka y Balakrishnan, 2012 y sus referencias).

Nagatsuka y Balakrishnan (2012) han propuesto un método de estimación basado en la verosimilitud de los datos transformados

$$w_i = \frac{x_i - X(1)}{X(n) - X(1)}, \quad (2)$$

para $i = 1, \dots, n$, donde $X(1) \leq X(2) \leq \dots \leq X(n)$ denotan los datos x_1, \dots, x_n de la muestra aleatoria ordenados de menor a mayor. Los autores muestran que su método no solamente proporciona estimadores bien definidos sino que tienen buenas propiedades, como sesgo y error cuadrático promedios (ECP) relativamente bajos, aunque tiene el inconveniente de requerir de la evaluación numérica de una integral indefinida doble. Métodos basados en la misma transformación (2) han sido propuestos para estimar los parámetros de la distribución de Weibull con traslado (Nagatsuka et al., 2013) y de la distribución inversa de Gauss (Nagatsuka y Balakrishnan, 2013).

En este artículo proponemos una ligera modificación del método de Nagatsuka y Balakrishnan (MNB) para mejorar los sesgos de los estimadores, así como un segundo método que requiere de la evaluación de una integral indefinida simple, por lo que es más rápido que el MNB. Además, se discuten resultados experimentales sobre la comparación del segundo método propuesto con el MNB modificado y con los conocidos métodos de estimación por momentos y por máxima verosimilitud local (Cohen y Whitten, 1980). Al final del artículo se presentan las conclusiones y recomendaciones.

MÉTODOS DE ESTIMACIÓN

Con el objetivo de describir con claridad los métodos de estimación que se han implantado y probado en la presente investigación, en esta sección presentaremos brevemente, y para el caso de estimación de parámetros de la distribución lognormal con traslado, el método de estimación por momentos (MMO), el método de máxima verosimilitud local (ML), el método de Nagatsuka y Balakrishnan (MNB), el MNB modificado (MMNB) y el nuevo método propuesto (NM).

Método de Momentos

Como se describe en Singh et al. (1990), el MMO para la distribución lognormal se puede implantar a partir de las siguientes ecuaciones para la esperanza (μ_X), varianza (σ_X^2) y coeficiente de asimetría (G_X) de una $\text{logN}(a, \mu, \sigma^2)$:

$$\mu_X = a + w^{1/2} e^\mu, \sigma_X^2 = w(w-1)e^{2\mu}, G_X = (w+2)(w-1)^{1/2}, \quad (3)$$

donde $w = \exp(\sigma^2)$. Las ecuaciones (3) se pueden resolver para a , μ , σ^2 , obteniéndose los estimadores

$$\hat{\mu}_M = \frac{1}{2} \ln \left(\frac{\hat{\sigma}_X^2}{\hat{w}(\hat{w}-1)} \right), \hat{\sigma}_M^2 = \ln(\hat{w}) \hat{a}_M = \hat{\mu}_X - \frac{\hat{\sigma}_X}{\theta}, \quad (4)$$

donde

$$\theta = \frac{1-B^{2/3}}{B^{1/3}}, B = \frac{(\hat{G}_X^2 + 4)^{1/2} - \hat{G}_X}{2}, \hat{w} = \theta^2 + 1, \quad (5)$$

y los estimadores de la esperanza, varianza y coeficiente de asimetría se obtienen a partir de los datos x_1, x_2, \dots, x_n de la muestra aleatoria de acuerdo con las siguientes ecuaciones

$$\hat{\mu}_X = \frac{\sum_{i=1}^n x_i}{n}, \hat{\sigma}_X^2 = \frac{\sum_{i=1}^n x_i^2 - n\hat{\mu}_X^2}{n-1}, \hat{G}_X = \frac{n \left(\sum_{i=1}^n x_i^3 - 3\hat{\mu}_X \sum_{i=1}^n x_i^2 + 2n\hat{\mu}_X^3 \right)}{(n-1)(n-2)\hat{\sigma}_X^3}. \quad (6)$$

Como es bien conocido, el MMO proporciona estimadores consistentes, aunque en este caso puede proporcionar un estimador \hat{a}_M que es mayor que $x_{(1)}$ (el mínimo valor de la muestra), por lo que debe ser rechazado y decimos que el MMO no pudo estimar los parámetros de la distribución lognormal.

Método de Máxima Verosimilitud Local

Como es muy conocido, para aplicar el método de máxima verosimilitud, deben encontrarse los estimadores que maximizan la función de verosimilitud o, equivalentemente, el logaritmo de la función de verosimilitud. En el caso de la $\text{logN}(a, \mu, \sigma^2)$, puede verificarse a partir de (1) que el logaritmo de la función de verosimilitud tiene la forma

$$l(a, \mu, \sigma^2; x_1, \dots, x_n) = -\frac{n \ln(2\pi)}{2} - \sum_{i=1}^n \ln(x_i - a) - \frac{n \ln(\sigma^2)}{2} - \frac{\sum_{i=1}^n [\ln(x_i - a) - \mu]^2}{2\sigma^2}. \quad (7)$$

Como se mencionó anteriormente, la función definida en (7) puede no tener un máximo global, por lo que Cohen y Whitten (1980) proponen utilizar estimadores que cumplen las condiciones de optimización de primer orden. Al igualar a cero las derivadas parciales la función definida en (7), se puede comprobar fácilmente que las condiciones de optimización de primer orden proporcionan estimadores que satisfacen las ecuaciones

$$\hat{\mu}_L = \frac{\sum_{i=1}^n \ln(x_i - \hat{a}_L)}{n}, \hat{\sigma}_L^2 = \frac{\sum_{i=1}^n [\ln(x_i - \hat{a}_L) - \hat{\mu}_L]^2}{n}, \sum_{i=1}^n \frac{1}{x_i - \hat{a}_L} + \frac{1}{\hat{\sigma}_L^2} \sum_{i=1}^n \frac{\ln(x_i - \hat{a}_L) - \hat{\mu}_L}{x_i - \hat{a}_L} = 0. \quad (8)$$

Si no existen estimadores que cumplen las condiciones (8) decimos que el ML no pudo estimar los parámetros de la distribución lognormal y, por otro lado, si existe más de una solución, Singh et al. (1990) sugieren tomar la solución que tiene \hat{a}_L más cercana a $x_{(1)}$.

Método de Nagatsuka y Balakrishnan

Como se mencionó anteriormente, el MNB considera la transformación de los datos definida en (2), mediante la cual la densidad conjunta de W_1, W_2, \dots, W_n depende sólo del parámetro σ^2 , por lo que la función de verosimilitud de los datos transformados resulta

$$L(\sigma^2; w_1, \dots, w_n) = \frac{n!}{(2\pi)^{n/2}} \sigma^{-n} \int_0^\infty \int_0^\infty v^{n/2} \prod_{i=1}^n (u + vw_i)^{-1} \exp\left(-\frac{1}{2\sigma^2} \sum_{i=1}^n \ln(u + vw_i)^2\right) dv du, \quad (9)$$

$0 = w_{(1)} \leq w_{(2)} \leq \dots \leq w_{(n)} = 1$. El MNB consiste en encontrar (numéricamente) el valor de $\hat{\sigma}_N^2$ que maximiza la verosimilitud definida en (9), y considerando a éste como el estimador de σ^2 se calculan los estimadores restantes como se indica a continuación.

En primer lugar se hace una estimación inicial de los parámetros a , μ considerando la propuesta no paramétrica de Cooke (1979),

$$\hat{a}_{init} = 2x_{(1)} - \sum_{i=1}^n \left[\left(1 - \frac{i-1}{n}\right)^n - \left(1 - \frac{i}{n}\right)^n \right] x_{(i)}, \hat{\mu}_{init} = \frac{1}{n} \sum_{i=1}^n \ln(x_i - \hat{a}_{init}) \quad (10)$$

y a continuación se calculan los estimadores finales de acuerdo con

$$\hat{a}_N = x_{(1)} - \exp(\hat{\mu}_{init} + \hat{\sigma}_N E[Z_{(1)}]), \hat{\mu}_N = \frac{1}{n} \sum_{i=1}^n \ln(x_i - \hat{a}_N), \quad (11)$$

donde $\hat{\mu}_{init}$ está definido en (10), $\hat{\sigma}_N$ maximiza (9) y $Z_{(1)}$ denota al valor más pequeño de una muestra aleatoria de tamaño n de una distribución normal estándar.

Método de Nagatsuka y Balakrishnan Modificado

Como se menciona en Nagatsuka y Balakrishnan (2012), la idea de proponer el estimador \hat{a}_N definido en (11) es la de reducir el sesgo imponiendo $x_{(1)} - \hat{a}_N = E[X_{(1)} - a]$, y aproximando $E[X_{(1)} - a]$ por $\exp(\hat{\mu}_{init} + \hat{\sigma}_N E[Z_{(1)}])$, pero $E[X_{(1)} - a]$ no es exactamente igual a $\exp(\mu + \sigma E[Z_{(1)}])$, sino más bien

$$g(\mu, \sigma) \stackrel{\text{def}}{=} E[X_{(1)} - a] = e^\mu \int_0^\infty (1 - \Phi(\ln(t); \sigma^2))^n dt, \quad (12)$$

donde $\Phi(x; \sigma^2)$ denota a la función de distribución acumulada de una distribución normal con esperanza 0 y varianza σ^2 , por lo que la modificación que proponemos consiste en considerar $\hat{a}_{N'} = x_{(1)} - g(\mu_{init}, \hat{\sigma}_N)$ como estimador de a , y calcular los otros estimadores como en el MNB con $\hat{a}_{N'}$ en lugar de \hat{a}_N .

Nuevo Método Propuesto

Este método consiste en resolver la ecuación

$$x_{(1)} - \hat{a}_N = g(\hat{\mu}_N, \hat{\sigma}_N), \quad (13)$$

donde $g(\mu, \sigma)$ está definida en (12), y

$$\hat{\mu}_N = \frac{\sum_{i=1}^n \ln(x_i - \hat{a}_N)}{n}, \hat{\sigma}_N^2 = \frac{\sum_{i=1}^n [\ln(x_i - \hat{a}_N) - \hat{\mu}_N]^2}{n}. \quad (14)$$

Nótese que al reemplazar las ecuaciones (14) en (13), esta última se convierte en una ecuación en una sola variable (\hat{a}_N). Un valor inicial apropiado para resolver numéricamente (13) puede ser la estimación no paramétrica de Cooke definida en (10).

Para el lector interesado en experimentar con los métodos descritos en este artículo, mencionamos que desde la página <http://ciep.itam.mx/~davidm/sofdop.htm> se puede descargar la interfaz *Simple Analyzer* que permite probar los procedimientos ML, MMNB y NM desde una hoja de Excel. Nótese que para correr los procedimientos de esta interfaz debe haberse instalado primero el software *Random Number Generators* que se descarga desde la misma página mencionada anteriormente.

RESULTADOS EXPERIMENTALES

En esta sección discutimos los resultados de nuestros experimentos con los cinco métodos considerados para la estimación de parámetros de una $\log N(a, \mu, \sigma^2)$, mencionando primero algunos detalles sobre el diseño de los experimentos y los algoritmos utilizados para la implementación de los métodos de estimación.

Implementación y Diseño de los Experimentos

Todos los algoritmos utilizados en nuestros experimentos fueron codificados en C++ y exportados a una librería dll, para ser luego llamados desde Excel utilizando Visual Basic for Applications. Para la solución de ecuaciones o maximización de funciones se utilizó la librería Numerical Recipes Software (NRS) que se distribuye con el libro de Press et al. (2007).

El algoritmo para el MMO se codificó en C++ a partir de las ecuaciones (3), reportando que el método falló cuando se obtuvo $\hat{a}_M > x_{(1)}$, mientras que para el ML se solucionó el sistema de ecuaciones (6) utilizando el algoritmo de Brent de la librería NRS, reportando que el método falló si se excedió un máximo número de 100 iteraciones. Para la implementación del MNB se utilizó el procedimiento de integración de Takahashi y Mori (1974), como se sugiere en Nagatsuka y Balakrishnan (2012), y para la maximización de (9) se utilizó el algoritmo de Brent de la librería NRS, codificando directamente en C++ las ecuaciones (10) y (11). Para implementar el algoritmo para MNBM sólo se necesitó adicionar la integración definida en la ecuación (12), para la que se utilizó el algoritmo de Romberg de la librería NRS. Por último, para el MRG se resolvió la ecuación (13) utilizando el algoritmo de Brent de la librería NRS y el algoritmo de Romberg para calcular la integral definida en (12).

En todos los experimentos se fijaron arbitrariamente $\mu = 2$ y $a = 20$, y se consideraron cuatro valores diferentes para el coeficiente de asimetría $G_x = 0.3, 1, 1.5, 2$, y ocho valores diferentes para el tamaño de la muestra, $n = 10, 20, 30, 50, 75, 100, 200, 400$. Para cada una de las 32 combinaciones se calculó σ^2 a partir de G_x y se generaron $M = 2,000$ muestras del tamaño correspondiente. A cada muestra se aplicó cada uno de los cinco métodos de estimación y se calcularon las medidas de desempeño que se detallan a continuación.

Considerando las $M = 2,000$ repeticiones de cada experimento de estimación de los parámetros a, μ, σ^2 , se calcularon el sesgo empírico (S), la raíz del error cuadrático promedio ($RECP$) para cada parámetro, y el promedio de la probabilidad de un valor mayor para el estadístico de Kolmogorov-Smirnov (PKS) definidos por:

$$S = \frac{1}{M} \sum_{i=1}^M (\hat{\theta}_i - \theta), RECP = \sqrt{\frac{1}{M} \sum_{i=1}^M (\hat{\theta}_i - \theta)^2}, PKS = \frac{1}{M} \sum_{i=1}^M P[KS > KSO]_i, \quad (15)$$

donde θ es el verdadero valor del parámetro (a, μ ó σ^2), y $\hat{\theta}_i$ es el valor estimado en la repetición i . Por otro lado, $P[KS > KSO]$ es la probabilidad de obtener un valor del estadístico de Kolmogorov-Smirnov más alto que el observado en la repetición i , bajo la hipótesis nula de que los datos siguen una distribución lognormal con parámetros iguales a los correspondientes estimadores calculados. Es conveniente remarcar que el estadístico de Kolmogorov-Smirnov se sugiere en los textos de simulación (e.g., Kelton et al., 2013) como un buen indicador de la bondad de ajuste de una distribución a un conjunto de datos observados.

RESULTADOS Y DISCUSIÓN

Nuestro primer conjunto de experimentos consistió en comparar el desempeño del MNB versus el MMNB. Estos experimentos mostraron que MMNB mejoró MNB en casi todas las 32 combinaciones de coeficiente de

asimetría y tamaño de muestra. En general, los porcentajes de mejora fueron positivos y menores del 10%, excepto para el caso de la *RECP* con $n = 200$, 400 y $G = 2$, como mostramos en la Tabla 1 por claridad (hemos omitido las *RECP* para σ porque son las mismas que para MNB y MMNB).

En nuestro segundo conjunto de experimentos comparamos el desempeño de MMO, ML, MMNB y NM. Como podemos apreciar de la Tabla 2, MMO y ML fallaron en algunas de las 2000 repeticiones de cada experimento de estimación, aunque los porcentajes de falla tienden a decrecer a medida que crece el tamaño de muestra, por supuesto que MMNB y NM nunca fallaron (y los resultados se omiten en la Tabla 2 por ser 0), lo que muestra la conveniencia de utilizar MMNB o MN cuando los otros métodos fallan.

Tabla 1. *RECP* para a y \square exhibidos por MNB y MMNB.

n	$G = 0.3$				$G = 2$			
	a MNB	a MMNB	μ MNB	μ MMNB	a MNB	a MMNB	μ MNB	μ MMNB
10	0.2817	0.2813	0.8327	0.8270	0.1184	0.1157	0.2663	0.2550
20	0.2632	0.2628	0.6927	0.6896	0.0863	0.0841	0.1798	0.1725
30	0.2536	0.2532	0.6324	0.6301	0.0717	0.0698	0.1460	0.1403
50	0.2432	0.2429	0.5752	0.5737	0.0572	0.0555	0.1127	0.1086
75	0.2360	0.2358	0.5413	0.5401	0.0495	0.0481	0.0949	0.0916
100	0.2303	0.2301	0.5165	0.5157	0.0423	0.0413	0.0801	0.0777
200	0.2191	0.2189	0.4715	0.4708	0.0435	0.0444	0.0641	0.0651
400	0.2083	0.2080	0.4334	0.4323	0.0750	0.0766	0.1013	0.1032

Tabla 2. Número de fallas en 2000 repeticiones de un experimento de estimación.

n	$G = 0.3$		$G = 0.5$		$G = 1.5$		$G = 2$	
	MMO	ML	MMO	ML	MMO	ML	MMO	ML
10	764	55	420	147	294	233	211	348
20	636	643	171	66	85	10	39	15
30	540	560	88	96	26	25	17	15
50	392	418	40	23	23	2	15	0
75	282	312	31	3	50	0	22	0
100	237	273	47	0	71	0	63	0
200	61	86	48	0	87	0	146	0
400	13	23	36	0	129	0	224	0

Tabla 3. RECP para la estimación de a bajo cuatro diferentes métodos.

n	$G = 0.3$				$G = 2$			
	MMO	ML	NM	MMNB	MMO	ML	NM	MMNB
10	3.94	0.32	1.39	0.28	10.54	0.16	0.77	0.12
20	29.97	0.62	0.69	0.26	2.38	0.14	0.49	0.08
30	31.51	0.65	0.84	0.25	8.96	0.21	0.29	0.07
50	68.84	0.75	0.75	0.24	0.26	0.10	0.13	0.06
75	14.69	0.71	0.60	0.24	0.18	0.06	0.08	0.05
100	11.01	0.64	0.58	0.23	0.16	0.05	0.07	0.04
200	4.26	0.55	0.63	0.22	0.11	0.03	0.05	0.04
400	2.40	0.42	0.74	0.21	0.09	0.02	0.04	0.08

Para facilitar la presentación, en las tablas 3, 4, 5 y 6 reportamos los resultados sólo para los casos $G = 0.3$ y $G = 2$; nuestras conclusiones no cambian sustancialmente al excluir los otros valores. Como se observa de

la Tabla 3 y la Fig. 1, el MMNB muestra la RECP más pequeña para la estimación de a , excepto cuando $G = 2$ y $n \geq 200$, donde ML se desempeñó mejor. En general, NM y ML tuvieron un desempeño similar y el MMO tuvo un mal desempeño (los correspondientes RECP se omiten en la Fig. 1 por ser demasiado grandes). Los resultados para los sesgos de las estimaciones de μ son similares y se han omitido.

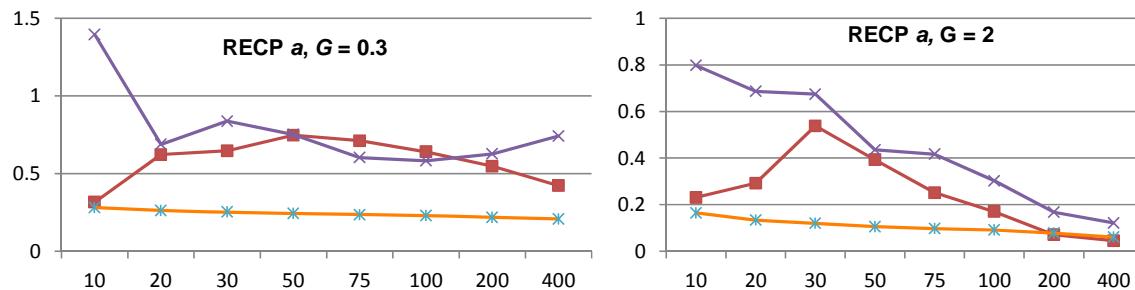


Fig. 1. Desempeño de tres métodos de estimación para a . —ML —NM —MMNB

Tabla 4. RECP para la estimación de μ bajo cuatro diferentes métodos.

n	G = 0.3				G = 2			
	MMO	ML	NM	MMNB	MMO	ML	NM	MMNB
10	0.58	1.50	1.37	0.83	0.55	0.74	0.64	0.25
20	0.59	0.65	0.95	0.69	0.41	0.74	0.31	0.17
30	0.53	0.49	0.80	0.63	0.34	0.19	0.22	0.14
50	0.52	0.43	0.63	0.57	0.24	0.13	0.15	0.11
75	0.48	0.39	0.54	0.54	0.20	0.10	0.12	0.09
100	0.47	0.36	0.50	0.52	0.18	0.08	0.10	0.08
200	0.39	0.31	0.44	0.47	0.14	0.05	0.08	0.07
400	0.29	0.25	0.43	0.43	0.11	0.04	0.06	0.10

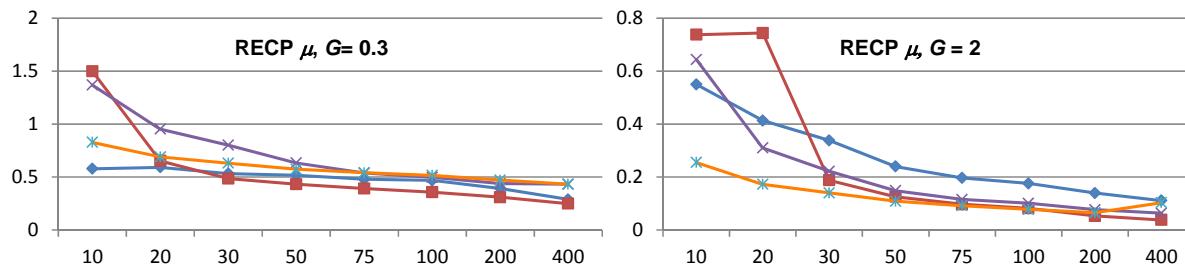


Fig. 2. Desempeño de cuatro métodos de estimación para μ . —MMO —ML —NM —MMNB

Como puede apreciarse de la Tabla 4 y la Fig. 2, el ML (cuando no falla) muestra la RECP más pequeña para la estimación de μ cuando $G = 0.3$, $n \geq 20$, y $G = 2$, $n \geq 100$, y el MMNB se desempeñó mejor en los otros casos. Los resultados para los sesgos de las estimaciones de μ son similares y se omiten.

Como puede apreciarse de la Tabla 5 y de la Fig. 3, el MMNB mostró la RECP más pequeña para la estimación de σ en todos los casos, excepto para $G = 2$, $n \geq 200$ donde el ML se desempeñó mejor (cuando no falló). Estos resultados confirman la intuición de que el MMNB proporciona buenos estimadores para σ , ya que maximiza una verosimilitud que depende sólo de σ . Sin embargo, como puede apreciarse de la Tabla 6 y la Fig. 4, el MMNB no se desempeñó bien desde el punto de vista del PKS, ya que proporciona probabilidades relativamente pequeñas, especialmente cuando el tamaño de muestra es grande; en este caso el ML (cuando no falla) se desempeña como el mejor, excepto para los casos $G = 0.3$, $n = 10$, y $G = 2$, $n \leq 20$, donde el NM se desempeñó como el mejor. Los resultados para los sesgos de las estimaciones de μ fueron similares que para σ y fueron omitidos.

Tabla 5. RECP para la estimación de σ bajo cuatro diferentes métodos.

n	$G = 0.3$				$G = 2$			
	MMO	ML	NM	MMNB	MMO	ML	NM	MMNB
10	1.734	29.652	39.869	2.604	0.506	4.452	4.497	0.516
20	1.409	8.582	26.747	1.663	0.438	6.839	1.172	0.380
30	1.118	1.643	20.409	1.248	0.381	0.368	0.437	0.303
50	0.939	1.120	13.601	0.922	0.325	0.259	0.271	0.222
75	0.842	0.893	9.869	0.763	0.277	0.205	0.228	0.177
100	0.702	0.709	8.156	0.607	0.258	0.175	0.203	0.167
200	0.581	0.520	5.214	0.444	0.199	0.116	0.161	0.384
400	0.445	0.403	3.572	0.347	0.162	0.082	0.131	0.555

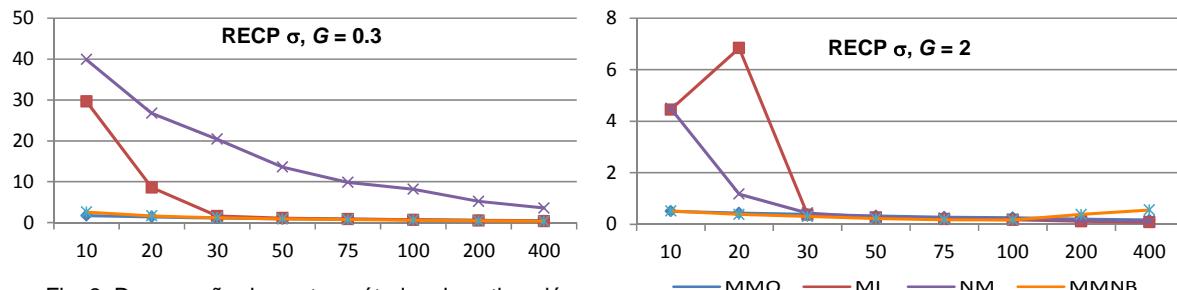
Fig. 3. Desempeño de cuatro métodos de estimación para σ .

Tabla 6 PKS promedio para los cuatro diferentes métodos.

n	$G = 0.3$				$G = 2$			
	MMO	ML	NM	MMNB	MMO	ML	NM	MMNB
10	0.873	0.105	0.462	0.310	0.824	0.148	0.691	0.606
20	0.853	0.827	0.482	0.196	0.777	0.009	0.788	0.644
30	0.844	0.856	0.513	0.134	0.753	0.845	0.795	0.666
50	0.839	0.849	0.574	0.078	0.717	0.833	0.774	0.679
75	0.824	0.834	0.587	0.044	0.695	0.821	0.756	0.657
100	0.820	0.829	0.587	0.020	0.682	0.818	0.750	0.582
200	0.796	0.808	0.571	0.002	0.626	0.796	0.698	0.071
400	0.755	0.766	0.499	< 1E-03	0.543	0.756	0.630	1E-04

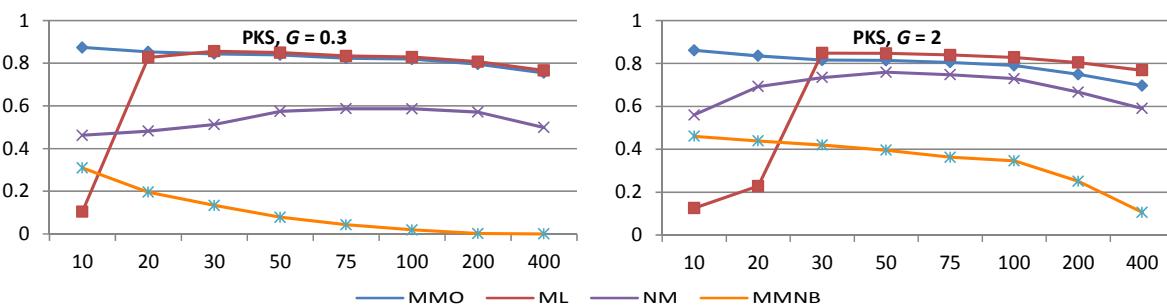


Fig. 4. Desempeño de cuatro métodos para la probabilidad de obtener un valor más grande que el estadístico de Kolmogorov-Smirnov.

Para facilitar la presentación de los resultados, no se han reportado los anchos medios de las estimaciones definidas en (15). Sin embargo las $M = 2000$ repeticiones utilizadas para cada experimento de estimación han permitido tener errores de estimación relativamente pequeños (menor al 10% del valor de la estimación en cada caso).

CONCLUSIONES

Como se reporta en la bibliografía, los resultados de nuestros experimentos confirman que el método de máxima verosimilitud local proporciona estimadores con buenas propiedades desde los puntos de vista del sesgo y del error cuadrático promedio cuando el tamaño de muestra no es muy pequeño y el método no falla. Más aún, nuestros resultados también muestran que una conclusión similar es cierta desde el punto de vista de la probabilidad de obtener un valor más grande que el estadístico de Kolmogorov-Smirnov.

Cuando el método de máxima verosimilitud local falla o el tamaño de muestra es pequeño, el método de Nagatsuka y Balakrishnan tiene el mejor desempeño desde el punto de vista del sesgo y del error cuadrático promedio, y el método propuesto en este artículo tiene el mejor desempeño desde el punto de vista de la probabilidad de obtener un valor más grande que el estadístico de Kolmogorov-Smirnov.

De acuerdo con nuestros resultados experimentales, para ajustar un conjunto de datos a una distribución lognormal con traslado, se sigue recomendando el uso del método de máxima verosimilitud local, si éste no falla; de otra forma, se recomienda utilizar el método de Nagatsuka y Balakrishnan (con la modificación propuesta en este artículo) si se ha decidido que la distribución a utilizar para ajustar los datos es la lognormal con traslado; en cambio, si se desea seleccionar la distribución que se ajusta mejor al conjunto de datos, utilizando el criterio del estadístico de Kolomogorov-Smirnov, se recomienda el nuevo método (NM) propuesto en este artículo.

Debido a que el método NM requiere (en cada iteración) de la evaluación de una integral simple, a diferencia del método de Nagatsuka y Balakrishnan (que requiere de la evaluación de una integral doble), el NM es más fácil de implementar y se ejecuta en tiempos de corrida significativamente más pequeños.

Por último, teniendo en cuenta que el ECP teórico de un estimador es igual al cuadrado de su sesgo más su varianza, aunque no hemos demostrado matemáticamente que el NM proporciona estimadores consistentes, los resultados experimentales indican que la RECP tiende a disminuir, a medida que el tamaño de muestra se incrementa, evidenciando que tanto el sesgo como la varianza de los estimadores tienden a hacerse pequeños (posiblemente cero).

AGRADECIMIENTOS

Este trabajo se ha desarrollado con el apoyo de la Asociación Mexicana de Cultura A.C. y de la Escuela Superior de Economía y Negocios de El Salvador.

REFERENCIAS

- Cohen, A. C. y B. J. Whitten, “*Estimation in the three-parameter lognormal distribution*”, Journal of the American Statistical Association 75, 399-404 (1980)
- Cooke, P., “*Statistical inference for bounds of random variables*”, Biometrika 66, 367-374 (1979)
- Hill, B.M., “*Three-parameter lognormal distribution and Bayesian analysis of a point-source epidemic*”, Journal of the American Statistical Association 58, 72–84 (1963)
- Kelton, W. D., J. S. Smith y D. T. Sturrock, “*Simio and Simulation: Modeling. Analysis, Applications*” (3^a ed.), Simio LLC, Sewickley (2013)
- Muñoz, D. F. y D. Villafuerte, “*Análisis de la entrada en simulación estocástica*”, Información Tecnológica 26, 13-22 (2015)
- Nagatsuka, H. y N. Balakrishnan, “*A consistent parameter estimation in the three-parameter lognormal distribution*”, Journal of Statistical Planning and Inference 142, 2071-2086 (2012)
- Nagatsuka, H., T. Kamakura y N. Balakrishnan, “*A consistent method of estimation for the three-parameter Weibull distribution*”, Computational Statistics & Data Analysis 58, 210-226 (2013)

Nagatsuka, H. y N. Balakrishnan, "A consistent method of estimation for the parameters of the three-parameter inverse Gaussian distribution". Journal of Statistical Computation and Simulation 83, 1915-1931 (2013)

Press, W. H., B. P., Teukolsky, S. A., Vetterling y B. P. Flannery, "Numerical Recipes The Art of Scientific Computing" (3^a ed.), Cambridge University Press, New York (2007)

Raynal, J. A. y L. G. García, "Análisis de caudales máximos anuales usando la distribución GVE para tres poblaciones", Información Tecnológica 16, 91-99 (2005)

Singh, V. P., J. F. Cruise y Ming Ma, "A comparative evaluation of the estimators of the three-parameter lognormal distribution by Monte Carlo simulation", Computational Statistics & Data Analysis 10, 71-85 (1990)

Takahasi, H. y M. Mori, "Double exponential formulas for numerical integration", Publications of the Research Institute for Mathematical Sciences Kyoto University 9, 721–741 (1974)

Vallejos, O. S. y D. M. Aedo, "Weibull_fit.xls: programa para el ajuste óptimo de la función de densidad de probabilidad Weibull de tres parámetros", Información Tecnológica 21, 91-99 (2010)