
Estimación de los parámetros de la distribución Lambda generalizada a través del método de los momentos y el programa MatLab¹

Using the Method of Moments for estimating parameters of Generalized Lambda Distribution and its implementation in Matlab

Camilo Andre Castillo^a
cacastillot@correo.udistrital.edu.co

Luis Alejandro Másmela^b
lasmela@udistrital.edu.co

Luis Fernando Villarraga^c
lfvillarraga@udistrital.edu.co

Resumen

Muchos de los problemas en diferentes áreas del conocimiento tienen solución por medio de la construcción de modelos estadísticos. Dichos modelos se soportan comúnmente en distribuciones de probabilidad y más exactamente en cómo se distribuyen los datos que se toman como base para su construcción. El interés de ajustar distribuciones a conjuntos de datos pretende describir el comportamiento de ellos mediante la distribución encontrada. El artículo presenta una familia de estas distribuciones, la *distribución Lambda generalizada* (DLG) y un programa en MatLab implementando el método de momentos para estimar los parámetros de la distribución que se ajuste a conjuntos de datos o aproxime algunas distribuciones conocidas.

Palabras clave: distribución Lambda Generalizada, Matlab, método de momentos.

¹Castillo, C.A., Másmela, L.A., Villarraga, L.F. (2015). Estimación de los parámetros de la distribución Lambda generalizada a través del método de los momentos y el programa MatLab *Comunicaciones en Estadística*, **8**(2), 193-209.

^aMatemático. Facultad de Ciencias y Educación. Universidad Distrital Francisco José de Caldas. Colombia.

^bProfesor asistente. Facultad de Ciencias y Educación. Universidad Distrital Francisco José de Caldas. Colombia.

^cProfesor asistente. Facultad de Ciencias y Educación. Universidad Distrital Francisco José de Caldas. Colombia.

Abstract

Many of the problems in different areas of knowledge can be solved through the construction of statistical models. Such models are commonly supported on probability distributions and exactly how the data are taken as the basis for its construction are distributed. Interest distributions adjusting data sets intended to describe their behavior by distributing found. The paper presents a family of these distributions, the *Lambda* widespread distribution (DLG) and a MatLab program implementing the method of moments to estimate the distribution parameters to fit data sets or approaching some popular distributions.

Keywords: Lambda Generalized Distribution, Matlab, Moments Method.

1. Introducción

La distribución Lambda generalizada tiene sus orígenes en la distribución lambda de Tukey¹, la cual está definida por la función cuantil (Ramberg & Schmeiser (1974)) generalizaron a una familia de distribuciones de cuatro parámetros para generar variables aleatorias utilizando el método de Montecarlo. Algunas aplicaciones de la DLG se encuentra en el modelamiento de fenómenos biológicos y físicos, Silver (1977), pruebas de bondad de ajuste en regresiones lógicas, Pregibon (1980), valoración de opciones, Corrado (2001).

Para estimar los parámetros de la DLG se acude a métodos clásicos de la teoría, el método de máxima verosimilitud (MV), percentiles (MP), método de mínimos cuadrados (MMC) y el método de los momentos (MM), entre otros. Ejemplos de cómo proceder pueden consultarse en textos como los de Wackerly & Scheaffer (2010), Casella & Berger (2002), Mood (1950).

En el momento de utilizar algún método de los ya nombrados para la estimación de parámetros en la DLG, se puede encontrar en la literatura formas de proceder. El método de L-momentos (Karvanen & Nuutinen 2008), método de percentiles (Másmela & Rodríguez 2012), el método de máxima verosimilitud (Su 2007), entre otros. Obtener por métodos analíticos un estimador conlleva una labor compleja. Por lo tanto, los métodos numéricos y los paquetes informáticos son herramientas útiles para aproximarse a ellos. Un *software* versátil para la implementación de métodos numéricos es MatLab, en particular el GUIDE permite construir programas que implementan estos métodos, y por esta razón se utilizará para programar el método de momentos para la estimación de parámetros de la DLG.

¹Nombrada así en honor a Jhon Tukey en 1947.

2. Distribución Lambda generalizada y restricciones para su espacio de parámetros

La distribución Lambda de Tukey depende de un solo parámetro y su forma funcional es esta

$$Q(p) = \begin{cases} \frac{p^\lambda - (1-p)^\lambda}{\lambda} & \text{si } \lambda \neq 0 \\ \frac{\log(p)}{1-p} & \text{si } \lambda = 0 \end{cases} \quad y \quad p \neq 1 \tag{1}$$

Donde $0 \leq p \leq 1$.

Al ser generalizada por (Ramberg & Schmeiser 1974) se conoce como familia de distribución lambda generalizada con parámetros $\lambda_1, \lambda_2, \lambda_3, \lambda_4$, o $DLG(\lambda_1, \lambda_2, \lambda_3, \lambda_4)$ y se suele especificar en términos de la función percentil:

$$Q(y) = Q(y; \lambda_1, \lambda_2, \lambda_3, \lambda_4) = \lambda_1 + \frac{y^{\lambda_3} - (1-y)^{\lambda_4}}{\lambda_2}, \tag{2}$$

Donde $0 \leq y \leq 1$.

Los parámetros λ_1 y λ_2 son, respectivamente, parámetros de localización y escala, λ_3 y λ_4 determinan sesgo y curtosis. A partir de (2) se encuentra una expresión para la función de densidad de probabilidad (f.d.p.) de la $DLG(\lambda_1, \lambda_2, \lambda_3, \lambda_4)$ presentada a continuación:

$$f(x) = \frac{\lambda_2}{\lambda_3 y^{\lambda_3-1} + \lambda_4 (1-y)^{\lambda_4-1}}, \quad y = Q^{-1}(x). \tag{3}$$

La figura 1 presenta algunas gráficas de f.d.p. y la figura 2 ejemplifica el hecho de ser λ_1 un parámetro de localización.

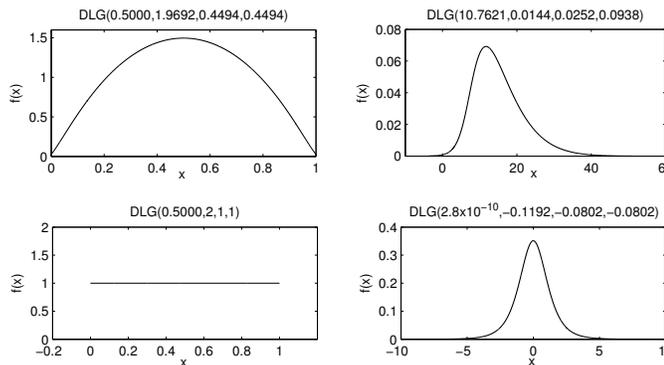


Figura 1: Distribuciones $Beta(1,1)$, $Gamma(5,3)$, $uniforme(0,1)$ y $normal\ estandar$ aproximadas a través de la DLG. Fuente: Elaboración propia.

La expresión (3) no siempre especifica una f.d.p. válida. Para que una función $f(x)$ especifique una f.d.p. debe cumplir:

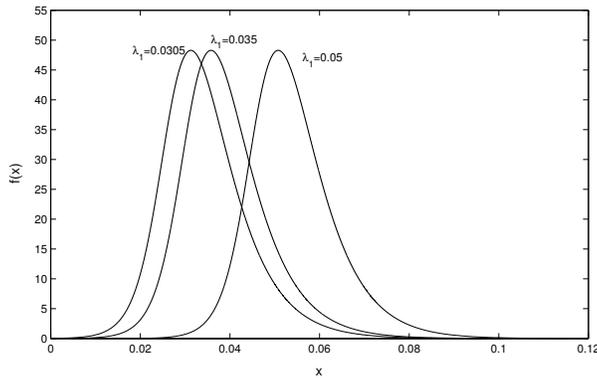


Figura 2: DLG con parámetros $\lambda_2 = 1.3673$, $\lambda_3 = 0.0045$, $\lambda_4 = 0.0102$ y diferentes λ_1 . Fuente: Elaboración propia.

$$f(x) \geq 0 \quad , \quad \int_{-\infty}^{\infty} f(x)dx = 1. \tag{4}$$

La función presentada en (3) junto a las condiciones en (4) implican una DLG válida siempre que:

$$\frac{\lambda_2}{\lambda_3 y^{\lambda_3-1} + \lambda_4 (1-y)^{\lambda_4-1}} \geq 0, \quad \int_{-\infty}^{\infty} f(Q(y))dQ(y) = 1 \tag{5}$$

La integral en (5) converge a 1 para cualquier valor de $\lambda_1, \lambda_2, \lambda_3, \lambda_4$. Ahora la primera condición presenta restricciones sobre el espacio de parámetros $\lambda_2, \lambda_3, \lambda_4$, la fracción en (5) es no negativa cuando el numerador y denominador tienen el mismo signo, para $0 \leq y \leq 1$. El signo del denominador depende solamente de los parámetros λ_3 y λ_4 y es determinado al dividir el espacio (λ_3, λ_4) en varias regiones. Las siguientes regiones hacen de la DLG una función de densidad:

$$\text{Región 1} = \{(\lambda_3, \lambda_4) \mid \lambda_3 \leq -1, \lambda_4 \geq 1\} \tag{6}$$

$$\text{Región 2} = \{(\lambda_3, \lambda_4) \mid \lambda_3 \geq 1, \lambda_4 \leq -1\} \tag{7}$$

$$\text{Región 3} = \{(\lambda_3, \lambda_4) \mid \lambda_3 \geq 0, \lambda_4 \geq 0\} \tag{8}$$

$$\text{Región 4} = \{(\lambda_3, \lambda_4) \mid \lambda_3 \leq 0, \lambda_4 \leq 0\} \tag{9}$$

Las curvas que limitan las regiones 5 y 6 válidas y no válidas están dadas respectivamente por:

$$\frac{(1-\lambda_3)^{1-\lambda_3}}{(\lambda_4-\lambda_3)^{\lambda_4-\lambda_3}}(\lambda_4-1)^{\lambda_4-1} = \frac{-\lambda_3}{\lambda_4} \tag{10}$$

$$\frac{(1-\lambda_3)^{1-\lambda_3}}{(\lambda_4-\lambda_3)^{\lambda_4-\lambda_3}}(\lambda_4-1)^{\lambda_4-1} = \frac{-\lambda_4}{\lambda_3} \tag{11}$$

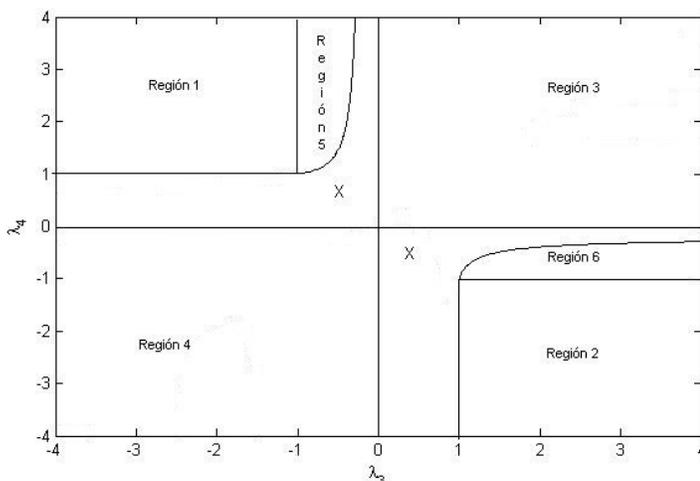


Figura 3: *Espacio (λ_3, λ_4) válido.* Fuente: *Elaboración propia.*

Análisis adicionales que pueden ser consultados en (Karian & Dudewicz 2010) detallan la construcción del espacio de parámetros para la DLG. La Figura 3 presenta el espacio (λ_3, λ_4) donde la DLG es una función de densidad válida a partir de las expresiones (6) a (11). La región designada por X no cumple (5) para todo $y \in [0, 1]$.

3. Momentos de la DLG

Karian & Dudewicz (2010) proponen especificar los cuatro primeros momentos teóricos centrales de la $DLG(\lambda_1, \lambda_2, \lambda_3, \lambda_4)$ a partir de los momentos no centrados de la $DLG(\lambda_1, \lambda_2, \lambda_3, \lambda_4)$. Por consecuencia directa de (3), si X es una variable aleatoria $DLG(\lambda_1, \lambda_2, \lambda_3, \lambda_4)$, entonces $Z = X - \lambda_1$ es $DLG(0, \lambda_2, \lambda_3, \lambda_4)$, mostrando a λ_1 como parámetro de localización, por tanto, se asume $\lambda_1 = 0$ para simplificar los análisis siguientes.

Para determinar los momentos no centrados de la $DLG(\lambda_1, \lambda_2, \lambda_3, \lambda_4)$ se toma Z una $DLG(0, \lambda_2, \lambda_3, \lambda_4)$, entonces $E(Z^k)$, está dado por:

$$E(Z^k) = \frac{1}{\lambda_2^k} \sum_{i=0}^k \left[\binom{k}{i} (-1)^i \beta(\lambda_3(k-i) + 1, \lambda_4 i + 1) \right] \tag{12}$$

Donde $\beta(a, b) = \int_0^1 x^{a-1} (1-x)^{b-1} dx$. La expresión (12) se conoce como función generadora de momentos de la DLG y presenta restricciones para su convergencia, pues interviene la función $\beta(a, b)$ y esta converge bajo ciertas restricciones en sus argumentos. Los k -ésimos momentos de la $DLG(\lambda_1, \lambda_2, \lambda_3, \lambda_4)$ existen, si y

solamente si, $\lambda_3 > -1/k$ y $\lambda_4 > -1/k$, condiciones que dependen de la cantidad de momentos, y como el interés es la estimación de los cuatro parámetros de la DLG, la cantidad mínima de ecuaciones que se requiere para utilizar el método de momentos son cuatro. Por tanto, se impone la condición $\lambda_3 > -1/4$ y $\lambda_4 > -1/4$. Si X es DLG($\lambda_1, \lambda_2, \lambda_3, \lambda_4$) con $\lambda_3 > -1/4$ y $\lambda_4 > -1/4$. Entonces los primeros cuatro momentos centrados, $\alpha_1, \alpha_2, \alpha_3, \alpha_4$ (media, varianza, sesgo, y curtosis, respectivamente) están dados por:

$$\alpha_1 = \mu = E(X) = \lambda_1 + \frac{A}{\lambda_2}, \quad (13)$$

$$\alpha_2 = \sigma^2 = E[(X - \mu)^2] = \frac{B - A^2}{\lambda_2^2}, \quad (14)$$

$$\alpha_3 = E(X - E(X))^3 / \sigma^3 = \frac{C - 3AB + 2A^3}{\lambda_2^3 \sigma^3}, \quad (15)$$

$$\alpha_4 = E(X - E(X))^4 / \sigma^4 = \frac{D - 4AC + 6A^2B - 3A^4}{\lambda_2^4 \sigma^4} \quad (16)$$

Donde

$$A = \frac{1}{1 + \lambda_3} - \frac{1}{1 + \lambda_4}, \quad (17)$$

$$B = \frac{1}{1 + 2\lambda_3} + \frac{1}{1 + 2\lambda_4} - 2\beta(1 + \lambda_3, 1 + 2\lambda_4), \quad (18)$$

$$C = \frac{1}{1 + 3\lambda_3} - \frac{1}{1 + 3\lambda_4} - 3\beta(1 + 2\lambda_3, 1 + \lambda_4) + 3\beta(1 + \lambda_3, 1 + 2\lambda_4), \quad (19)$$

$$D = \frac{1}{1 + 4\lambda_3} + \frac{1}{1 + 4\lambda_4} - 4\beta(1 + 3\lambda_3, 1 + \lambda_4) + 6\beta(1 + 2\lambda_3, 1 + 2\lambda_4) - 4\beta(1 + \lambda_3, 1 + 3\lambda_4). \quad (20)$$

4. Estimación de parámetros de la DLG vía método de momentos

El método de momentos diseñado por Karl Pearson (1857-1936), y presentado en (Pearson 1894), consiste en igualar los primeros k momentos muestrales $\hat{\alpha}_k$ con los correspondientes k momentos poblacionales $\mu'_k = E(X^k)$. Observe que $E(X^k)$ depende de $(\theta_1, \dots, \theta_k)$:

$$\begin{aligned}
\hat{\alpha}_1 &= \mu'_1(\theta_1, \dots, \theta_k) \\
\hat{\alpha}_2 &= \mu'_2(\theta_1, \dots, \theta_k) \\
&\vdots \\
\hat{\alpha}_k &= \mu'_k(\theta_1, \dots, \theta_k)
\end{aligned}
\tag{21}$$

Para X_1, X_2, \dots, X_n muestra de una población con f.d.p. $f(x|\theta_1, \dots, \theta_k)$ el estimador $(\hat{\theta}_1, \dots, \hat{\theta}_k)$ vía método de momentos de $(\theta_1, \dots, \theta_k)$ es obtenido por la solución de (21) para $(\theta_1, \dots, \theta_k)$ en términos de $(\hat{\alpha}_1, \dots, \hat{\alpha}_k)$. Los cuatro primeros estadísticos $\hat{\alpha}_k$ obtenidos a partir de la muestra están dados por:

$$\hat{\alpha}_1 = \bar{X} = \sum_{i=1}^n \frac{X_i}{n} \tag{22}$$

$$\hat{\alpha}_2 = \hat{\sigma}^2 = \sum_{i=1}^n \frac{(X_i - \bar{X})^2}{n} \tag{23}$$

$$\hat{\alpha}_3 = \sum_{i=1}^n \frac{(X_i - \bar{X})^3}{n\hat{\sigma}^3} \tag{24}$$

$$\hat{\alpha}_4 = \sum_{i=1}^n \frac{(X_i - \bar{X})^4}{n\hat{\sigma}^4} \tag{25}$$

Para estimar los parámetros de la DLG mediante el método de momentos se soluciona el siguiente sistema de ecuaciones:

$$\alpha_i = \hat{\alpha}_i \quad \text{con} \quad i = 1, 2, 3, 4 \tag{26}$$

para $\lambda_1, \lambda_2, \lambda_3$ y λ_4 .

Las ecuaciones (17) a (20) no dependen de λ_1 y λ_2 . Además α_3 y α_4 tienen en su denominador el término $\lambda_2\sigma$, por (14) se tiene que $\lambda_2^i\sigma^i = (B - A^2)^{i/2}$ para $i = 3$ y 4 , así α_3 y α_4 solo dependen de λ_3 y λ_4 . Por tanto, λ_3 y λ_4 se pueden obtener a partir del subsistema:

$$\alpha_3 = \hat{\alpha}_3 \quad , \quad \alpha_4 = \hat{\alpha}_4. \tag{27}$$

Hallados los valores λ_3 y λ_4 por medio de (27) se procede a utilizar (13) y (14) para encontrar λ_1 y λ_2 , respectivamente. El sistema de ecuaciones (27) es difícil de solucionar por métodos analíticos, por lo que se utilizan métodos numéricos para aproximar su solución. Algoritmos para encontrar una solución numérica a sistemas de ecuaciones son generalmente nombrados de "búsqueda". En este

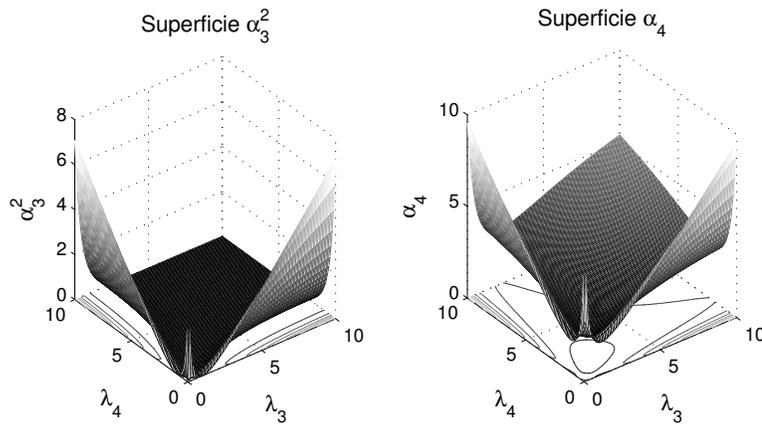


Figura 4: *Superficies. Fuente: Elaboración propia.*

artículo se presenta un método de búsqueda para solucionar el sistema (27) a partir de las curvas de nivel de las superficies de la figura 4.

Si (α_3, α_4) origina una $DLG(\lambda_1, \lambda_2, \lambda_3, \lambda_4)$ válida, entonces $(-\alpha_3, \alpha_4)$ genera una $DLG(\lambda_1, \lambda_2, \lambda_4, \lambda_3)$ válida con primeros cuatro momentos $\alpha_1 - \frac{2A}{\lambda_2}, \alpha_2, -\alpha_3, \alpha_4$, considerando el espacio (α_3^2, α_4) asociado a la $DLG(\lambda_1, \lambda_2, \lambda_3, \lambda_4)$. Para la ecuación $\alpha_3 = \hat{\alpha}_3$ se toma la curva de nivel $\hat{\alpha}_3^2$ de la superficie α_3^2 teniendo la curva de nivel $\alpha_3^2 = \hat{\alpha}_3^2$ en el plano (λ_3, λ_4) asociada a cada $\hat{\alpha}_3$, para $\hat{\alpha}_4 = \alpha_4$ se toma la curva de nivel $\hat{\alpha}_4$ de la superficie α_4 . La figura 5 presenta diferentes curvas de nivel para cada una de las superficies α_3^2 y α_4 .

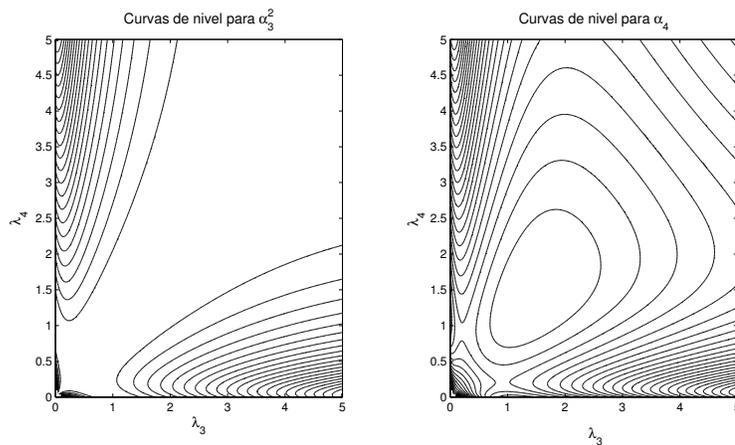


Figura 5: *Curvas de nivel de α_3^2 y α_4 . Fuente: Elaboración propia.*

Las soluciones del sistema resultante están dadas por los puntos de intersección de

las dos curvas de nivel. Se presentan las curvas $\alpha_3 = 0.03$ y $\alpha_4 = 2$, en la figura 6 las curvas tienen varios puntos de intersección, esto es, no solo un par de valores (λ_3, λ_4) resuelven el sistema y, por tanto, diferentes DLG se ajustan a un mismo conjunto de datos.

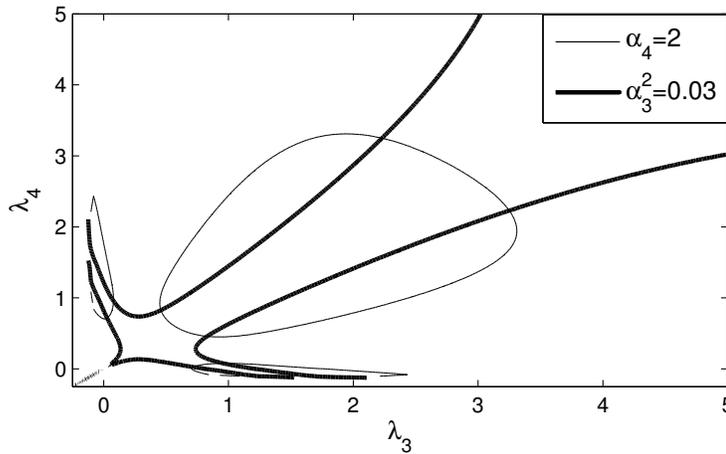


Figura 6: Curvas de nivel $\alpha_3^2 = 0.03$ y $\alpha_4 = 2$. Fuente: Elaboración propia.

5. Programa para estimar parámetros de la DLG

La solución del sistema (27) está dada por el punto (λ_3, λ_4) , que es intersección de las curvas $\alpha_3^2 = \hat{\alpha}_3^2$ y $\alpha_4 = \hat{\alpha}_4$. Una solución numérica para calcular tal punto se obtiene a partir de la diferencia entre $\hat{\alpha}_i$ y α_i , $i = 3, 4$. Cuanto menor sea esta diferencia, la solución numérica se aproximará más a la solución. El algoritmo propuesto para la solución de (27) se describe a continuación.

5.1. Algoritmo de búsqueda

1. Graficar las curvas para $\alpha_3^2 = \hat{\alpha}_3^2$ y $\alpha_4 = \hat{\alpha}_4$.
2. Tomar las coordenadas para un punto $(\lambda_{3_0}, \lambda_{4_0})$ próximo a una intersección de las curvas del paso 1.
3. Calcular $|\hat{\alpha}_i - \alpha_i|$ para $i = 3, 4$ con el paso 2 y nombrar **Error** al mayor de los dos valores.
4. Nombrar $\epsilon = 0.2$ y construir dos vectores L_3 y L_4 con primera entrada $\lambda_{3_0} - \epsilon$ hasta $\lambda_{3_0} + \epsilon$ y $\lambda_{4_0} - \epsilon$ hasta $\lambda_{4_0} + \epsilon$, respectivamente, con saltos de $\gamma = \epsilon/1000$.

5. Evaluar los valores i-ésimos de los vectores del paso 4 en:

$$\alpha_3 = E(X - E(X))^3 / \sigma^3 = \frac{C - 3AB + 2A^3}{\lambda_2^3 \sigma^3}$$

$$\alpha_4 = E(X - E(X))^4 / \sigma^4 = \frac{D - 4AC + 6A^2B - 3A^4}{\lambda_2^4 \sigma^4},$$

comparar $|\hat{\alpha}_3 - \alpha_3|$, $|\hat{\alpha}_4 - \alpha_4|$ y **Error** tomar el mayor de los tres y nombrar **NuevoError**.

6. Llamar $(\lambda_{3_0}, \lambda_{4_0})$ a los k-ésimos valores que calculen el menor de todos los **NuevoError**, $\epsilon = \gamma$ y hacer **Error** igual a **NuevoError**.

7. Iterar tres veces más pasos 2 a 7 con las coordenadas del punto 6.

8. $(\hat{\lambda}_3, \hat{\lambda}_4) = (\lambda_{3_0}, \lambda_{4_0})$.

Al hacer una red en el punto 4 puede que los puntos sobre los que se calculan los errores no aproximen bien la solución y hacer una red fina con el fin de evitar el problema puede resultar en un largo tiempo de computación. Por ejemplo, pasar de una red 10×10 a una 25×25 aumenta el tiempo de computación por más de un factor 6 y pasar a una 250×250 aumenta el factor en un 625. Si el original 10×10 requiere 5 segundos de tiempo de computación, la red 250×250 requerirá 52 minutos de computación.

Un algoritmo más aceptable es buscar un intervalo próximo a la solución y subdividir tanto como se desee. Suponga que una búsqueda para (σ, ρ) conduce a un intervalo $[A_1, A_2] \times [B_1, B_2]$. En lugar de construir una red 10000×10000 , utilizar el algoritmo descrito implica que en la primera iteración se reduce el área de búsqueda en un 0.04% a la cuarta iteración se tendrá un mejor efecto que la red 10000×10000 y el tiempo de computación será menor; por lo tanto se toman un total de cuatro iteraciones en el paso 7.

El punto $(\hat{\lambda}_3, \hat{\lambda}_4)$ es una solución numérica para el sistema $\alpha_3^2 = \hat{\alpha}_3^2$ y $\alpha_4 = \hat{\alpha}_4$, con un *error absoluto* (Mathews & Fink 1999) igual a **Error**. Los puntos finales $\hat{\lambda}_1$, $\hat{\lambda}_2$, $\hat{\lambda}_3$, $\hat{\lambda}_4$ se presentan como solución numérica del sistema (26) y se utilizan para construir la gráfica $(x, f(x))$, donde $f(x)$ es la función de densidad de probabilidad de la DLG $(\hat{\lambda}_1, \hat{\lambda}_2, \hat{\lambda}_3, \hat{\lambda}_4)$.

Como se observa en la figura 6, la solución al sistema presentado en (27) puede tener varias soluciones, se tomará la intersección más próxima al eje λ_4 . Se construye un programa llamado "Selección" utilizando como herramienta el GUIDE de MATLAB en el que se implementó el algoritmo. A continuación se presentan algunas imágenes y comentarios del programa de tal manera que expliquen su funcionamiento.

5.2. Programa

En Matlab se construyó un programa llamado “Seleccion”, que implementa el algoritmo presentado para estimar los parámetros de una DLG que se ajuste a un conjunto de datos, junto con la gráfica de f.d.p. de dicha DLG. La apariencia inicial del programa propuesto se muestra en la siguiente figura.

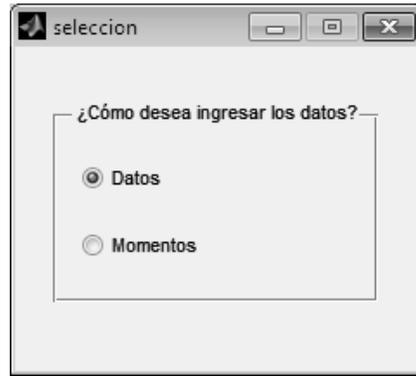


Figura 7: Ventana inicial. Fuente: Matlab.

En esta ventana se elige cómo se ingresan los momentos muestrales $\hat{\alpha}_1$, $\hat{\alpha}_2$, $\hat{\alpha}_3$ y $\hat{\alpha}_4$, si se requiere que el programa los calcule a partir de un conjunto de datos o si el mismo usuario los ingresa. Si la elección es que el programa los calcule, se elige “Datos” y emerge la figura 8.

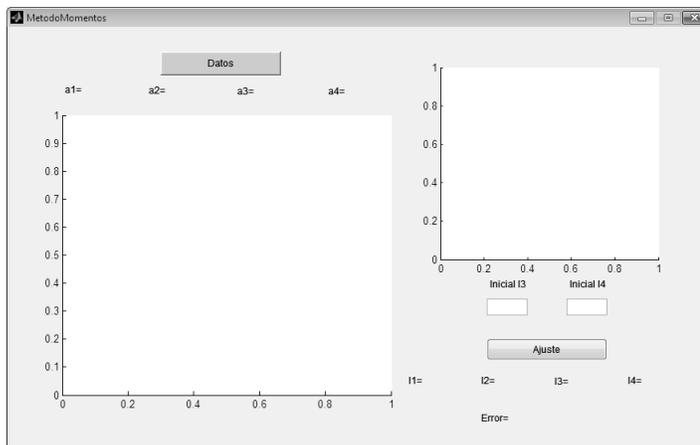


Figura 8: Método Momentos. Fuente: Matlab.

En ella se presiona en el botón “Datos”, se busca el archivo .txt con los datos a describir mediante la DLG y abrir. La figura 9 presenta lo anterior para el archivo datosnormales.txt que contiene un conjunto de 1000 datos con distribución normal

estándar y en forma de columna.

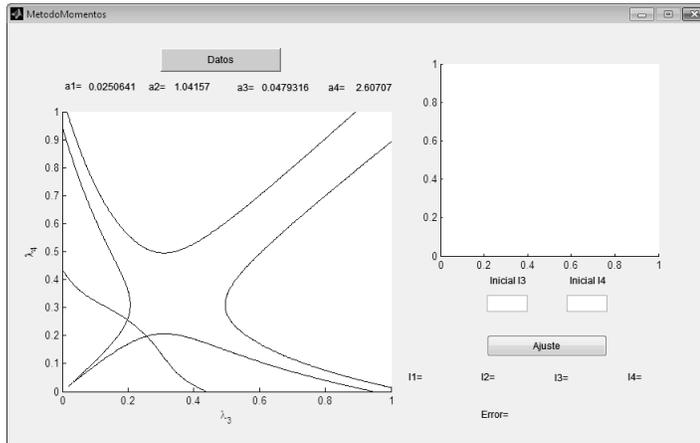


Figura 9: *Curvas $\alpha_3 = 0.0479316$ y $\alpha_4 = 2.60707$. Fuente: Matlab.*

En los espacios “Inicial I3” e “Inicial I4” se ingresan las coordenadas de un punto de la forma (λ_3, λ_4) cercano a una intersección. En la figura 10 se observa que los valores iniciales para comenzar el algoritmo son $\lambda_{30} = 0.2$ y $\lambda_{40} = 0.25$, obteniendo los valores $\hat{\lambda}_1 = -0.111667$, $\hat{\lambda}_2 = 0.284952$, $\hat{\lambda}_3 = 0.197614$, $\hat{\lambda}_4 = 0.256231$, el error absoluto de $8.12248e^{-10}$ y la gráfica de la función de densidad de probabilidad de una DLG(-0.111667,0.284952,0.197614,0.256231).

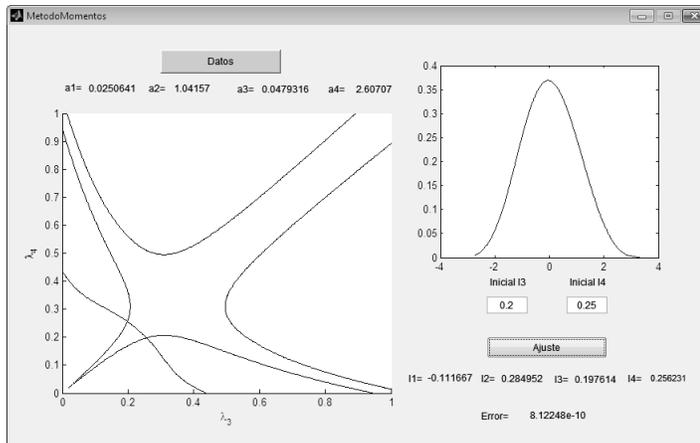


Figura 10: *Parámetros estimados para el conjunto de datos. Fuente: Matlab.*

Si el usuario decide ingresar los momentos directamente, se selecciona “Momentos” y la ventana emergente es la figura 11.

Se ingresan los momentos en los espacios $a1$, $a2$, $a3$ y $a4$ y se da presiona el botón

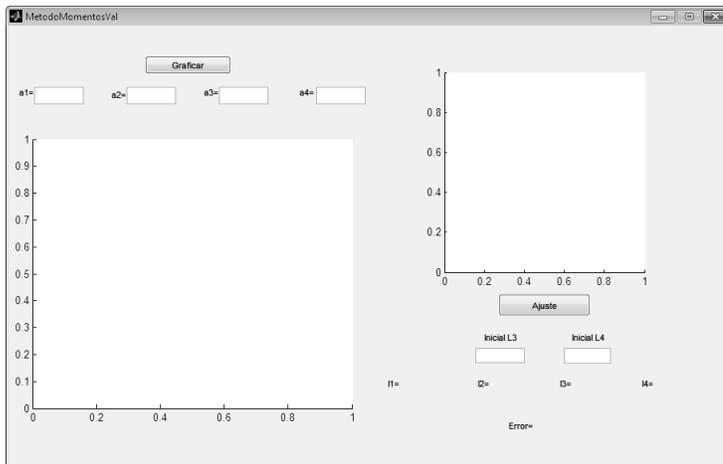


Figura 11: Momentos. Fuente: Matlab.

“Graficar”, obteniendo las curvas $\alpha_3^2 = a3^2$ y $\alpha_4 = a4$. En la figura 12 se observan las curvas para $a3 = 0.01$ y $a4 = 1.8$.

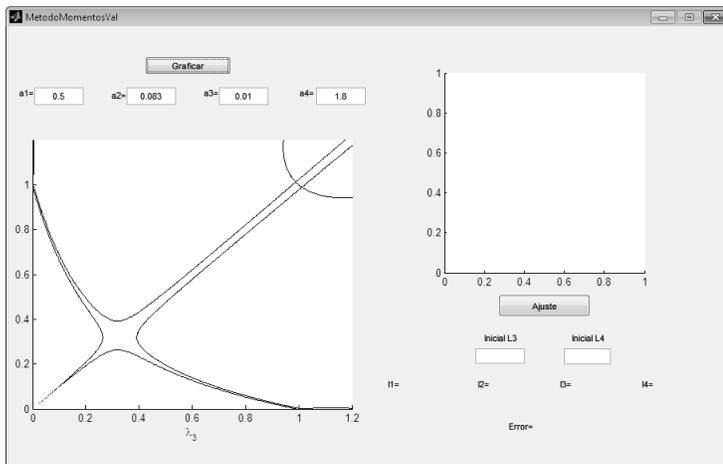


Figura 12: Curvas α_3^2 y α_4 por momentos. Fuente: Matlab.

Se ingresa el valor “Inicial l3” e “Inicial l4” próximos a la intersección de las curvas, y se presiona “Ajuste” para obtener los parámetros de la DLG, el error de la solución y la gráfica de $(x, f(x))$. La figura 13 muestra la ventana resultante.

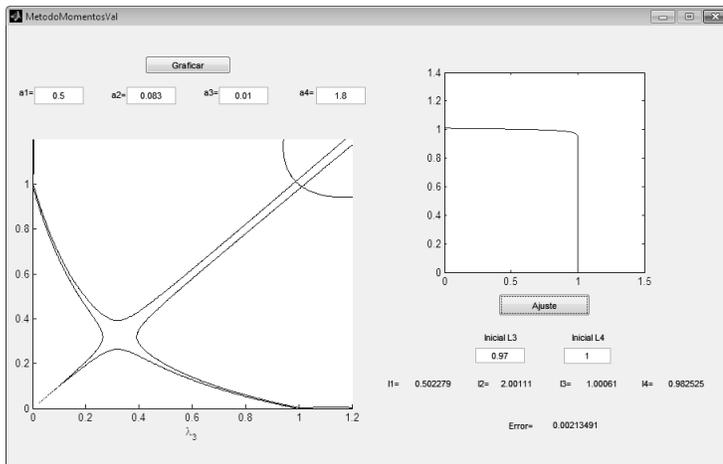


Figura 13: *Ajuste a partir de conjunto de datos. Fuente: Matlab.*

6. Ejemplos

Se tomaron dos ejemplos para variables aleatorias con diferente distribución. El primer ejemplo se tomó con un conjunto de 1000 datos generados con Excel cuya distribución es normal estándar, la figura 10 presenta las dos curvas $\alpha_3 = \hat{\alpha}_3$ y $\alpha_4 = \hat{\alpha}_4$ y la gráfica de la función de densidad de probabilidad de una DLG para este conjunto de datos. La figura 14 muestra la función de densidad de probabilidad de la DLG(-0.1116,0.2849,0.1976,0.2562) y el histograma de frecuencias relativas del conjunto de datos.

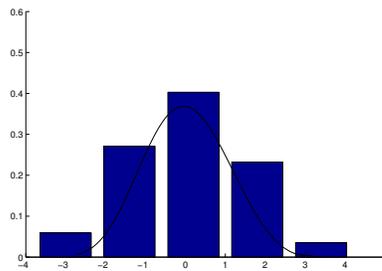


Figura 14: *Histograma. Fuente: Elaboración propia.*

El último ejemplo hace uso del procedimiento presentado para ajustar la DLG a una distribución Weibull (1,5), la figura 13 presenta los estimadores para la distribución Weibull con parámetros $\alpha = 1$, $\beta = 5$ cuyos momentos alrededor de la media son $\hat{\alpha}_1 = 0.9181$, $\hat{\alpha}_2 = 0.0442$, $\hat{\alpha}_3 = -0.2541$, y $\hat{\alpha}_4 = 2.8802$, y la función de densidad de probabilidad de la DLG(0.9934,1.0491,0.2121,0.1061). La figura 15

presenta las dos funciones de densidad de probabilidad.

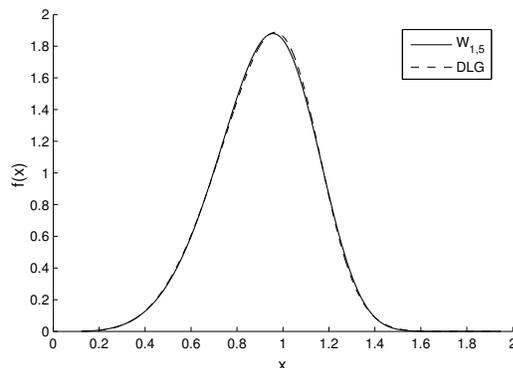


Figura 15: *Distribuciones Weibull y DLG. Fuente: Elaboración propia.*

Una forma de verificar el buen ajuste de la DLG a otras distribuciones es considerar la distancia entre las f.d.p. de la DLG $\hat{f}(x)$ y la distribución a aproximar $f(x)$. Dicha distancia es determinada por $\sup |\hat{f}(x) - f(x)|$. Para la programación de dicha distancia se utiliza esta aproximación:

$$\sup |\hat{f}(x) - f(x)| \approx \max |\hat{f}(x_i) - f(x_i)|, \quad (28)$$

Donde los $x_i = Q(y_i)$ y pertenecen al soporte de la DLG (Karian & Dudewicz 2010). Al verificar el ajuste de la DLG a la distribución Weibull, se obtuvo

$$\sup |\hat{f}(x) - f(x)| = 0.0353$$

con soporte $[0.0413, 1.9475]$.

Cabe anotar que el uso primordial para la DLG es la generación de números pseudo-aleatorios para problemas de simulación y que esté definida a partir de su función percentil permite mediante el uso del teorema de la función de distribución inversa transformar variables aleatorias uniforme $(0,1)$ a otras distribuciones específicas.

Como se observa, el programa “**Selección**” que implementa la estimación de parámetros de la familia DLG vía método de momentos, permite ajustar distribuciones estadísticas a diferentes conjuntos de datos o distribuciones conocidas, como se ilustró con los dos ejemplos anteriores.

7. Conclusiones

Recurrir al método de momentos para la estimación de los parámetros de la DLG, implica obtener la solución de un sistema de ecuaciones que por métodos analíticos

presenta gran dificultad. Por ello, el algoritmo que se implementa para obtener las estimaciones se basa en métodos numéricos y en su respectiva programación.

El conocimiento de la DLG, las restricciones por su forma funcional y las restricciones del método de momentos, permitieron construir un programa en Matlab llamado “Seleccion” para estimar los parámetros de una DLG que se ajuste a un conjunto de datos.

Los ajustes a conjuntos de datos utilizando el programa “Seleccion” que estima los parámetros de la DLG($\lambda_1, \lambda_2, \lambda_3, \lambda_4$) vía método de momentos de forma numérica presenta una buena aproximación tanto a dichos conjuntos, como a distribuciones conocidas, haciendo del algoritmo y el programa “Seleccion” una herramienta útil para estimar los parámetros de la DLG.

Recibido: 11 de junio del 2015

Aceptado: 1 de septiembre del 2015

Referencias

- Casella, G. & Berger, R. L. (2002), *Statistical inference*, Duxbury Pacific Grove, CA, New York.
- Castillo, C. (2015), Estimación de parámetros de la distribución lambda generalizada vía método de momentos, Master’s thesis, Universidad Distrital Francisco José de Caldas.
- Corrado, C. J. (2001), ‘Option pricing based on the generalized lambda distribution’, *Journal of Futures Market*, February **21**, 213–236.
- Karian, Z. A. & Dudewicz, E. J. (2010), *Handbook of fitting statistical distributions with R*, CRC Press, New York.
- Karvanen, J. & Nuutinen, A. (2008), ‘Characterizing the generalized lambda distribution by l-moments’, *Computational Statistics & Data Analysis* **52**, 1971–1983.
- Másmela, L. A. & Rodríguez, H. F. (2012), ‘Estimación de parámetros de la distribución lambda generalizada a partir del método de percentiles’, *Comunicaciones en Estadística* **5**(1), 81–95.
- Mathews, J. H. & Fink, K. D. (1999), *Numerical methods using MATLAB*, Prentice hall Upper Saddle River, NJ, California, USA.
- Mood, A. M. (1950), *Introduction to the Theory of Statistics*, McGraw-Hill, New York.
- Pearson, K. (1894), ‘Contributions to the mathematical theory of evolution’, *Philosophical Transactions of the Royal Society of London. A* **1**, 71–110.

- Pregibon, D. (1980), 'Goodness of link tests for generalized linear models', *Journal of the Royal Society. Applied statistics* **29**, 15–23.
- Ramberg, J. S. & Schmeiser, B. W. (1974), 'An approximate method for generating asymmetric random variables', *Communications of the ACM* **17**(2), 78–82.
- Silver, E. A. (1977), *Operational Research Quarterly* **28**, 743–746.
- Su, S. (2007), 'Numerical maximum log likelihood estimation for generalized lambda distributions', *Computational Statistics & Data Analysis* **51**(8), 3983–3998.
- Tukey, J. W. (n.d.), The practical relationship between the common transformations of percentages of counts and of amounts, Technical report, Princeton University.
- Wackerly, Dennis D, M. W. & Scheaffer, R. L. (2010), *Estadística matemática con aplicaciones*, Cengage Learning Editores, Florida, USA.
- Walpole, R. E., Myers, R. H., Myers, S. L. & Ye, K. (1993), *Probability and Statistics for engineers and scientists*, Macmillan New York, New York.