

---

## Estimación de parámetros de la Distribución Lambda generalizada a partir del método de percentiles

Parameter estimation in the generalized Lambda distribution by using the percentile method

Luis Alejandro Másmela Caita<sup>a</sup>  
lmasmela@udistrital.edu.co

Héctor Fabián Rodríguez Mayorga<sup>b</sup>  
hfrodriguez@gmail.com

---

### Resumen

La Distribución Lambda Generalizada es una distribución de cuatro parámetros definida mediante su función percentil. Su forma funcional permite modelar diversos conjuntos de datos además de una amplia gama de distribuciones. Para ello se requiere de la estimación de sus cuatro parámetros a través de diferentes métodos entre los que se tienen: método de momentos, método de mínimos cuadrados, método de máxima verosimilitud, entre otros. En este documento se pretende hacer una descripción del método de estimación de la distribución de interés, denominado método de percentiles. Se ilustra además su aplicación, ajustando esta distribución a un conjunto de datos además de aproximar una distribución teórica particular.

**Palabras clave:** Ajuste de datos, aproximación de distribuciones, distribución Lambda Generalizada, estimación de parámetros, método de percentiles.

### Abstract

The Generalized Lambda Distribution (GLD) is a four-parameter distribution defined by a percentile function. Its functional form allows to model different data sets and a wide range of distributions. The four parameters estimation may be carried out through different methods such as: moments method, least squares method, maximum likelihood method, among others. This paper wants to describe the method for estimating the distribution of interest, called the Method of Percentiles. An illustration is shown, fitting the distribution to data sets well approximate to a particular theoretical distribution.

**Key words:** Approximation of distributions, data fit, generalized Lambda distribution, method of percentiles, parameter estimation.

---

<sup>a</sup>Docente, Universidad Distrital Francisco José de Caldas.

<sup>b</sup>Estudiante Universidad Distrital Francisco José de Caldas.

## 1. Introducción

Familias de distribuciones de probabilidad generalizadas que, bajo una única forma funcional, abarquen como casos particulares otras más, variando los valores de sus parámetros, se convierten en una herramienta deseable cuando de modelamiento se trata. Distribuciones de este tipo encuentran un protagonismo esencial en áreas como la generación de variables aleatorias para estudios de simulación, estudios a través de muestreo de Monte Carlo para robustez de procedimientos de tipo estadístico o representación de datos cuando el modelo fundamental es desconocido. Varias de estas familias de distribuciones generalizadas han sido desarrolladas, estudiadas y utilizadas en las áreas mencionadas. Algunos ejemplos, entre otros, son el sistema de Pearson y el sistema de Johnson (Johnson & Kotz 1970), o la distribución Lambda de Tukey (Tukey 1960).

La Distribución Lambda Generalizada (*DLG*), propuesta originalmente por Ramberg, Tadikamalla, Dudewicz y Mykytka en (Ramberg et al. 1979), es una generalización de cuatro parámetros de la familia Lambda de Tukey que ha demostrado ser útil en un número de aplicaciones diferentes, como lo son el modelamiento biológico, fenómenos físicos, ingeniería y economía, entre otros. Dado que puede tomar variedad de formas, la *DLG* ofrece gran flexibilidad en el modelado de una amplia gama de datos. Debido a su versatilidad, sin embargo, la estimación de los parámetros adecuados para la *DLG* puede ser un problema difícil. Por tal razón es de gran importancia utilizar un método que se adapte de buena manera a los datos y ofrezca una excelente precisión en la estimación de dichos parámetros.

Diversos métodos de estimación de sus parámetros se han propuesto en la literatura. El método de Momentos Karian et al. (1996), el método de L-momentos Karvanen & Nuutinen (2008), el método de Maxima Verosimilitud Steve (2007), entre otros, son métodos que buscan estimar los cuatro parámetros de los cuales depende la *DLG*. El método en el que aquí nos centramos corresponde al Método de Percentiles presentado por Karian & Dudewicz (1999) y obtenido por los autores a partir de verificar algunas restricciones que tiene el método de momentos.

El documento que se desarrolla a continuación, presenta, en la Sección 2, una corta descripción de la *DLG*, su espacio de parámetros y sus formas funcionales. En la Sección 3 se presenta una descripción del método de interés, esto es, el método de percentiles. Por último, en la Sección 4 se implementa dicho método ajustando la distribución *DLG* a un conjunto de datos; de igual manera, como aplicación se muestra la aproximación a una distribución teórica particular.

## 2. Distribución Lambda Generalizada

La familia de Distribución Lambda Generalizada,  $DLG(\lambda_1, \lambda_2, \lambda_3, \lambda_4)$ , es una distribución de cuatro parámetros. Se especifica a través de su función percentil

$$F^{-1}(y) = F^{-1}(y; \lambda_1, \lambda_2, \lambda_3, \lambda_4) = \lambda_1 + \frac{y^{\lambda_3} - (1-y)^{\lambda_4}}{\lambda_2}, \quad (1)$$

donde  $0 \leq y \leq 1$ . En esta distribución los parámetros  $\lambda_1$  y  $\lambda_2$  son, respectivamente, parámetros de localización y escala, mientras que  $\lambda_3$  y  $\lambda_4$ , juntos, determinan el sesgo y la curtosis.

Un resultado cuya prueba es bastante sencilla y que puede consultarse en Karian & Dudewicz (2011), permite obtener una expresión para la función de densidad de la *DLG* a partir de su función percentil. La expresión para la función de densidad de la *DLG*( $\lambda_1, \lambda_2, \lambda_3, \lambda_4$ ) está dada por

$$f(x) = \frac{\lambda_2}{\lambda_3 y^{\lambda_3-1} + \lambda_4 (1-y)^{\lambda_4-1}}, \text{ con } x = F^{-1}(y). \quad (2)$$

Para que la función en (1) especifique una distribución válida, se deben establecer condiciones sobre ella. En particular, una función  $f(\cdot)$  es una función de densidad de probabilidad si y solo si satisface el par de condiciones:

$$f(x) \geq 0, \quad (3)$$

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

De (2) se observa que, para la familia *DLG*, las condiciones (3) y (4) se satisfacen si y solo si,

$$\frac{\lambda_2}{\lambda_3 y^{\lambda_3-1} + \lambda_4 (1-y)^{\lambda_4-1}} \geq 0, \quad (5)$$

$$\int_{-\infty}^{\infty} f(F^{-1}(y)) dF^{-1}(y) = 1. \quad (6)$$

Ya que la segunda condición (6) es sencilla de probar, nos centraremos en la primera condición, esta permite hacer una descripción del espacio de parámetros de la *DLG*. Debido a que se requiere que

$$\frac{\lambda_2}{\lambda_3 y^{\lambda_3-1} + \lambda_4 (1-y)^{\lambda_4-1}} \geq 0, \quad (7)$$

esta condición se cumple siempre que, tanto numerador como denominador, en la expresión a mano izquierda de (7) tengan el mismo signo. Es así como, al poder realizar una elección libre de  $\lambda_2$ , dicha condición se satisface si la función definida a continuación,

$$g(y; \lambda_3, \lambda_4) := \lambda_3 y^{\lambda_3 - 1} + \lambda_4 (1 - y)^{\lambda_4 - 1},$$

tiene el mismo signo que  $\lambda_2$  para todo  $0 \leq y \leq 1$ . Observamos aquí que el espacio de parámetros de la *DLG* queda supeditado al conjunto de valores que puede tomar  $\lambda_3$  y  $\lambda_4$ ; de esta forma, al referirnos al espacio de parámetros de la *DLG*, hablaremos de regiones el *espacio*– $(\lambda_3, \lambda_4)$ . Un análisis completo de este espacio puede estudiarse en Karian & Dudewicz (2011).

La figura (1) ilustra las regiones para las cuales la *DLG* es válida. Dichas regiones aparecen en el gráfico etiquetadas como Regiones 1 a 6 y muestran aquellos valores para los cuales, las parejas  $(\lambda_3, \lambda_4)$  determinan una *DLG* válida.

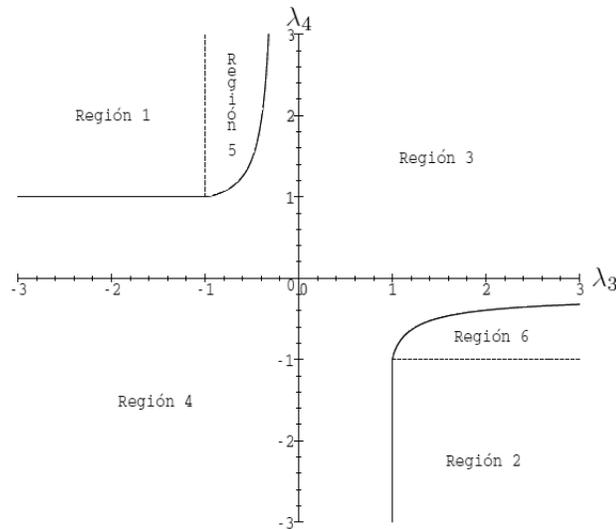


Figura 1: Regiones válidas para la *DLG*

La obtención de gráficas de la  $f(x)$  para la familia *DLG*, se realiza de manera diferente, se construyen parejas de la forma  $(F^{-1}(y), f(x))$  recordando que  $x = F^{-1}(y)$ . Así se toma una secuencia de valores de  $y$  (tales como 0.01, 0.02, 0.03, ..., 0.99, los cuales darán los puntos del 1%, 2%, 3%, ..., 99%), y se encuentran los valores de  $x$  correspondientes a cada uno de estos mediante la ecuación (1), luego se halla  $f(x)$  para cada  $x$  con base en (2), utilizando los valores de  $y$ . De esta forma, los pares  $(x, f(x))$  se grafican y se unen mediante una curva suave. La gráfica en la figura (2) aproxima, asignando valores a los parámetros  $\lambda_1, \lambda_2, \lambda_3$  y  $\lambda_4$ , la distribución normal estándar y la distribución Weibull con parámetros  $\alpha = 1$  y  $\beta = 5$ .

### 3. Ajuste de la *DLG* mediante el método de percentiles

En el marco del ajuste de datos a través de distribuciones, surge el método de percentiles, como respuesta a la restricciones que tiene el método de momentos para conjuntos de datos o distribuciones sobre las cuales no se puede garantizar la existencia de los cuatro momentos indispensables en este método. El método se basa en el concepto de *percentil*, también *cuartil* o *decil*, a través del cual puede caracterizarse una distribución por su localización utilizando la mediana y su dispersión mediante el rango intercuartil.

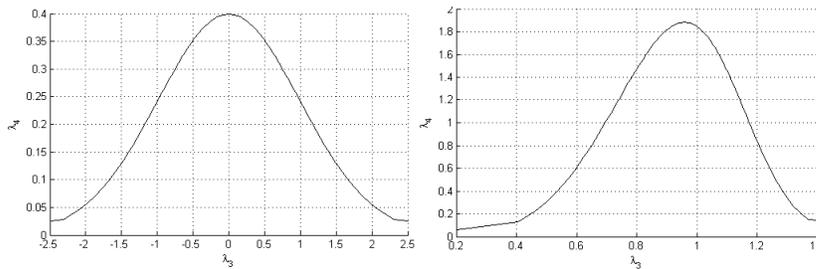


Figura 2: Aproximación de la distribución normal estándar y Weibull a través de la *DLG*

El método propuesto en Karian & Dudewicz (1999), busca ajustar una *DLG* a un conjunto de datos, encontrando el valor de los cuatro parámetros  $\lambda_1, \lambda_2, \lambda_3, \lambda_4$  mediante un método que iguala cuatro estadísticas muestrales a sus cuatro contrapartes para la *DLG*, resolviendo para los valores  $\lambda_1, \lambda_2, \lambda_3, \lambda_4$  y esperando que el resultado genere una distribución *DLG* válida.

Para describir este proceso, partimos de dos definiciones importantes: la definición de percentil en una distribución teórica y su respectiva contraparte en una muestra o conjunto de datos. La literatura estadística, define el  $p$ -ésimo percentil en una distribución, partiendo de una variable aleatoria  $X$  continua, con función de probabilidad  $f(x)$ , función de distribución acumulada  $F(x)$  y siendo  $0 < p < 1$ . El percentil  $(100p)$ -ésimo de la distribución de  $X$  se define como el valor  $\pi_p$  tal que,

$$F(\pi_p) = P(X \leq \pi_p) = \int_{-\infty}^{\pi_p} f(t)dt.$$

Su respectivo estimador puede definirse como aparece en la expresión (8), partiendo de un conjunto de datos  $X_1, X_2, \dots, X_n$ , y denotando por  $\hat{\pi}_p$  el  $(100p)$ -ésimo percentil de los datos. Para el cálculo de  $\hat{\pi}_p$  se escribe  $(n+1)p$  como  $r + (a/b)$ , donde  $r$  es un entero positivo y  $a/b$  es una fracción propia, posiblemente cero. Si  $Y_1, Y_2, \dots, Y_n$  son las estadísticas de orden de  $X_1, X_2, \dots, X_n$ , entonces  $\hat{\pi}_p$  puede obtenerse como,

$$\hat{\pi}_p = Y_r + \frac{a}{b} (Y_{r+1} - Y_r). \quad (8)$$

El método propio, define inicialmente los siguientes cuatro parámetros,

$$\rho_1 = \pi_{0.5}, \quad (9)$$

$$\rho_2 = \pi_{1-u} - \pi_u, \quad (10)$$

$$\rho_3 = \frac{\pi_{0.5} - \pi_u}{\pi_{1-u} - \pi_{0.5}}, \quad (11)$$

$$\rho_4 = \frac{\pi_{0.75} - \pi_{0.25}}{\rho_2}, \quad (12)$$

siendo  $u$  un número entre 0 y  $1/4$ . Los parámetros definidos mediante las expresiones (9) a (12) caracterizan una distribución de la siguiente forma: si  $u = 0.1$ ,  $\rho_1$  define la mediana;  $\rho_2$  define el rango inter-decil, esto es, el rango entre el percentil 10 y el percentil 90;  $\rho_3$  es el cociente del peso de las colas izquierda y derecha y da una medida relativa de los pesos de ambas colas, en el numerador la distancia entre la mediana y el percentil 10 mientras que en el denominador se tiene la distancia entre el percentil 90 y la mediana; por último,  $\rho_4$  es el factor peso-cola o cociente del rango intercuartil y el rango interdecil, el cual es menor que 1 y mide qué tan grande es el peso de la cola (valores próximos a 1 indican que la distribución no se extiende en gran medida en sus colas, mientras que valores próximos a 0 indican que las colas son largas).

Partiendo de la función percentil de la *DLG* en (1), se define  $\rho_1, \rho_2, \rho_3$  y  $\rho_4$  para la *DLG* de la siguiente manera:

$$\rho_1 = F^{-1}\left(\frac{1}{2}\right) = \lambda_1 + \frac{\left(\frac{1}{2}\right)^{\lambda_3} - \left(\frac{1}{2}\right)^{\lambda_4}}{\lambda_2}, \quad (13)$$

$$\rho_2 = F^{-1}(1-u) - F^{-1}(u) = \frac{(1-u)^{\lambda_3} - u^{\lambda_4} + (1-u)^{\lambda_4} - u^{\lambda_3}}{\lambda_2}, \quad (14)$$

$$\rho_3 = \frac{F^{-1}\left(\frac{1}{2}\right) - F^{-1}(u)}{F^{-1}(1-u) - F^{-1}\left(\frac{1}{2}\right)} = \frac{(1-u)^{\lambda_4} - u^{\lambda_3} + \left(\frac{1}{2}\right)^{\lambda_3} - \left(\frac{1}{2}\right)^{\lambda_4}}{(1-u)^{\lambda_3} - u^{\lambda_4} + \left(\frac{1}{2}\right)^{\lambda_4} - \left(\frac{1}{2}\right)^{\lambda_3}}, \quad (15)$$

$$\rho_4 = \frac{F^{-1}\left(\frac{3}{4}\right) - F^{-1}\left(\frac{1}{4}\right)}{\rho_2} = \frac{\left(\frac{3}{4}\right)^{\lambda_3} - \left(\frac{1}{4}\right)^{\lambda_4} + \left(\frac{3}{4}\right)^{\lambda_4} - \left(\frac{1}{4}\right)^{\lambda_3}}{(1-u)^{\lambda_3} - u^{\lambda_4} + (1-u)^{\lambda_4} - u^{\lambda_3}}. \quad (16)$$

Las siguientes consecuencias se obtienen directamente de las definiciones de  $\rho_1, \rho_2, \rho_3, \rho_4$ :

1. Ya que  $\lambda_1$  puede asumir cualquier valor real, puede observarse de (13) que  $\rho_1$  puede tomar también cualquier valor real.
2. Ya que  $0 < u < 1/4$ , se tiene que  $u < 1-u$  y de (14) se puede ver que  $\rho_2 \geq 0$ .

3. El numerador y el denominador de  $\rho_3$  en (15) son ambos positivos; por tanto,  $\rho_3 \geq 0$ .
4. En (16), ya que de la restricción sobre  $u$ , el denominador de  $\rho_4$  puede ser tan grande o igual que su numerador, esto permite concluir que  $0 \leq \rho_4 \leq 1$ .

Puede verificarse además, al considerar  $\rho_3 = \rho_3(u, \lambda_3, \lambda_4)$  y  $\rho_4 = \rho_4(u, \lambda_3, \lambda_4)$  como funciones de  $u$ ,  $\lambda_3$  y  $\lambda_4$  que,

$$\rho_3(u, \lambda_3, \lambda_4) = \frac{1}{\rho_3(u, \lambda_4, \lambda_3)} \quad (17)$$

y

$$\rho_4(u, \lambda_3, \lambda_4) = \rho_4(u, \lambda_4, \lambda_3). \quad (18)$$

### 3.1. El Espacio— $(\rho_3, \rho_4)$ de la $DLG(\lambda_1, \lambda_2, \lambda_3, \lambda_4)$

El ajuste de la  $DLG(\lambda_1, \lambda_2, \lambda_3, \lambda_4)$  a un conjunto de datos  $X_1, X_2, \dots, X_n$  mediante el método de percentiles, se basa en resolver el sistema de ecuaciones  $\hat{\rho}_i = \rho_i$  ( $i = 1, 2, 3, 4$ ) para  $\lambda_1, \lambda_2, \lambda_3$  y  $\lambda_4$ , donde  $\hat{\rho}_i$  se obtiene a partir de los datos reemplazando en las expresiones (9) a (12)  $\pi_i$  por  $\hat{\pi}_i$ . El planteamiento de este sistema, tiene la ventaja de que el subsistema  $\hat{\rho}_3 = \rho_3$  y  $\hat{\rho}_4 = \rho_4$  está dado en función de solamente  $\lambda_3$  y  $\lambda_4$ . Al resolver este subsistema y obtener las soluciones para  $\lambda_3$  y  $\lambda_4$ , puede utilizarse la ecuación  $\hat{\rho}_2 = \rho_2$ , de donde, con los valores anteriores es posible obtener el valor para  $\lambda_2$ . Por último, con estos tres valores obtenidos, se utiliza la ecuación  $\hat{\rho}_1 = \rho_1$  para tener  $\lambda_1$  y de esta forma los valores para los cuatro parámetros que definen la  $DLG$ .

Para la solución del subsistema  $\hat{\rho}_3 = \rho_3$ ,  $\hat{\rho}_4 = \rho_4$ , es necesario especificar un valor para  $u$ . Si tomamos un valor particular de  $u$  se debe tener que  $(n+1)u \geq 1$  para poder calcular  $\hat{\pi}_u$  y  $\hat{\pi}_{1-u}$  y por tanto  $\hat{\rho}_2, \hat{\rho}_3$  y  $\hat{\rho}_4$ . Si  $u$  es muy pequeño, entonces el método se debe restringir a muestras grandes ( $n \geq 99$  para el caso de  $u = 0.01$ ), al tratar de resolver el problema tomando valores de  $u$  próximos a  $1/4$ , el valor de  $\rho_4$  se aproxima a 1 debido a que numerador y denominador de  $\rho_4$  se aproximan mucho entre sí y el objetivo de tomar  $\rho_4$  como una medida para el peso de las colas de la distribución no tiene sentido. De esta forma, el valor para  $u$  se asume como  $u = 0.1$ .

Debido a que la solución de las ecuaciones referidas anteriormente, arrojan como resultado valores para los parámetros  $\lambda_1, \lambda_2, \lambda_3$  y  $\lambda_4$  y de manera particular resultan dependiendo directamente de la solución del subsistema  $\hat{\rho}_3 = \rho_3, \hat{\rho}_4 = \rho_4$  que depende exclusivamente de los valores de los parámetros  $\lambda_3$  y  $\lambda_4$ , el interés recae en analizar si nuestro subsistema permite o no soluciones. Para cumplir con tal objetivo se analiza el mapeo del *espacio*— $(\lambda_3, \lambda_4)$  (más exactamente las regiones de este espacio en donde la  $DLG$  es una distribución válida) en el *espacio*— $(\rho_3, \rho_4)$ , lo que significa analizar la transformación definida por las ecuaciones (15) y (16).

El análisis parte de observar cómo las regiones válidas para el espacio de parámetros de la *DLG* se mapea en el *espacio*— $(\rho_3, \rho_4)$  mediante la transformación:

$$\rho_3 = \frac{(1-u)^{\lambda_4} - u^{\lambda_3} + \left(\frac{1}{2}\right)^{\lambda_3} - \left(\frac{1}{2}\right)^{\lambda_4}}{(1-u)^{\lambda_3} - u^{\lambda_4} + \left(\frac{1}{2}\right)^{\lambda_4} - \left(\frac{1}{2}\right)^{\lambda_3}},$$

$$\rho_4 = \frac{\left(\frac{3}{4}\right)^{\lambda_3} - \left(\frac{1}{4}\right)^{\lambda_4} + \left(\frac{3}{4}\right)^{\lambda_4} - \left(\frac{1}{4}\right)^{\lambda_3}}{(1-u)^{\lambda_3} - u^{\lambda_4} + (1-u)^{\lambda_4} - u^{\lambda_3}}.$$

Las figuras (3), (4) y (5) muestran cómo se mapean las regiones 1, 2, 5 y 6, la Región 3 y la Región 4 del *espacio*— $(\lambda_3, \lambda_4)$  en el *espacio*— $(\rho_3, \rho_4)$  respectivamente. Se puede observar a través del análisis de las figuras (3), (4) y (5), que existen regiones del *espacio*— $(\rho_3, \rho_4)$  que no son cubiertas a través del mapeo, para algunos casos el sistema de ecuaciones  $\hat{\rho}_3 = \rho_3$ ,  $\hat{\rho}_4 = \rho_4$ , no tiene solución, en otros casos existen múltiples soluciones. Se puede observar además que la mayor parte del *espacio*— $(\rho_3, \rho_4)$  se cubre por valores de  $(\lambda_3, \lambda_4)$  en las regiones 3 y 4, mientras que valores en otras regiones no se caracterizan por esto. Por tal motivo se dará preferencia a la búsqueda de soluciones numéricas en las regiones 3 y 4 (ver figura 1).

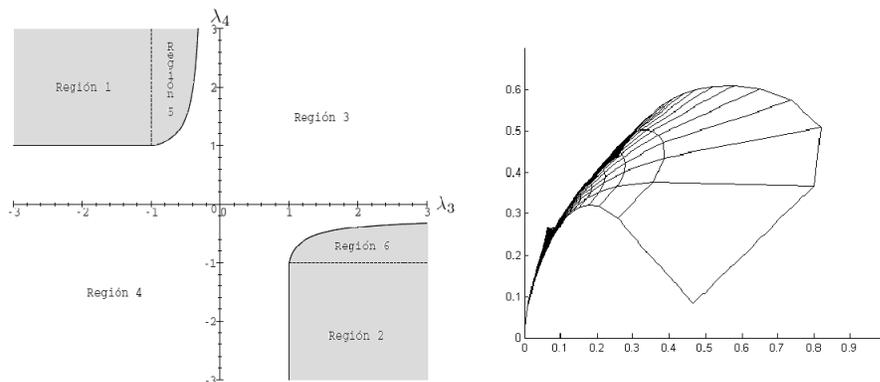


Figura 3: Mapeo de las Regiones 1, 2, 5 y 6 en el *espacio*— $(\rho_3, \rho_4)$

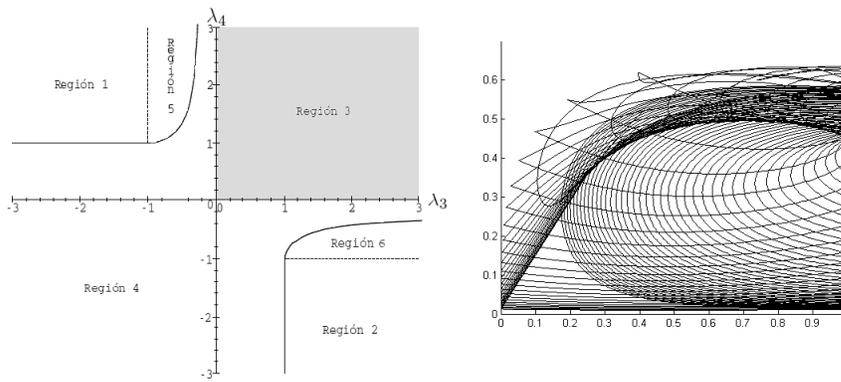


Figura 4: Mapeo de la Región 3 en el espacio  $(\rho_3, \rho_4)$

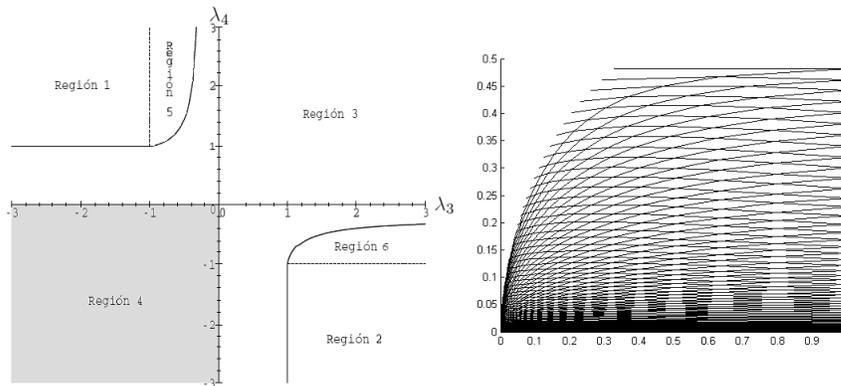


Figura 5: Mapeo de la Región 4 en el espacio  $(\rho_3, \rho_4)$

La existencia de múltiples soluciones se puede argumentar debido a que, al considerar las soluciones asociadas a valores fijos  $(\hat{\rho}_3^*, \hat{\rho}_4^*)$ , cualquier solución a  $\rho_3 = \hat{\rho}_3^*$  y  $\rho_4 = \hat{\rho}_4^*$  puede ser simultáneamente localizada sobre dos superficies.

Suponiendo que  $(\hat{\rho}_3^*, \hat{\rho}_4^*) = (0.4, 0.5)$ , una solución debe estar localizada en la intersección de la superficie  $\rho_4 = 0.5$  y el plano en 0.5 (ver figura 6). De esta forma, todos los puntos en la intersección pueden ser posibles soluciones. De manera similar puede considerarse la intersección de la superficie  $\rho_3 = 0.4$  y el plano horizontal en 0.4. Las soluciones buscadas se localizarán en la intersecciones de las curvas. La gráfica en la figura 7 muestra una serie de cortes que corresponden a las soluciones buscadas y que se ubican en la Región 3. Es posible encontrar otras soluciones dadas en otras regiones ampliando el rango de esta gráfica.

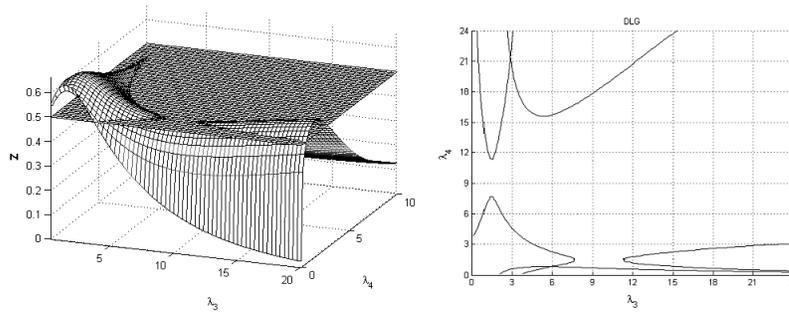


Figura 6: Intersección de la superficie  $\rho_4 = 0.5$  y el plano en 0.5

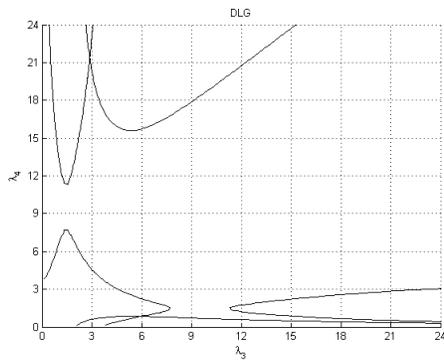


Figura 7: Soluciones en las intersecciones de las curvas  $\rho_3 = 0.4$  y  $\rho_4 = 0.5$  en la Región 3

#### 4. Implementación del método de percentiles

Las ecuaciones  $\hat{\rho}_3 = \rho_3$  y  $\hat{\rho}_4 = \rho_4$  no se pueden resolver en forma cerrada y, en consecuencia, debemos utilizar métodos numéricos. Para proporcionar el acceso a este método, sin tener que pasar a través de algoritmos numéricos para encontrar las soluciones, Karian & Dudewicz (2011) ofrecen varios tipos de tablas que dan soluciones a  $\hat{\rho}_3 = \rho_3$  y  $\hat{\rho}_4 = \rho_4$  en las Regiones 3 y 4. Los cálculos dependen de encontrar soluciones aproximadas donde  $\rho_3, \rho_4$  son funciones de  $\lambda_3$  y  $\lambda_4$ .

A partir de tablas que pueden ser consultadas en Karian & Dudewicz (2011), Apéndice D, el siguiente algoritmo muestra cómo obtener los valores numéricos de  $\lambda_1, \lambda_2, \lambda_3, \lambda_4$  para el ajuste de la *DLG*.

#### 4.1. Algoritmo DLG-P: Ajustando una distribución DLG para datos, método de percentiles

**DLG-P-1.** A partir de un conjunto de datos de tamaño  $n$ , ordenar los datos de menor a mayor y luego usar (9) a (12) para calcular  $\hat{\rho}_1, \hat{\rho}_2, \hat{\rho}_3, \hat{\rho}_4$ .

**DLG-P-2.** Teniendo  $\hat{\rho}_3, \hat{\rho}_4$ , utilizar la tabla (ver Karvanen & Nuutinen (2008), apéndice D) más cercana para encontrar las posibles soluciones de  $(\lambda_3, \lambda_4)$ ; si  $\hat{\rho}_3 > 1$ , usar  $(1/\hat{\rho}_3, \hat{\rho}_4)$  en lugar de  $(\hat{\rho}_3, \hat{\rho}_4)$  e intercambiar  $\lambda_3$  y  $\lambda_4$ .

**DLG-P-3.** Después de haber obtenido  $\lambda_3$  y  $\lambda_4$  de las tablas, reemplazar estos valores en la ecuación  $\hat{\rho}_2 = \rho_2$  para determinar  $\lambda_2$  (recordar que  $u$  debe ser igual a 0.1 para el uso de las tablas).

**DLG-P-4.** Reemplazar  $\lambda_2, \lambda_3$  y  $\lambda_4$  en la ecuación  $\hat{\rho}_1 = \rho_1$ , para obtener  $\lambda_1$ .

Para la estimación de  $\lambda_1, \lambda_2, \lambda_3$  y  $\lambda_4$  con una mayor precisión de la que sería posible a través del algoritmo DLG-P, podemos utilizar un software adecuado. Esta opción nos brinda la facilidad para manipular símbolos matemáticos, la implementación de los métodos numéricos y la ventaja de poder generar un entorno gráfico. El algoritmo que hemos utilizado se basa en el método de Newton. En la aplicación creada aquí, como aporte y complemento al estudio del método (ver Anexo), se calcula  $\hat{\rho}_1, \hat{\rho}_2, \hat{\rho}_3, \hat{\rho}_4$  con base en el conjunto de datos de interés, y se grafican las curvas  $\hat{\rho}_3 = \rho_3, \hat{\rho}_4 = \rho_4$ , con el fin de encontrar valores iniciales de manera visual, para luego correr el algoritmo (por ejemplo, la figura 7 indica la presencia de soluciones cerca de los puntos  $(\lambda_3, \lambda_4) = (3, 21)$  y  $(\lambda_3, \lambda_4) = (6, 1)$ ). Como es de esperar, los valores iniciales son cruciales para el éxito de la obtención de las estimaciones finales. En algunas ocasiones, un cálculo aproximado, como el que se puede obtener gráficamente en la figura 7, proporciona un buen punto de partida. En el caso donde la gráfica no brinda la información suficiente, los valores en las tablas anteriormente mencionadas se pueden utilizar como un valor inicial para obtener resultados de alta precisión.

El ejemplo 1, permite ilustrar la implementación del método de percentiles aplicado a un conjunto de datos aleatorios generados mediante una distribución Beta con parámetros  $a = 2$  y  $b = 4$ .

**Ejemplo 1.** El conjunto de datos fue generado aleatoriamente a partir de una distribución Beta, con parámetros  $a = 2$  y  $b = 4$ .

La manera de abordar este ejemplo a través del método de percentiles, usando el Algoritmo DLG-P e implementando la aplicación permite, inicialmente, hacer el cálculo de  $\hat{\rho}_1, \hat{\rho}_2, \hat{\rho}_3, \hat{\rho}_4$  obteniendo

$$\hat{\rho}_1 = 0.2964, \quad \hat{\rho}_2 = 0.5309, \quad \hat{\rho}_3 = 0.5903, \quad \hat{\rho}_4 = 0.527.$$

El programa permite obtener los valores  $\lambda_1 = 0.1587, \lambda_2 = 0.7038, \lambda_3 = 0.0270$  y  $\lambda_4 = 0.1770$  (ver Anexo). Por tanto, tenemos el ajuste para los datos mediante

Tabla 1: Datos Generados aleatoriamente mediante una distribución  $Beta(2,4)$ 

0.6051	0.2752	0.3323	0.4359	0.5481	0.1937	0.3905	0.0328	0.0675	0.3936
0.2676	0.3229	0.5479	0.433	0.106	0.1173	0.4723	0.1737	0.665	0.28
0.4971	0.4079	0.0753	0.3907	0.1393	0.4846	0.1963	0.2042	0.1905	0.3509
0.1178	0.6547	0.2853	0.3373	0.6722	0.0578	0.1224	0.2271	0.238	0.6324
0.153	0.2511	0.5763	0.1621	0.3623	0.5121	0.1664	0.4277	0.379	0.3708
0.1913	0.5113	0.2272	0.3075	0.1795	0.1742	0.4236	0.2463	0.1378	0.7063
0.7693	0.0842	0.0935	0.5488	0.1567	0.3559	0.1854	0.6297	0.4205	0.196
0.3904	0.4618	0.3923	0.2384	0.1145	0.6855	0.1569	0.2182	0.409	0.0986
0.4354	0.2016	0.6084	0.6811	0.0843	0.0912	0.0919	0.2065	0.2427	0.5364
0.1614	0.3361	0.6303	0.3635	0.1847	0.2467	0.3242	0.2599	0.6573	0.5127

la  $DLG(0.1587, 0.7038, 0.0270, 0.1770)$ . La figura 8 muestra el histograma para el conjunto de datos y su ajuste a través de la  $DLG$  por el método de percentiles.

Cabe anotar que varios ajustes para un mismo conjunto de datos pueden obtenerse a través de este método. Pruebas de bondad de ajuste pueden utilizarse para la evaluación de las  $DLG$ s obtenidas.

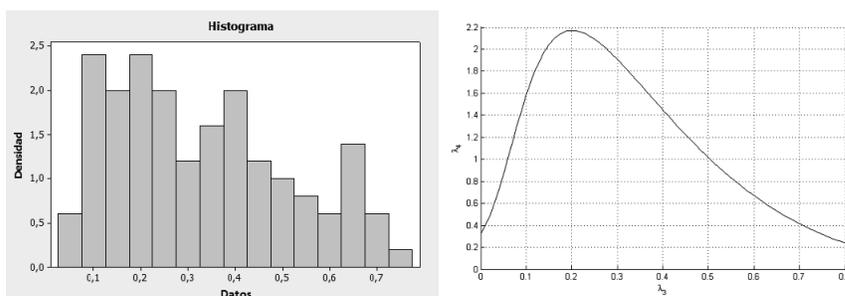


Figura 8: Histograma para los datos y su ajuste  $DLG(0.1587, 0.7038, 0.0270, 0.1770)$

Otra utilidad del método es hacer aproximaciones de distribuciones conocidas. En el ejemplo que se expone a continuación, se aproxima una distribución normal estándar a través de una  $DLG$  utilizando el método de percentiles.

**Ejemplo 2.** La función de densidad de probabilidad (fdp) de la distribución normal estándar está dada por,

$$f(x) = \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{(x)^2}{2}\right], \quad -\infty < x < \infty.$$

Para esta distribución los parámetros  $\rho_1, \rho_2, \rho_3, \rho_4$  están dados por,

$$\rho_1 = 0, \quad \rho_2 = 2.56, \quad \rho_3 = 1, \quad \rho_4 = 0.53.$$

Al igualar estos valores a los parámetros de la  $DLG$  y solucionar el sistema obtenido se obtiene la aproximación observada en la figura 9 dada por

$$DLG(0, 0.2142, 0.1488, 0.1488),$$

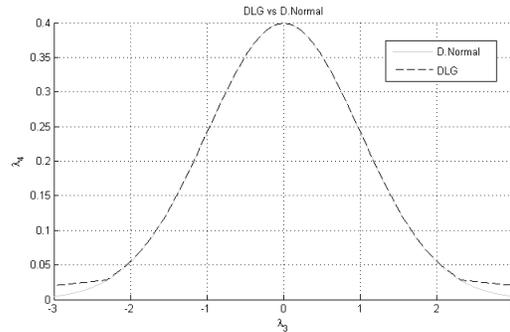


Figura 9: Aproximación de la  $N(0,1)$  a partir de la DLG

Para este tipo de aproximaciones los autores del método Karian & Dudewicz (1999) definen una forma de evaluar el error en el ajuste.

## 5. Conclusiones

La Distribución Lambda Generalizada es un modelo de distribución versátil a la hora de modelar conjuntos de datos. El estudio de su espacio de parámetros y formas funcionales que esta distribución puede generar, permitió observar una amplia variedad de formas que puede aproximar. Además, el que su forma funcional dependa de cuatro parámetros contribuye a su gran versatilidad en cuanto a estas aproximaciones, logrando realizar esta tarea mediante diferentes métodos de estimación. Dentro de los varios métodos de estimación que existen para sus parámetros, se estudió el método denominado de percentiles. Debido a que para la gran mayoría de las distribuciones continuas podemos encontrar sus percentiles teóricos, el método permite una buena precisión cuando de aproximar distribuciones se trata. Asimismo, bajo ciertas condiciones de los percentiles, el método permite cubrir conjuntos de datos sobre los cuales otros métodos no pueden ser implementados.

**Recibido: 8 de diciembre de 2011**

**Aceptado: 1 de abril de 2012**

## Referencias

Johnson, N. L. & Kotz, S. (1970), *Distribution in Statistics: Continuous Univariate Distributions*, New York: John Wiley.

- Karian, Z. & Dudewicz, E. (1999), 'Fitting the generalized lambda distribution to data: a method based on percentiles', *Communications in Statistics - Simulation and Computation*, **28**, 793 – 819.
- Karian, Z. & Dudewicz, E. (2011), *Handbook of Fitting Statistical Distributions with R*, Boca Ratón: CRC Press.
- Karian, Z., Dudewicz, E. & McDonald, P. (1996), 'The extended generalized lambda distribution (egld) system for fitting distributions to data with moments, ii:tables', *The American Journal of Mathematical and Management Sciences* **16**, 271–332.
- Karvanen, J. & Nuutinen, A. (2008), 'Characterizing the generalized lambda distribution by l-moments', *Computational Statistics & Data Analysis* **52**, 1971 – 1983.
- Ramberg, J. S., Tadikamalla, P. R., Dudewicz, E. J. & Mykytka, E. F. (1979), 'A probability distribution and its uses in fitting data', *Technometrics* **21**, 201 – 214.
- Steve, S. (2007), 'Numerical maximum log likelihood estimation for generalized lambda distribution', *Computational Statistics and Data Analysis* **51**, 3983 – 3998.
- Tukey, J. (1960), The practical relation between the common transformation of percentages of counts and of amounts, Technical Report 36, Princeton University.

A. Anexo

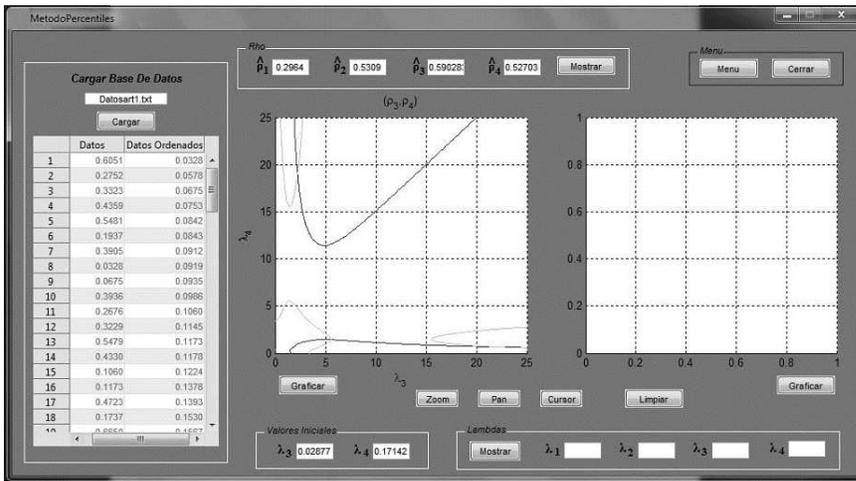


Figura 10: Método de percentiles

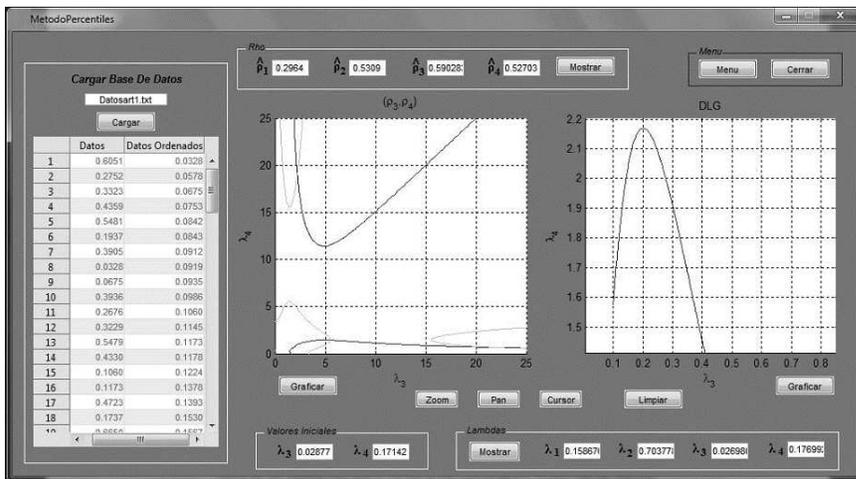


Figura 11: Método de percentiles