https://repository.uaeh.edu.mx/revistas/index.php/sahagun/issue/archive

Ingenio y Conciencia Boletín Científico de la Escuela Superior Ciudad Sahagún

Publicación semestral No. 13 (2020) 23-28

Simulación de la distribución uniforme generalizada

Simulation of The Generalized Uniform Distribution

Isidro J. González-Hernández [a], Isaías Simón-Marmolejo [b], Rafael Granillo-Macías [c], Francisca Santa-Robles [d], Carlos Rondero-Guerrero [e], Carlos A. Soto-Campos [f]

Abstract:

The continuous uniform distribution is one of the simplest models in statistical analysis but it has been the basis for developing new distribution families.This paper presents two generalizations or extensions of the continuous uniform distribution developed by Jose and Krishna; Sankaran and Jayakumar, in both cases the authors added new parameters to generate new distribution families. To identify the flexibility of both generalizations, a simulation of the probability density function, the distribution function, and the hazard rate function was carried out. Besides, the simulation results were compared with other existing distribution functions to identify to which other models the two generalizations fit. The results show that the two generalizations have behaviors similar to Gamma, Beta, Johnson, and Pearson distribution functions.

Keywords:

Marshall-Olkin distribution, uniform distribution, generalized uniform distribution

Resumen:

La distribución uniforme continua es uno de los modelos más simples en el análisis estadístico pero ha sido la base para desarrollar nuevas familias de distribución. En este trabajo se presentan dos generalizaciones o extensiones de la distribución uniforme continua desarrolladas por Jose y Krishna; Sankaran y Jayakumar, en ambos casos los autores agregaron nuevos parámetros para generar nuevas familias de distribución. Para identificar la flexibilidad de ambas generalizaciones se llevó a cabo una simulación de la función de densidad de probabilidad, de la función de distribución y de la función de tasa de riesgo. Además, los resultados de la simulación fueron comparados con otras funciones de distribución existentes para identificar a que otros modelos se ajustan las dos generalizaciones. Los resultados muestran que las dos generalizaciones tienen comportamientos similares a funciones de distribución Gamma, Beta, Johnson y Pearson.

Palabras Clave:

D*istribución Marshall-Olkin, distribución uniforme, distribución uniforme generalizada*

Introducción

Cualquier análisis estadístico de un proceso o sistema depende en gran medida del modelo estadístico utilizado para representar el fenómeno en estudio. Por lo tanto, cuanto mayor sea la clase de modelos estadísticos disponibles para analizar el estadístico en cuestión, más fácil será elegir un modelo. Sin embargo, los datos de

[a] Autor de Correspondencia, Universidad Autónoma del Estado de Hidalgo, Escuela Superior Ciudad Sahagún, https://orcid.org/0000-0003-2805-6674, Email: igonzalez@uaeh.edu.mx

[b] Universidad Autónoma del Estado de Hidalgo, Escuela Superior Ciudad Sahagún, https://orcid.org/0000-0003-2116-6192, Email:isaíasm@uaeh.edu.mx

[c] Universidad Autónoma del Estado de Hidalgo, Escuela Superior Ciudad Sahagún: https://orcid.org/0000-0002-1015-667X, Email: rafaelgm@uaeh.edu.mx

[d] Universidad Autónoma del Estado de Hidalgo, Escuela Superior Ciudad Sahagún: https://orcid.org/0000-0002-3301-9790, Email: profe_7739@uaeh.edu.mx

[e] Universidad Autónoma del Estado de Hidalgo, Instituto de Ciencias Básicas e Ingeniería: https://orcid.org/0000-0003-0663-8366, Email:ronderocar@gmail.com

[f] Universidad Autónoma del Estado de Hidalgo, Instituto de Ciencias Básicas e Ingeniería https://orcid.org/0000-0001-7874-0306, Email:csoto@uaeh.edu.mx

muchos problemas importantes y prácticos no siguen ninguno de los modelos de probabilidad disponibles. 1
En este sentido, muchos investigadores del área de estadística aplicada se han interesado en desarrollar e introducir nuevas familias de distribuciones o simplemente han generalizado algunas de las distribuciones ya existentes para describir la vida útil de algunos dispositivos o para describir el comportamiento de un conjunto de datos reales. Sin embargo, estas distribuciones tienen un rango limitado de aplicabilidad y no pueden representar todas las situaciones encontradas en las aplicaciones 2.
Por ejemplo, la distribución exponencial a menudo se describe como flexible y su función de riesgo es constante. Por otra parte, las limitaciones de las distribuciones ya conocidas a menudo despiertan el interés de los investigadores por encontrar nuevas distribuciones ampliando las ya existentes. El procedimiento de expandir una familia de distribuciones para mayor flexibilidad o construir modelos de covariables es una técnica bien conocida en la literatura. Por ejemplo, la familia de distribuciones de Weibull contiene la distribución exponencial y se construye tomando potencias de variables aleatorias distribuidas exponencialmente 2.
Una de las distribuciones continuas con el modelo más simple y utilizado es la distribución uniforme de probabilidad continua. La distribución uniforme continua representa una situación en la que todos los resultados en un rango entre un valor mínimo y máximo son igualmente probables. Desde una perspectiva teórica, esta distribución es clave en el análisis de riesgos; muchos algoritmos en diferentes software usan una muestra de esta distribución (entre cero y uno) para generar muestras aleatorias de otras distribuciones, es decir, los valores aleatorios uniformes entre (0,1) son utilizados en el método de la transformada inversa para generar variables aleatorias que se ajustan a una función de distribución de probabilidad específica 3, 4.
Bajo este contexto, la distribución uniforme continua ha sido generalizada para introducir nuevas familias de distribuciones. En este artículo abordamos dos generalizaciones de la distribución uniforme continua desarrolladas en los trabajos de Jose y Krishna, y Sankaran y Jayakumar, en cada una de las generalizaciones simulamos la Función de Densidad de Probabilidad (FDP), la Función de Distribución (FD) y la Función de Tasa de Riesgo (FTR). Además, se utilizará el software ExpertFit para identificar si los resultados de las diferentes simulaciones se ajustan a otras distribuciones de probabilidad ya conocidas. 5, 3
El resto del trabajo está organizado como sigue. En la Sección 2 presentamos el trabajo seminal de Marshall y Olkin (1997) [6]. En la Sección 3 se realiza la simulación del modelo generalizado de la distribución uniforme presentado por Jose y Krishna. En la Sección 4 se realiza la simulación del modelo generalizado de la distribución uniforme presentado por Sankaran y Jayakumar. La sección 5 se presenta una discusión con base en la comparación de las simulaciones utilizando el software ExpertFit. Finalmente se presentan las conclusiones. 5, 3

Trabajo seminal de Marshall y Olkin

Una idea interesante para generalizar una distribución, es un método conocido en la literatura como distribución extendida de Marshall y Olkin. El artículo seminal de Marshall y Olkin, fue el enfoque utilizado en los trabajos de Jose y Krishna, y Sankaran y Jayakumar para generalizar la distribución uniforme continua. 6,5,3
Marshall y Olkin propusieron un método para agregar un nuevo parámetro a una familia de distribución y generar una nueva familia de distribuciones. Donde definen a una variable aleatoria con función de distribución como $G(x)$ y función de densidad de probabilidad como $g(x)$. En su trabajo seminal los autores comenzaron con una función de supervivencia $\bar{F}(x) = P(X > x)$ de una variable aleatoria X y un parámetro $\alpha > 0$, de tal forma que, se consideró una familia de funciones de supervivencia dada por

$$\bar{\mathrm{G}}(\mathrm{x}, \alpha) = \frac{\alpha \bar{\mathrm{F}}(\mathrm{x})}{1 - (1 - \alpha)\bar{\mathrm{F}}(\mathrm{x})} \tag{1}$$

donde, $-\infty < x < \infty$, y $\alpha > 0$

Siendo la función de densidad de probabilidad correspondiente

$$\mathrm{g}(\mathrm{x}, \alpha) = \frac{\alpha \mathrm{f}(\mathrm{x})}{[1 - (1 - \alpha)\bar{\mathrm{F}}(\mathrm{x})]^2} \tag{2}$$

donde $-\infty < x < \infty,\ \alpha > 0$ y $f(x)$ es la correspondiente función de densidad de probabilidad para $\bar{F}(x)$. La función de tasa de riesgo está dada por

$$\mathrm{h}(\mathrm{x}, \alpha) = \frac{\mathrm{r}(\mathrm{x})}{1 - (1 - \alpha)\bar{\mathrm{F}}(\mathrm{x})} \tag{3}$$

donde

$$r(x) = \frac{f(x)}{\bar{F}(x)}$$

Simulación de la Distribución Uniforme Extendida de Marshall-Olkin, Jose y Krishna (2011)

Desde la perspectiva de Jose y Krishna 5, los autores reportan una función de distribución uniforme extendida, introduciendo un parámetro $\theta > 0$, quedando expresada la distribución uniforme como $U(0,\theta)$. Considerando a la función de sobrevivencia como, $\bar{F}\ (x) = 1 - (x/\theta)$, la cual está asociada con una nueva distribución que denotan como MOEU (α, θ). Por tanto, sustituyendo en la función de supervivencia (1) se obtiene

$$\overline{G}(x, \alpha, \theta) = \frac{\alpha(\theta - x)}{\alpha\theta + (1 - \alpha)x}, \tag{4}$$

donde $0 < x < \theta, \quad \alpha > 0$.

De tal forma que, la función de distribución es,

$$G(x, \alpha, \theta) = \frac{x}{\alpha\theta + (1 - \alpha)x}, \tag{5}$$

donde $0 < x < \theta, \quad \alpha > 0$.

La correspondiente función de densidad de probabilidad está dada por

$$g(x, \alpha, \theta) = \frac{\alpha\theta}{[\alpha\theta + (1 - \alpha)x]^2}, \qquad 0 \tag{6}$$

donde $0 < x < \theta, \quad \alpha > 0$.

La función de tasa de riesgo de una variable aleatoria X con distribución MOEU (α, θ) es

$$h(x, \alpha, \theta) = \frac{\theta}{[\alpha\theta + (1 - \alpha)x](\theta - x)} \tag{7}$$

donde $0 < x < \theta, \quad \alpha > 0$.

Un mecanismo conocido para simular números aleatorios que se ajusten al comportamiento de una función de distribución específica, consiste en analizar la función inversa de la función de distribución. Para este caso se utilizará $G(x, \alpha, \theta)$(5), en el caso de la MOEU (α, θ), la función inversa queda expresada como

$$X = F^{-1}\left[\frac{x}{\alpha\theta + (1 - \alpha)x}\right] \tag{8}$$

y mediante un cálculo directo se obtiene,

$$X = \frac{\alpha\theta Y}{1 - (1 - \alpha)Y} \tag{9}$$

donde $Y \sim U(0,1)$.

Para llevar a cabo la simulación de la MOEU se utilizaron diferentes valores de α y θ. Se simularon 1000 números utilizando el software de Minitab. A continuación, se presentan las gráficas de la función de densidad de probabilidad, la función de distribución y la función de tasa de riesgo.

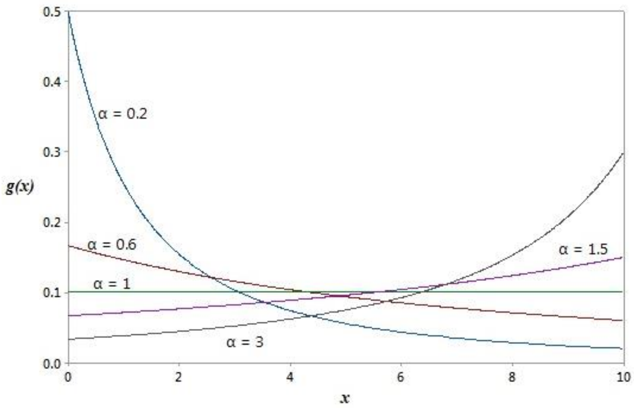

Figura 1. Simulación de la FDP de Jose y Krishna (2011), $\theta = 10$ y diferentes valores de α

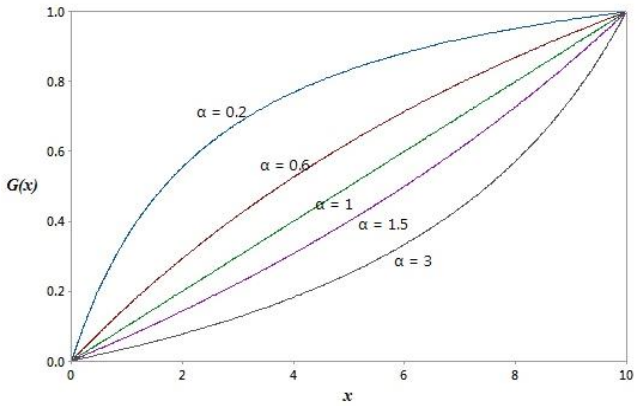

Figura 2. Simulación de la FD de Jose y Krishna (2011), $\theta = 10$ y diferentes valores de α

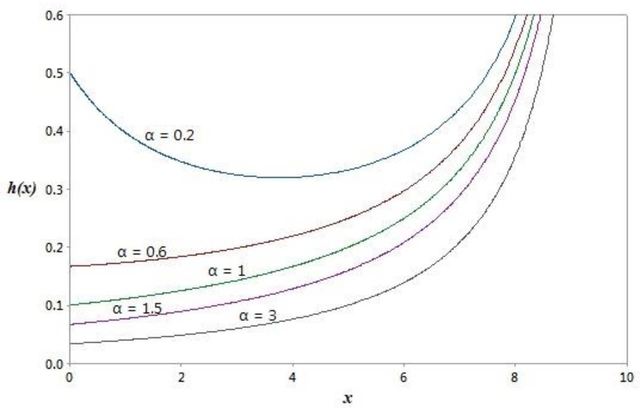

Figura 3. Simulación de la FTR de Jose y Krishna (2011), $\theta = 10$ y diferentes valores de α

Simulación de la Generalización de la Distribución Uniforme de Sankaran y Jayakumar (2016)

Desde la perspectiva de Sankaran y Jayakumar 3, se introduce lo que denominan la GUD (Generalized Uniform Distribution). Esta función se deduce de considerar la distribución binomial truncada negativa. De forma tal que, la función de sobrevivencia reportada por los mismos autores es,

$$\overline{G}(x,\alpha,\theta) = \frac{\alpha^{\theta}}{1-\alpha^{\theta}}\left[[F(x)+\alpha\overline{F}(x)]^{-\theta}-1\right] \qquad (10)$$

donde $\theta > 0$ y $\alpha > 0$.

De tal forma que, la función de distribución está dada por

$$G(x,\alpha,\theta) = \frac{1-\alpha^{\theta}[x(1-\alpha)+\alpha]^{-\theta}}{1-\alpha^{\theta}} \qquad (11)$$

donde $\theta > 0$ y $\alpha > 0$.

La función de densidad de probabilidad correspondiente está dada por

$$g(x,\alpha,\theta) = \frac{(1-\alpha)\theta\alpha^{\theta}}{(1-\alpha^{\theta})[x(1-\alpha)+\alpha]^{\theta+1}} \qquad (12)$$

para $0 < x < 1$, $\theta > 0$ y $\alpha > 0$.

La función de tasa de riesgo de una variable aleatoria X con distribución GUD es

$$h(x,\alpha,\theta) = \frac{\theta(1-\alpha)}{(\alpha+(1-\alpha)x)[1-[\alpha+(1-\alpha)x]^{\theta}]} \qquad (13)$$

Para simular la GUD, también se utilizará el método de la función inversa de la función de distribución. Para este caso se utilizará $G(x,\alpha,\theta)$(11). De tal forma que, la función inversa queda expresada como

$$X = \frac{\alpha}{1-\alpha}\left(\left[1-Y\left(1-\alpha^{\theta}\right)\right]^{-\frac{1}{\theta}}-1\right)$$

Para la simulación de la GUD se utilizaron diferentes valores de α y θ. Se simularon 1000 números utilizando el software de Minitab. A continuación, se presentan las gráficas de la función de densidad de probabilidad, la función de distribución y la función de tasa de riesgo para este modelo.

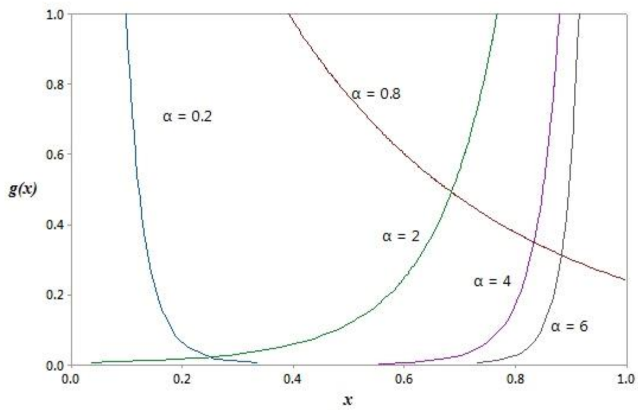

Figura 4. Simulación de la FDP de Sankaran y Jayakumar (2016), $\theta = 10$ y diferentes valores de α

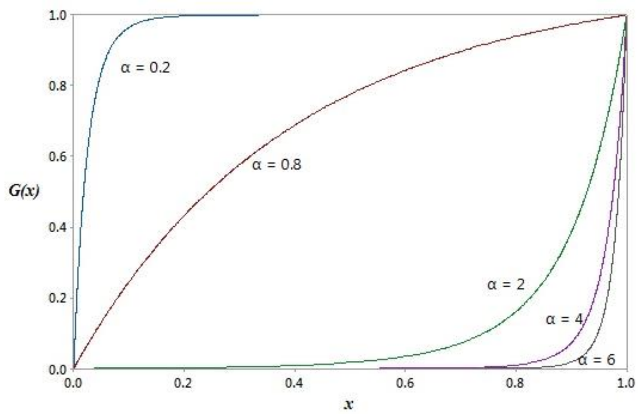

Figura 5. Simulación de la FD de Sankaran y Jayakumar (2016), $\theta = 10$ y diferentes valores de α

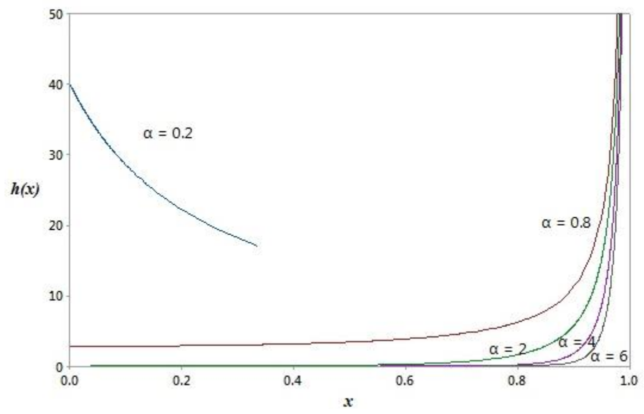

Figura 6. Simulación de la FTR de Sankaran y Jayakumar (2016), $\theta = 10$ y diferentes valores de α

Discusión

A lo largo de la historia, las distribuciones comúnmente utilizadas para modelar un conjunto de datos que pertenecen a un proceso o sistema han sido la distribución normal (variables continuas) y de Poisson (variables discretas). Sin embargo, con la cantidad de datos que se están generando actualmente por los avances de la tecnología, estas distribuciones a menudo se consideran inapropiadas o proporcionan ajustes inadecuados al conjunto de datos analizados. De tal forma que, la generalización de lagunas distribuciones de probabilidad existentes se han ampliado agregando desde uno hasta tres parámetros para obtener una nueva familia de distribuciones para modelar o ajustar un conjunto de datos.

Además, la selección de una distribución apropiada es el primer paso esencial para realizar un análisis de datos. En este sentido se analizaron los datos simulados de la MOUE y GUD para los diferentes valores de α y θ para identificar a que otra función de distribución se ajustan los datos, para este caso, se utilizó el software ExperFit para identificar a que otras distribuciones se ajustan a los datos simulados. La Tabla 1 muestra otros modelos de distribución a los que se pueden ajustar las diferentes simulaciones de la distribución MOEU.

Tabla 1. Ajuste de los datos de la MOEU con ExpertFit

MOEU	**Mejor ajuste con ExpertFit**		
Simulación	**Modelo**	**Puntaje relativo (*p*)**	**Parámetros**
$\alpha = 0.2$ $\theta = 10$	Gamma(E)	96.88	U = 4.131 e -4 E = 3.02481 F = 0.81927
$\alpha = 0.6$ $\theta = 10$	Beta	100	LE = 0.00115 UE = 10.4346 F1 = 0.83568 F2 = 1.32676
$\alpha = 1$ $\theta = 10$	Beta	98.44	LE = 0.00000 UE = 10.0342 F1 = 0.90697 F2 = 0.96795
$\alpha = 1.5$ $\theta = 10$	Beta	100	LE = 0.00000 UE = 10.0112 F1 = 1.01467 F2 = 0.84736
$\alpha = 3$ $\theta = 10$	Johnson SB	98.68	LE = 0.00000 UE = 10.0403 F1 = -0.51560 F2 = 0.54300

U = Ubicación, E = Escala, F = Forma,
LE = Lower endpoint, UP = Upper endpoin

La Tabla 2 presenta otros posibles modelos de distribución a los que se pueden ajustar las diferentes simulaciones de la distribución GUD.

Tabla 2. Ajuste de los datos de la GUD con ExpertFit

MOEU	**Mejor ajuste con ExpertFit**		
Simulación	**Modelo**	**Puntaje relativo (*p*)**	**Parámetros**
$\alpha = 0.2$ $\theta = 10$	Pearson Type VI	100	U = 0.00000 E = 0.18353 F1 = 0.91476 F2 = 7.02999
$\alpha = 0.8$ $\theta = 10$	Beta	100	LE = 0.00000 UE = 1.40331 F1 = 0.86409 F2 = 3.15556
$\alpha = 2$ $\theta = 10$	Johnson SB	100	LE = 1.90 e-4 UE = 1.00722 F1 = -1.98210 F2 = 0.82172
$\alpha = 4$ $\theta = 10$	Johnson SB	100	LE = 0.53910 UE = 1.00287 F1 = -2.40741 F2 = 0.87641
$\alpha = 6$ $\theta = 10$	Johnson SB	100	LE = 0.72284 UE = 1.00178 F1 = -2.41705 F2 = 0.88095

U = Ubicación, E = Escala, F = Forma,
LE = Lower endpoint, UP = Upper endpoin

Como se puede observar en las tablas 1 y 2, las diferentes simulaciones se pueden ajustar a otros modelos de distribución ya existentes.

Conclusión

La distribución uniforme es una de las distribuciones de probabilidad más simples, pero es la que más se utiliza para generar otros modelos de distribución utilizando el método de la transformada inversa. Sin embargo, tiene pocas aplicaciones en la vida real. Este problema ha hecho que autores como Jose y Krishna y Sankaran y Jayakumar hayan desarrollado otras distribuciones basadas en la distribución uniforme agregando más parámetros con el objetivo de contar con nuevas distribuciones más flexibles que permitan modelar el comportamiento de un proceso o sistema. De tal forma que, en este trabajo se simularon los modelos propuestos por los autores mencionados y se realizó un ajuste de datos con el software ExpertFit, donde encontramos que las diferentes simulaciones se pueden ajustar a otras distribuciones como la Beta, Gamma, Johnson SB y Pearson Type VI. 5, 3

Referencias

[1] H. M. Okasha and M. Kayid, "A new family of Marshall-Olkin extended generalized linear exponential distribution," *J. Comput. Appl. Math.*, vol. 296, pp. 576–592, 2016.

[2] A. M. Alshangiti, M. Kayid, and B. Alarfaj, "A new family of Marshall-Olkin extended distributions," *J. Comput. Appl. Math.*, vol. 271, pp. 369–379, 2014.

[3] K. K. Sankaran and K. Jayakumar, "A New Extended Uniform Distribution," *Int. J. Stat. Distrib. Appl.*, vol. 2, no. 3, pp. 35–41, 2016.

[4] S. H. Abid and H. A. Hassan, "The Beta Marshall-Olkin Extended Uniform Distribution," *J. Saf. Eng.*, vol. 4, no. 1, pp. 1–7, 2015.

[5] K. K. Jose and E. Krishna, "Marshall-Olkin extended uniform distribution," *ProbStat Forum*, vol. 4, no. October, pp. 78–88, 2011.

[6] A. W. Marshall and I. Olkin, "A new method for adding a parameter to a family of distributions with application to the exponential and Weibull families," *Biometrika*, vol. 84, no. 3, pp. 641–652, 1997.