

# VARIABLES ALEATORIAS EN LA GESTIÓN EMPRESARIAL, SIMULACIÓN

María Hermenegildo Ch. (\*)

E-mail: [marivic17@yahoo.com](mailto:marivic17@yahoo.com)

## RESUMEN

En este artículo se presentan técnicas cuantitativas para simular un generador de una variable aleatoria no-uniforme, cuyos valores se reparten siguiendo una distribución de probabilidad conocida o no conocida, y sólo cuando esto no sea posible por uno de los métodos presentados, se trabajará la simulación empíricamente. También se presentan aplicaciones en la empresa de distintas distribuciones de probabilidad y los perfiles para valores específicos de sus parámetros, a fin de que el lector pueda identificar, antes de aplicar algún método, qué curva podría representar a la variable aleatoria.

Como ejemplo se muestra mediante un gráfico y con resultados estadísticos que los datos generados siguen realmente la distribución que representa a la muestra original. De esta forma, el lector puede usar esta información en un sistema dinámico de simulación de procesos con eventos discretos, que representa a un sistema real.

**Palabras Clave:** Distribución de Probabilidad, Generador, Variable Aleatoria, P-Valor, Prueba de Bondad de Ajuste.

## ABSTRACT

In this article I present quantitative techniques to simulate a non-Uniform random-variables generator whose values distribute following a known (or unknown) probability distribution. And only when this isn't possible with one of the presented methods, I will work the simulation empirically. Also, I present applications in the business of different probability distributions and its profiles for specific values of its parameters with the objective that the reader can identify the curve could represent to the random variable before applying a method.

Like a example I show a graphic with statistical results that the generated dates distribute really as the dates of the original sample. In this way, the reader can use this information in a dynamic system of simulation processes with discrete events which represent a real system.

**Key Words:** Probability Distribution, Generator, Random Variable, P-Value, Test of Goodness of Fit.

\* Magíster en Ciencias de la Computación por la Universidad de Cantabria (España), Estudios de Maestría en Ciencias (UNI) y Licenciada en Matemáticas (UNT). Docente Auxiliar de la Facultad de Ciencias Administrativas de la UNMSM, Docente de la Unidad de Postgrado FCA (UNMSM), Docente del Postgrado de la Facultad de Ingeniería Industrial y de Sistemas (UNI).

## INTRODUCCIÓN

En los procesos de gestión o procesos industriales se produce la llegada de entidades como clientes, piezas, documentos, llamadas. Las entidades arriban al azar en intervalos de tiempo y en un número finito. Cuando ocurre el proceso de llegada, las entidades demandan un tiempo de servicio para ser atendidas en diferentes locaciones (ventanillas, máquinas, oficinas) y tardarán un tiempo para trasladarse de una a otra locación; en muchas situaciones deberán formar una cola la cual tiene una disciplina FIFO (*first in, first out*) o LIFO (*last in, first out*) o una cola para aquellas entidades que tengan prioridades, por ejemplo un cliente que es mayor de 60 años es atendido en una ventanilla especial, o se le da prioridad a un documento muy importante que se debe tramitar con urgencia. También puede ocurrir que un trámite no se puede realizar en una determinada locación por lo que es rechazado pero igual transcurrió un tiempo entre hacer la consulta y la respuesta, o un producto que no pasa control de calidad, o un documento que está mal elaborado y que debe rehacerse; cada una de esas situaciones reales se dan al azar. Los tiempos de servicio, tiempo para hacer la consulta antes de dar la respuesta, número de arribos de entidades, tiempos entre arribos, número de piezas defectuosas, son ejemplos de variables aleatorias.

En los modelos de inventario de optimización se asume que la demanda de un producto es conocida, cuando en realidad no lo es. Generalmente estas variables siguen distribuciones teóricas o empíricas distintas a la distribución uniforme. Para simular este tipo de variables con mayor precisión se requiere trabajar con un generador de números aleatorios no-uniformes y una función de densidad hallada por un método específico. Para lograr este objetivo se usa la Prueba de Bondad de Ajuste Chi-cuadrada o Kolmogorov-Smirnov, y se elige la distribución de probabilidad que represente a la variable aleatoria. Si no es posible determinar esa distribución, se estima previamente un polinomio que represente a la muestra mediante una regresión no lineal, y se aplica el método de la transformada inversa o de aceptación y rechazo, que se utiliza cuando es imposible o muy difícil calcular la función inversa de la función de densidad. Solo en el caso que no se logre hallar la función de densidad se recomienda usar el método

empírico; es decir, se generan variables aleatorias a partir de un histograma de frecuencias relativas acumuladas.

El método de composición se apoya en el método de la transformada inversa y se usa cuando la función de densidad está definida por partes.

Debido a que es escaso encontrar textos en las librerías o páginas de internet donde se representen y muestren estas situaciones reales en forma práctica, me he sentido muy motivada para presentar este artículo, esperando que sea de gran ayuda a los analistas de simulación de procesos con eventos discretos y a los lectores en general.

## MÉTODOS

### 1. Distribuciones de probabilidad y sus aplicaciones

Se denotará a una variable aleatoria con  $x$  que puede ser discreta o continua. Las siguientes distribuciones: Binomial, Binomial Negativa, Uniforme Discreta, Geométrica, Poisson, Hipergeométrica, etc. representan a variables aleatorias discretas; y las distribuciones: Beta, Chi-Cuadrada, Erlang, Exponencial, Valor Extremo, Gamma, Logistic, Loglogistic, Lognormal, Normal, Pareto, Pearson 5, Pearson 6, Triangular, Uniforme, Weibull y otras más, representan a variables aleatorias continuas.

A continuación se presentan aplicaciones de distintas distribuciones de probabilidad y sus perfiles para valores específicos de sus parámetros.

La distribución Binomial, donde  $x$  representa el número de aciertos en  $n$  ensayos independientes, con número de ensayos  $n$  y probabilidad constante de ocurrencias del evento  $p$ , ha sido usada en juegos del azar, pero también es muy usada en investigación de mercado; por ejemplo, para representar muestras de un producto doméstico que posee cada familia. Asimismo, se usa para representar muestras de partes defectuosas de un determinado producto, y otras muestras donde la probabilidad del evento es conocida. Para  $n$  grande, la distribución Binomial puede ser aproximada por la distribución Normal y, si  $n$  es grande y la probabilidad  $p$  es chica, puede ser aproximada por la distribución de Poisson. Observe distintos perfiles de la distribución Binomial (ver Gráficos N° 1 al N° 4).

Gráfico N° 1

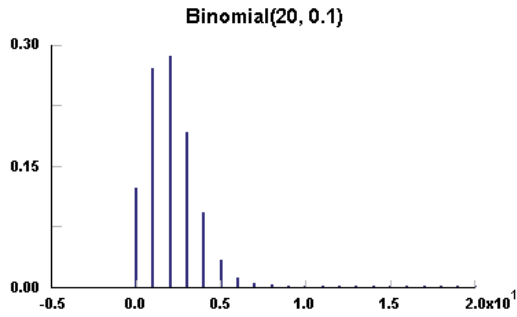


Gráfico N° 2

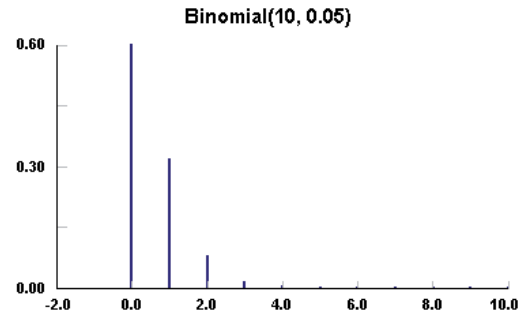


Gráfico N° 3

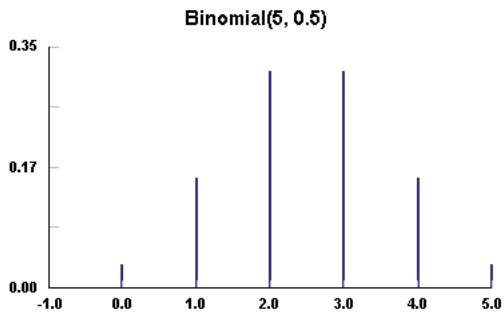
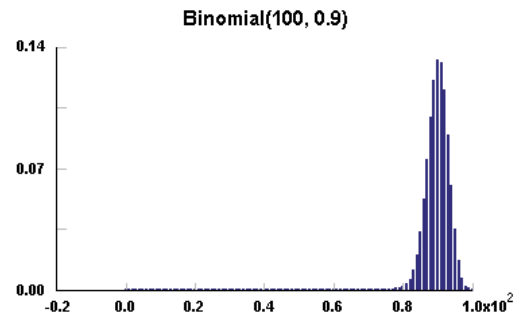


Gráfico N° 4



La distribución de Poisson, tiene muchas aplicaciones importantes que no tienen relación directa con la distribución Binomial. Esta distribución con parámetro  $\lambda$  (que sustituye a  $n.p$ ) el cual se interpreta como el número de ocurrencias, aciertos esperados, o promedio, generalmente se usa para representar ocurrencias no frecuentes de eventos cuya razón es constante. En la teoría de

colas es muy usada para representar, por ejemplo, el número de llegadas de clientes a un banco y arribos de llamadas telefónicas. También es usada para modelar defectos en manufactura de semiconductores, defectos en muchos aspectos de control de calidad, etc. Observe los Gráficos N° 5 al N° 8.

Gráfico N° 5

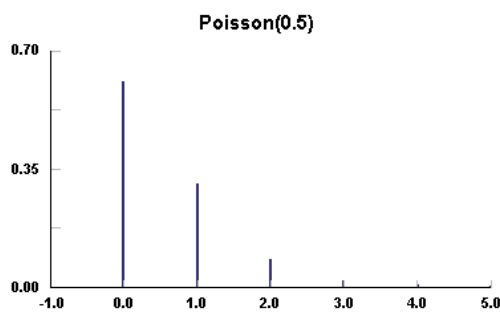


Gráfico N° 6

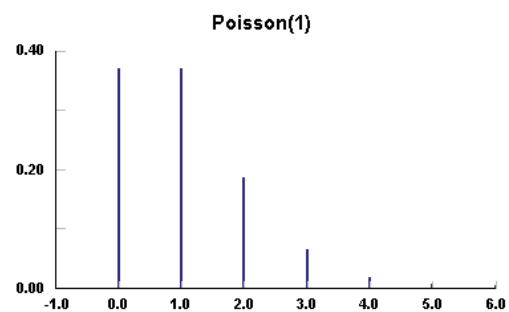


Gráfico N° 7

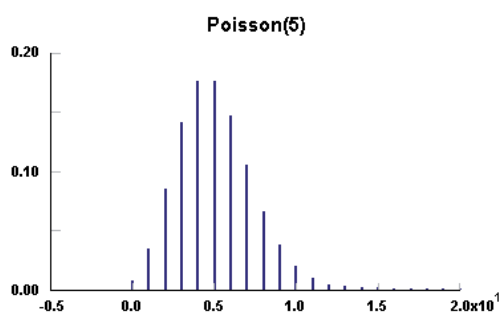


Gráfico N° 8

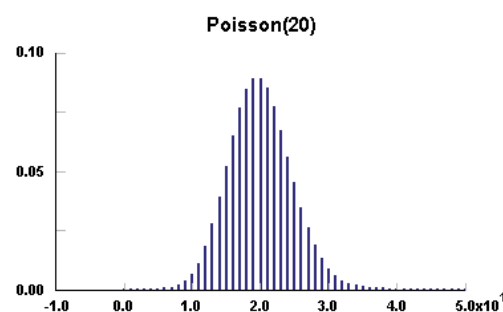
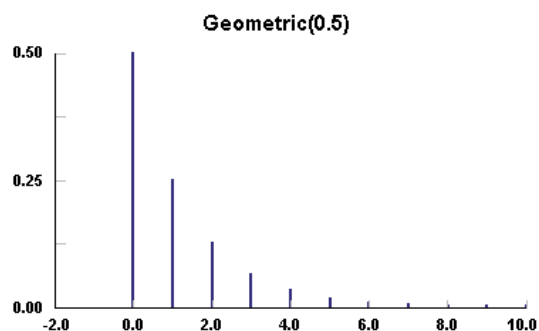


Gráfico N° 9

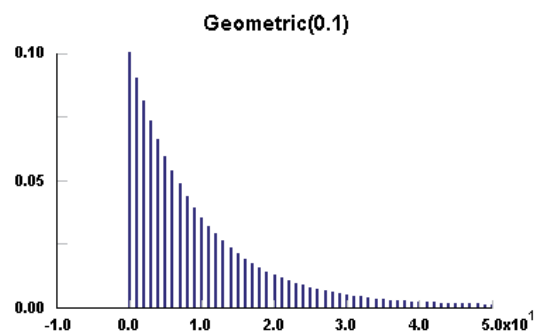


La distribución Geométrica, con parámetro probabilidad de ocurrencia  $p$  se usa para modelar la demanda de modelos de inventario, problemas de control de tickets, y modelos meteorológicos. Se utiliza en la distribución de los tiempos de espera, de manera que si los ensayos se realizan a intervalos regulares de tiempo, esta variable aleatoria proporciona el tiempo transcurrido hasta el primer éxito. Por ejemplo, encontrar la primera pieza defectuosa, la primera ocurrencia de un suceso, la llegada de un cliente a un lugar de servicio, la rotura de una cierta pieza, etc. (ver Gráficos N° 9 y N° 10).

La distribución discreta Uniforme con parámetros mínimo  $a$  y máximo  $b$ , algunas veces llamada distribución discreta rectangular, representa a eventos que puede tener un número finito y de igual probabilidad de llegada (ver Gráfico N° 11).

La distribución Uniforme continua acotada con parámetros mínimo  $a$  y máximo  $b$  y tiene muchas aplicaciones. Muchos generadores de números aleatorios proporcionan muestras de la distribución Uniforme sobre  $(0,1)$  y luego se usan estas muestras para generar números aleatorios de otras distribuciones. La distribución Uniforme se usa para representar una variable aleatoria con probabilidad constante (ver Gráfico N° 11).

Gráfico N° 10



Uno de los más importantes ejemplos de una distribución de probabilidad continua es la distribución Normal con parámetros media  $\mu$  y desviación estándar  $\sigma$ . Su estudio data sobre la naturaleza de errores experimentales. La distribución Normal es frecuentemente usada para representar datos simétricos, siendo no acotada en ambos extremos. Si el conjunto de datos tiene una cota inferior éstos pueden representarse mejor por la distribución Lognormal, Weibull o Gamma; si el conjunto de datos tiene una cota superior y una cota inferior, la distribución Beta puede ser usada, aunque en muchos casos se usa truncando los extremos.

De la distribución Normal se derivan las distribuciones *Chi-cuadrada*, *t de Student* y *F*, de mucha importancia para la contrastación de hipótesis estadísticas. El Gráfico N° 13 muestra tres distribuciones normales con media  $\mu=0$  y distintas desviaciones estándar, y el Gráfico N° 14 muestra la distribución normal con  $\mu=10$  y  $\sigma=1$ .

La distribución Normal se utiliza para procesos de control de calidad, para representar muestras de desempeño laboral, de calificaciones, de sistemas de pagos de los clientes, para medir tiempos de servicio, para medir el grado de satisfacción de los clientes de un determinado producto, para

Gráfico N° 11

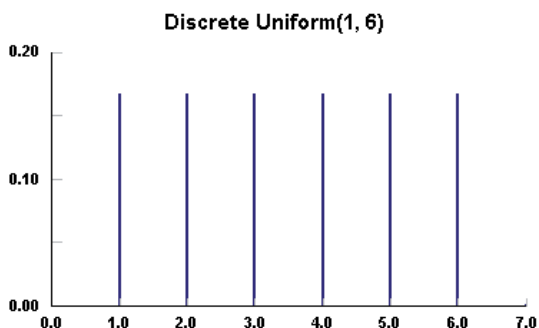
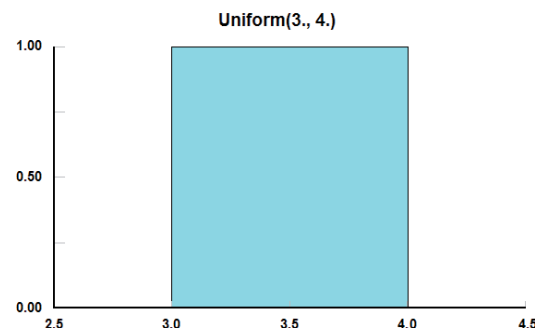


Gráfico N° 12



representar tallas, pesos, etc. Muchos sistemas comerciales de control de inventario encuentran que la demanda durante el período de abastecimiento está muy bien representada por la distribución Normal.

La distribución Exponencial con parámetros mínimo  $a$  y de escala  $\beta$ , es una distribución continua acotada inferiormente. El tiempo entre las llegadas en las instalaciones de servicio y el tiempo de falla de los componentes y sistemas eléctricos de modo frecuente se distribuye exponencialmente. La distribución Exponencial es usada para representar tiempos entre ocurrencias aleatorias que siguen un proceso Poisson, tal como los tiempos entre arribos a una locación específica en un modelo de colas o el tiempo que transcurre entre fallas, considerando que las fallas no dependen del tiempo transcurrido. También es usada para representar los tiempos de servicios de una operación específica cuando la duración es muy breve (con frecuencia no son representadas por esta distribución). Las distribuciones Exponencial y Gamma juegan un papel importante en teoría de colas, en problemas de confiabilidad y supervivencia. Observe el Gráfico N° 15.

La distribución Logistic es una distribución continua acotada con parámetros *media* (o de forma)  $\alpha$  y de escala  $\beta > 0$ , y es de naturaleza simétrica sobre su media. Esta distribución es muy parecida a la Normal con la diferencia que tiene colas más “gruesas” y por lo tanto mayor curtosis. Se usa en marketing, para estudiar la difusión de nuevos productos, para describir cómo las tecnologías se posicionan y compiten entre sí. En psicología se usa para describir procesos de aprendizaje. Es muy usada también para representar modelos de regresión logística. Observe el perfil de la distribución logística en el Gráfico N° 16.

La distribución Gamma, con parámetros mínimo, de forma  $\alpha$  y de escala  $\beta$ , es una distribución continua acotada inferiormente. La distribución Gamma se usa para representar problemas de confiabilidad, mantenimiento y tiempos de espera; tiempos de vida de elementos físicos, sistemas eléctricos y mecánicos. Su relación con la distribución Exponencial permite que la distribución Gamma se utilice en tipos similares de problemas. Observe distintos perfiles en los Gráficos N° 17 a N° 20.

Gráfico N° 13

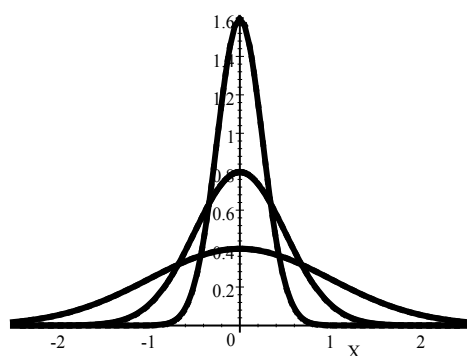


Gráfico N° 14

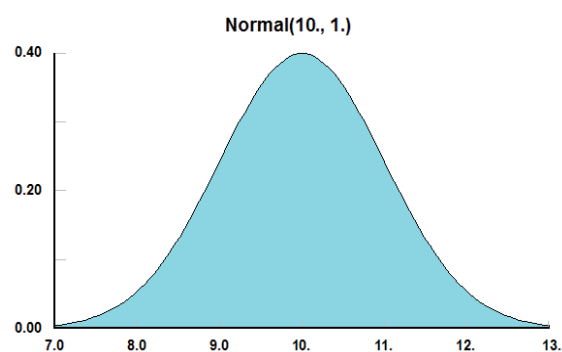


Gráfico N° 15

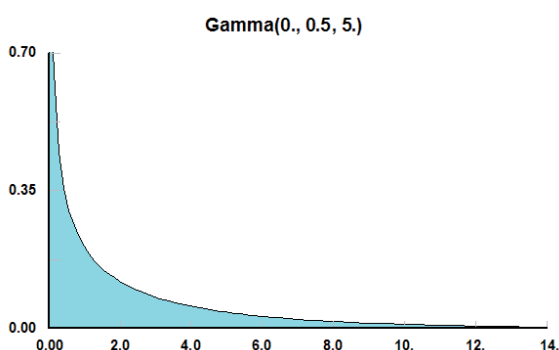


Gráfico N° 16

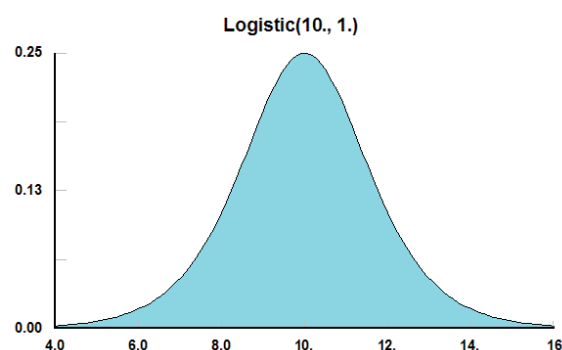


Gráfico N° 17

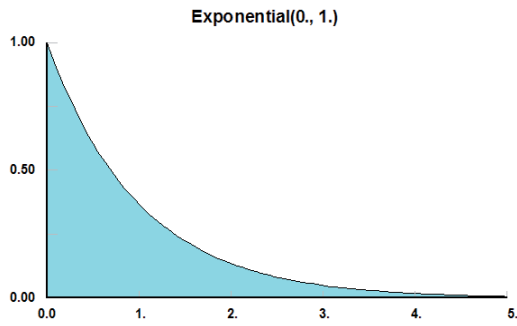


Gráfico N° 18

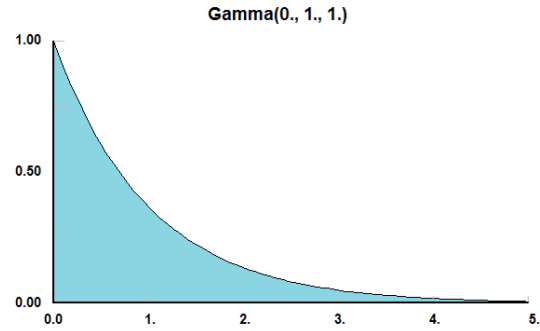


Gráfico N° 19

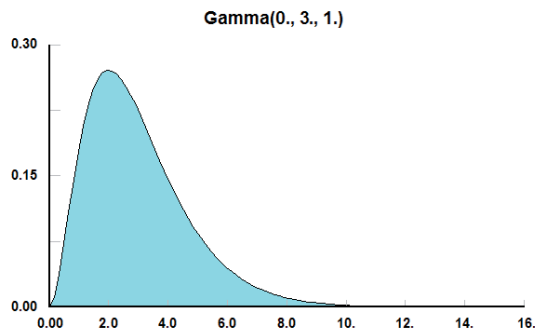
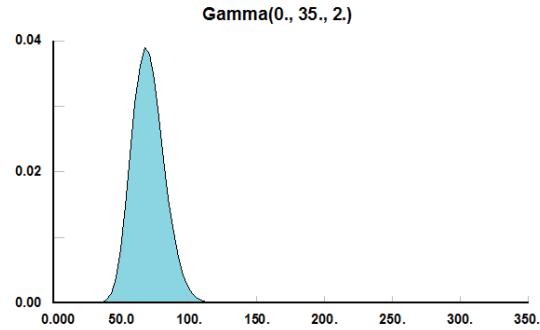


Gráfico N° 20



La distribución Beta, con parámetros *mínimo a*, *máximo b*, *de forma p* y *de escala q*, es una distribución continua con cota finita inferior y superior. La distribución Uniforme es un caso especial de la distribución Beta con  $p, q = 1$ . Muchas situaciones reales pueden ser acotadas de esta forma; la

distribución Beta puede ser usada empíricamente para estimar la distribución de muchos datos disponibles lo cual no indica que sea el mejor ajuste.

La distribución Beta se aplica para estimar el valor esperado y la desviación estándar en los

Gráfico N° 21

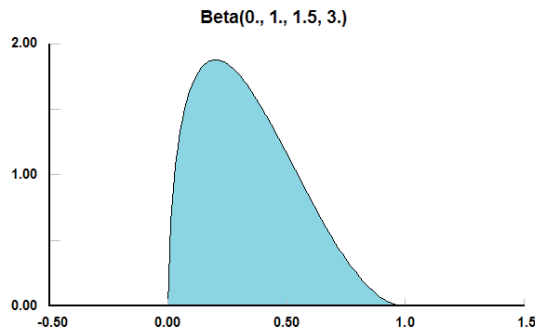


Gráfico N° 22

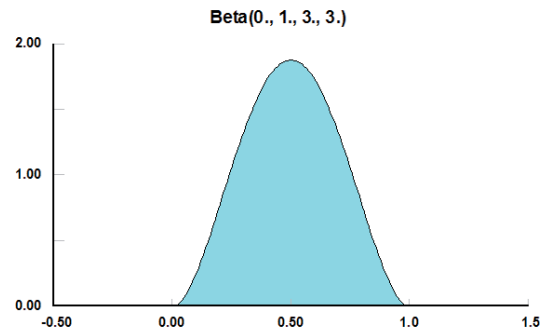


Gráfico N° 23

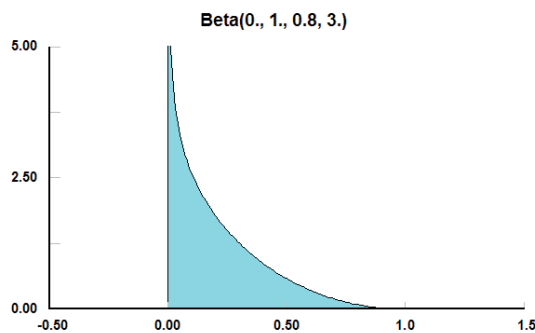


Gráfico N° 24

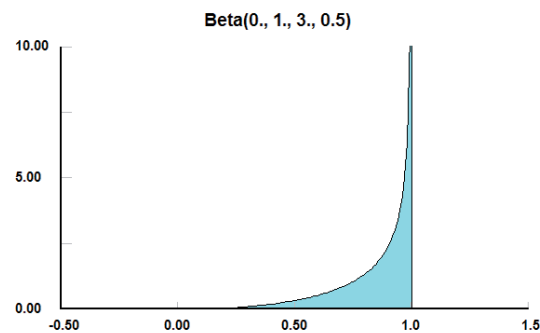


Gráfico N° 25

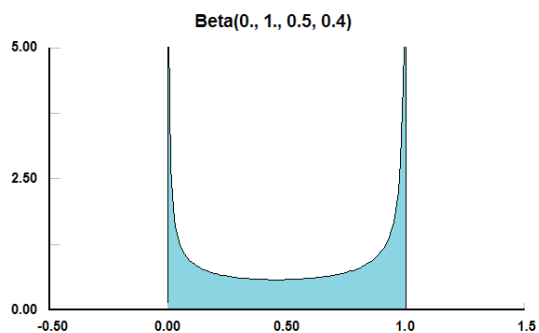
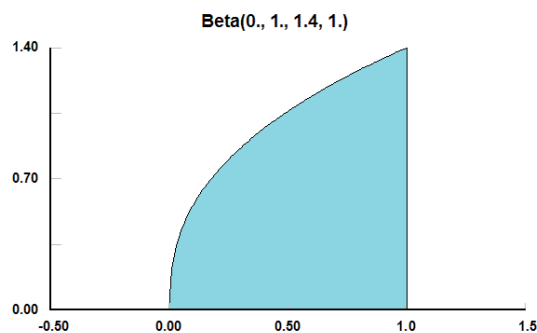


Gráfico N° 26



tiempos de actividades en análisis del PERT, sistemas de seguros individuales y teoría de riesgos colectivos, tiempo de vida de construcciones y otros. Observe los Gráficos del N° 21 y N° 26.

La distribución Erlang es una distribución continua acotada inferiormente con parámetros mínimo  $a$ , factor de forma  $m$  (entero positivo), y factor de escala  $\beta$ . Es un caso especial de la distribución Gamma donde el parámetro  $m$  es restringido a un entero positivo. Es muy usada en confiabilidad y en teoría de colas si los tiempos entre llegadas (por ejemplo de llamadas telefónicas) o de servicio no son exponenciales. La distribución Erlang es vista como la suma de  $m$  variables aleatorias dis-

tribuidas exponencialmente, cada una con media  $\beta$ . Observe los Gráficos N° 27 y N° 28.

La distribución Weibull es una distribución continua acotada inferiormente con parámetros: mínimo  $a$ , de forma  $\alpha$ , y de escala  $\beta$ . Es parecida a la distribución Gamma. La distribución Weibull es popular por su uso para modelos de resistencia de materiales, y para modelar cualquier situación. En particular, la distribución Weibull es usada para representar y: modelos de confiabilidad y análisis de supervivencia, y sistemas migratorios. Veamos ejemplos de distintos perfiles de la distribución Weibull en los Gráficos N° 29 al N° 32.

Gráfico N° 27

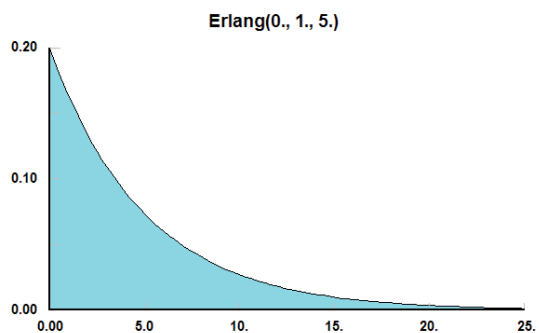


Gráfico N° 28

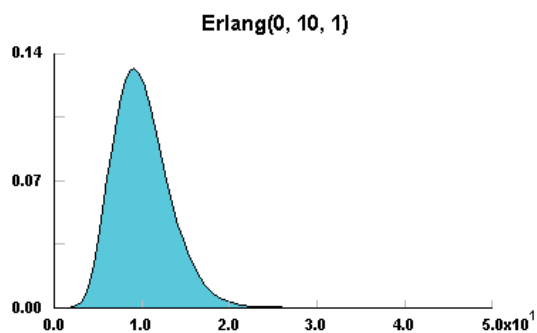


Gráfico N° 29

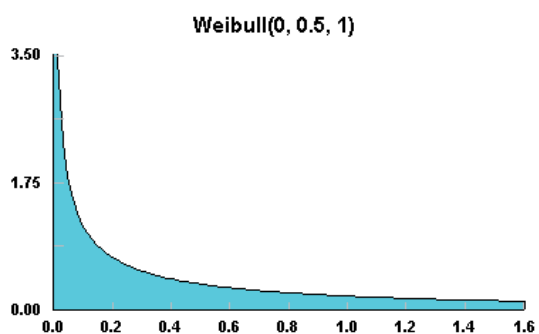


Gráfico N° 30

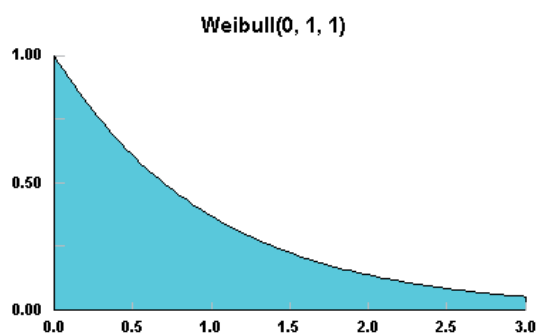
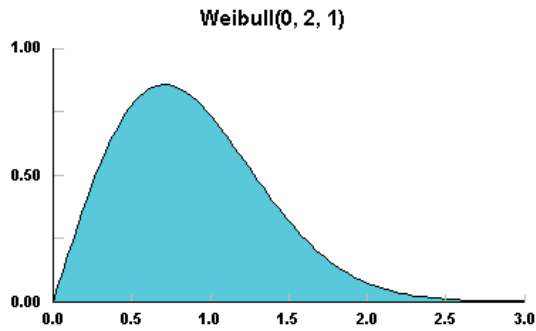


Gráfico N° 31



La distribución Lognormal es una distribución continua acotada inferiormente, con parámetros mínimo  $a$ , media  $\mu$  y desviación estándar  $\sigma$  de los logaritmos de los datos (que siguen una distribución Normal). Se usa en diferentes áreas, tiempos de vida en confiabilidad, y también para modelar peso, altura, etc. La distribución Lognormal se ajusta a ciertos tipos de fallos (desgaste de componentes metálicos), tiempo de vida de los aislamientos eléctricos y puede ser una buena representación de la distribución de los tiempos de reparación. Es también una distribución importante en la valoración de sistemas con reparación. Observe sus perfiles en los Gráficos N°33 y N°34.

## 2. Generación de Variables Aleatorias

En un modelo de simulación probabilístico existen variables aleatorias interactuando. Generalmente estas variables siguen distribuciones teóricas o empíricas distintas a la distribución uniforme. Para simular este tipo de variables es necesario trabajar con un generador de números aleatorios uniformes y una función hallada por un método específico, que permita obtener valores de la distribución de probabilidad deseada.

Gráfico N° 33

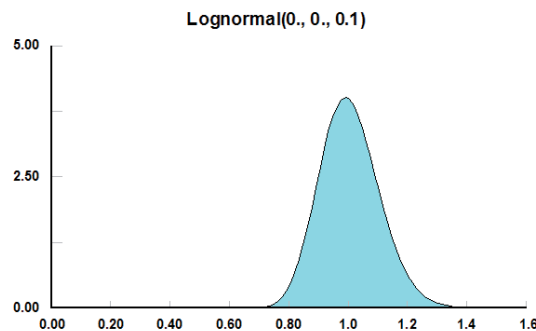
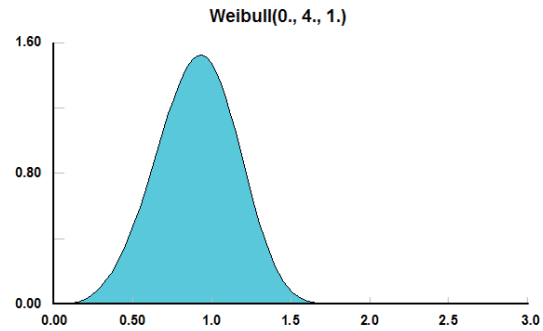


Gráfico N° 32



El muestreo a partir de distribuciones de probabilidad, o generación de variables aleatorias es un procedimiento que se conoce como simulación de Monte Carlo.

Para lograr nuestro objetivo se presentan los siguientes métodos:

### a) Método de la transformada inversa

Sea la función de densidad  $f(x)$  donde  $x \in [a, b] \subset \mathbb{R}$ , sea su función acumulada  $F(x)$ .

Pasos a seguir:

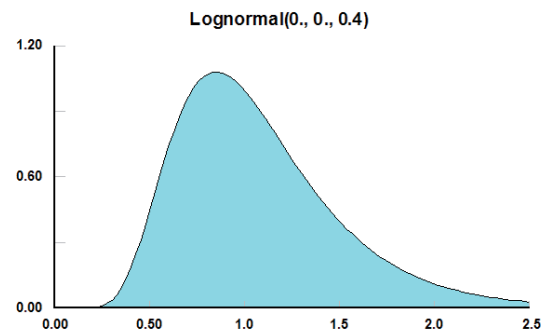
- Generar un número  $U \sim U(0,1)$
- Hacer  $U = F(x)$
- Resolver para  $x$ , es decir  $x = F^{-1}(U)$
- Se libera el valor de  $x$ .

Desventaja.

No siempre se puede encontrar la transformada inversa  $F^{-1}$ .

Con este método se determina un generador de variables aleatorias distribuidas exponencialmente, distribuidas uniformemente, distribuidas normalmente, y de otras distribuciones más, considerando sus parámetros.

Gráfico N° 34





### b) Método de aceptación y rechazo

Con frecuencia es posible calcular la función de densidad  $f(x)$  pero difícil calcular  $F(x)$  o la inversa  $F^{-1}(x)$ . La técnica llamada método de aceptación y rechazo ha sido desarrollada para tratar estos casos.

Sea  $f(x)$  acotada por  $M$ , con rango finito (o soporte)  $a \leq x \leq b$ .

Pasos a seguir:

- Generar dos números  $U_1 \sim U(0,1)$  y  $U_2 \sim U(0,1)$
- Generar una variable aleatoria distribuida uniformemente sobre  $a$  y  $b$ :  $x = a + (b - a)U_1$
- Hallar  $f(x)$
- Si  $M \cdot U_2 \leq f(x)$ , se acepta el valor de  $x$ , y luego se libera; de lo contrario, se rechaza y se regresa al paso 1.

Desventaja:

Mientras más alta sea la moda  $M$ , la probabilidad de que los números aleatorios generados sean aceptados es menor.

Se ha demostrado que en cualquier iteración del método de aceptación y rechazo, la probabilidad de que se genere un valor de la variable aleatoria es de  $\frac{1}{M(b-a)}$ .

### c) Método de composición

Sea  $x \in [a, b] \subset \mathbb{R}$ . Suponga que la distribución de densidad  $f(x)$  es representada como la combinación de dos distribuciones:  $f_1(x)$  para  $x \in [a, c]$  y  $f_2(x)$  para  $x \in [c, b]$ , luego:

$$f(x) = p \cdot f_1(x) + (1-p) \cdot f_2(x), \text{ donde } p = F_1(c)$$

Donde  $F_1(x)$  y  $F_2(x)$  son las distribuciones acumuladas de  $f_1(x)$  y  $f_2(x)$ , respectivamente.

Podemos generar valores de  $x$  que siguen la distribución  $f(x)$  considerando los pasos siguientes:

- Generar un número  $U \sim U(0,1)$ .
- Luego, si  $U < p$ , se hace  $F_1(x) = U$  y se despeja  $x$ ; de lo contrario seguir el paso 3.
- Si  $U \geq p$ , se hace:  $U = p + F_2(x)$ , y despejamos  $x$ .
- Se libera el valor de  $x$ .

Desventaja:

Como se basa en el método de la transformada inversa, no siempre se puede encontrar la

inversa de la distribución acumulada de cada distribución que lo compone.

El método de composición se generaliza cuando  $f(x)$  está formada por  $n$  distribuciones.

### 3. Estimación de Parámetros

Cuando los datos se distribuyen siguiendo una distribución de probabilidad conocida se procede a la identificación de la curva y a la estimación de sus parámetros correspondientes aplicando las Pruebas de Bondad de Ajuste: Chi-cuadrada para distribuciones discretas, y de Kolmogorov-Smirnov para distribuciones continuas, bajo la hipótesis nula  $H_0$  que afirma que "un conjunto de datos empíricos o de muestra no difieren significativamente de aquellos que se esperan de la distribución teórica específica, con un nivel de significancia  $\alpha$ ".

El  $p$ -value varía entre 0 y 1. En realidad mientras más alto sea el  $p$ -value, la distribución específica realizará un mejor ajuste a la muestra en estudio y, mientras más pequeño sea el  $p$ -value, tiende a rechazarse la distribución teórica.

#### CASOS EN ESTUDIO

##### Caso 1:

En un determinado banco un cliente se dispone a realizar un trámite financiero a través de ventanilla. El tiempo transcurrido entre el fin de la consulta del usuario y el inicio de la respuesta del sistema a esa consulta es una variable aleatoria continua. Un día normal empieza a las 9:00 horas y termina a las 17 horas.

##### Caso 2:

En el mismo banco se cuenta con un departamento de Mesa de Ayuda donde se atienden y canalizan los requerimientos de servicios informáticos. Los canales de atención con que cuenta son teléfono y a través de un Sistema de Mesa de Ayuda (Intranet).

Los tiempos entre arribos de las llamadas de requerimientos por parte de los usuarios en Mesa de Ayuda es una variable aleatoria.

El objetivo es determinar un generador de tiempos (variable aleatoria), considerando la distribución de probabilidad que realiza un buen ajuste a la muestra real, usando las técnicas cuantitativas adecuadas.

## RESULTADOS Y DISCUSIÓN

### 1. Caso 1

A continuación se trabaja con el caso 1:

La variable aleatoria es continua. Se ha tomado una muestra de  $n=70$  datos, y se obtiene las siguientes medidas descriptivas (tabla 1) y los histogramas de frecuencias absolutas y frecuencias relativas (ver Gráficos N° 35 y N° 36).

Tabla N°1

Número de observaciones	70
Mínimo	0.0416
Máximo	2.8907
Media	1.0210
Mediana	0.9876
Moda	1.0445
Desviación Estándar	0.6909
Sesgo	0.6520
Curtosis	-0.2693

Se observa que la muestra tiene sesgo positivo (sesgada hacia a la derecha); de esta forma se procede a proponer cuatro distribuciones teóricas: Exponencial, Logistic, Triangular y Pearson 5.

En el Gráfico N° 37 se observa el ajuste de la distribución Exponencial y se puede apreciar que en el valor mínimo de  $x$ , la frecuencia sería muy alta, y lo más probable es que sea rechazada. En la

Gráfico N° 37

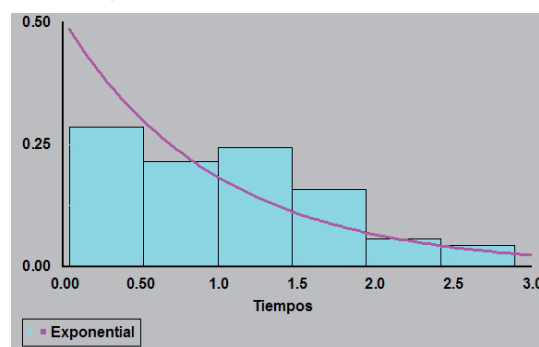


Gráfico N° 38

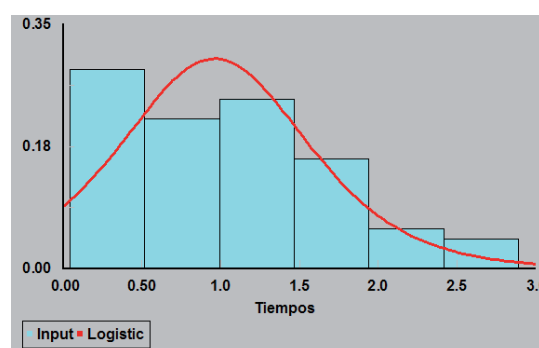


Gráfico N° 35

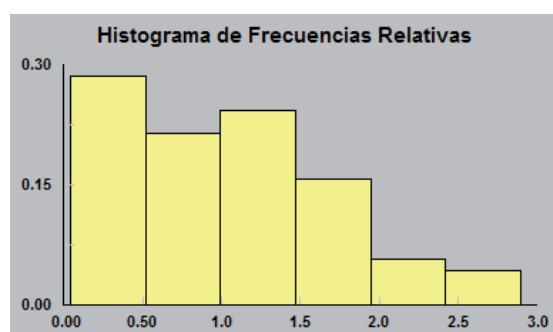


Gráfico N° 39

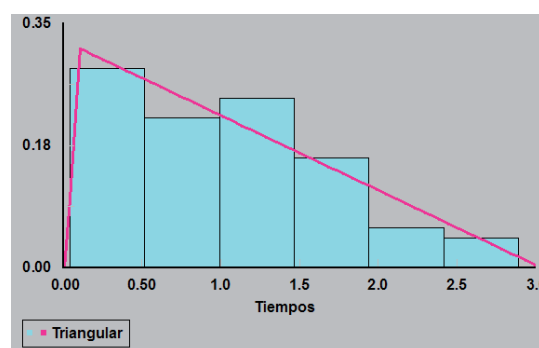


Gráfico N° 36

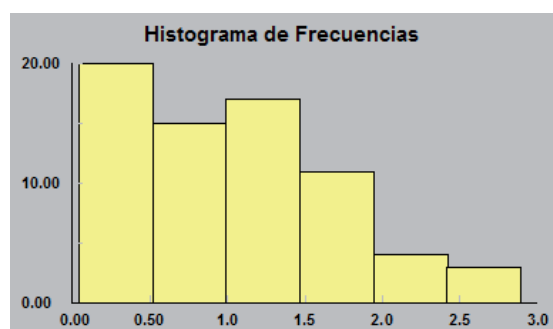
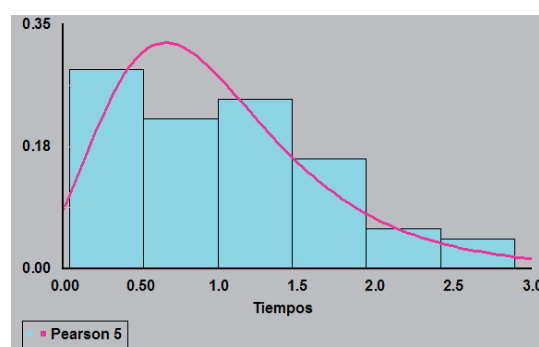


Gráfico N° 40



distribución Triangular (Gráfico N°39) se observa que los datos se distribuyen de esa forma, excepto que en el valor mínimo de  $x$  la función se acerca a 0, por lo que tendrían un pequeño  $p$ -value. Y las distribuciones Logistic y Pearson 5 (ver Gráficos N°38 y N°40), tendrían un  $p$ -value más alto.

Como se trata de una variable aleatoria continua usamos la Prueba de Bondad de Ajuste de Kolmogorov-Smirnov, para lo cual se trabajó con el software STAT::FIT y se obtuvieron los parámetros y  $p$ -value correspondientes (ver Tabla N°2) de las distribuciones de probabilidad, considerando también a las distribuciones Gamma y Weibull:

- Logistic(0.965, 0.395)
- Weibull(0.0231, 1.39, 1.08)
- Gamma(0.0115, 1.65, 0.61)
- Exponencial(0.0416, 0.979)
- Triangular(1.04e-002, 3.04, 0.107)
- Pearson 5(-1.2, 11.1, 22.5)

La distribución Logistic tiene el más alto  $p$ -value por lo tanto es la que mejor ajusta a la muestra. Se presenta el histograma de frecuencias con el ajuste de la distribución Logistic (ver Gráfico N°41).

Se presenta un histograma de la muestra real con  $n=10$  clases en el Gráfico N°42.

Para dos muestras generadas de tiempos que tarda un empleado en responder a la consulta (en minutos): para  $n=70$  datos (ver Gráfico N°43), y para  $n=65$  (ver Gráfico N°44), truncando la distribución Logistic (en ambos casos) para descartar previamente algún valor negativo (recuerde que es una distribución de naturaleza simétrica) se obtienen los siguientes histogramas: (Ver graficos N.º 43, N.º 44)

Aplicaremos ahora el método de aceptación y rechazo:

Vamos a hacer un ajuste de los datos mediante una regresión con un nivel de significancia  $\alpha=0.05$ . Considerando el histograma de frecuen-

Tabla N° 2

	Logistic	Weibull	Gamma	Exponencial	Triangular	Pearson 5
p-value	0.617	0.393	0.217	0.0268	0.556	0.57

Gráfico N°41

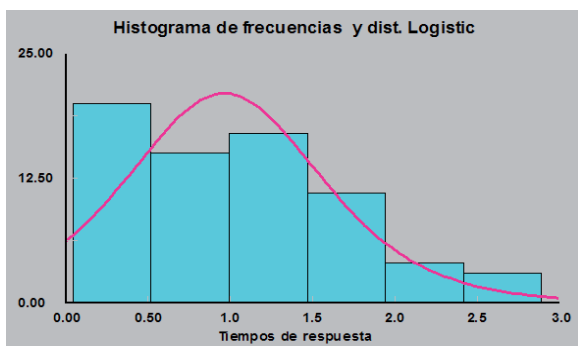


Gráfico N°43

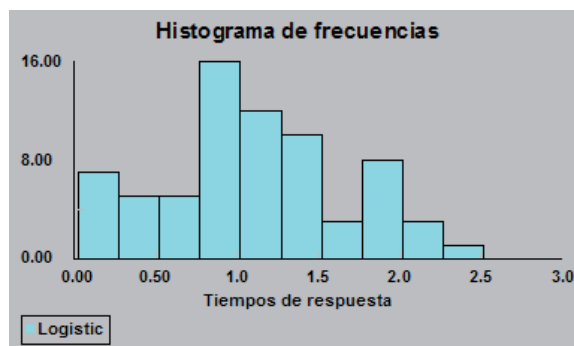


Gráfico N°42

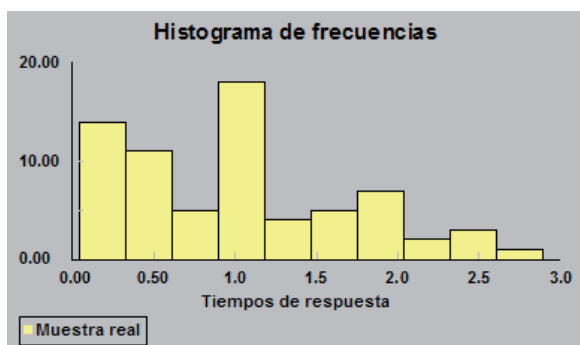
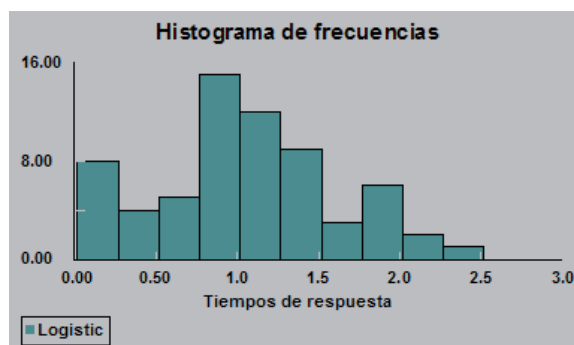


Gráfico N°44



cias con 10 clases (ver Gráfico N°44), siempre con  $n=70$ , se calculan las marcas de clase y las frecuencias con el STAT::FIT (observe la Tabla N° 3), recordando que  $a = \text{mínimo} = 0.0416$  y  $b = \text{máximo} = 2.8907$ .

Se considera como variable independiente a las marcas de clase y como variable dependiente a las frecuencias de clase, y se ajusta con un polinomio de grado 5:

$$f_3(x) = -5.7369x^5 + 43.825x^4 - 121.66x^3 + 147.77x^2 - 78.094x + 24.228$$

Con un coeficiente de determinación:  $R-Sq = 55.4\%$  (ver Gráfico N° 45).

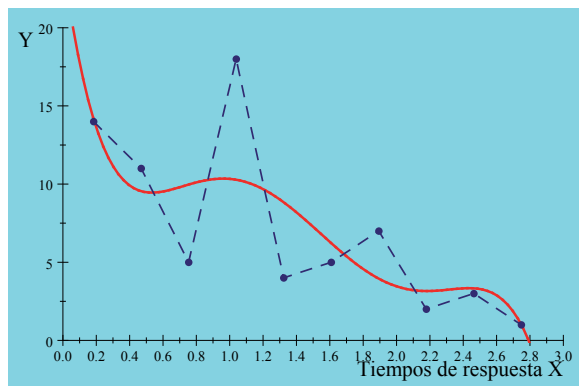
Función de densidad de  $f_5(x)$ :

$$f_5(x) = -0.28468x^5 + 2.1747x^4 - 6.0371x^3 + 7.3328x^2 - 3.8752x + 1.2023$$

Con función de densidad acumulada:

$$F(x) = -0.0474x^6 + 0.4349x^5 - 1.5093x^4 + 2.4443x^3 - 1.9376x^2 + 1.2023x - 0.0469$$

Gráfico N°45



Por el método de la transformada inversa, haciendo  $x=F^{-1}(U)$  donde  $U(0,1)$ .

Observando los tiempos iniciales (marcas de clase), que aún cuando se tiene errores de redondeo, el polinomio  $f_5(x)$  hace un buen ajuste de los datos. Observe los resultados generados en la Tabla N° 3, columna. 4.

En general, se puede obtener un ajuste de los datos con un polinomio de grado  $n$ ; por ejemplo con un polinomio de grado 3 (ver Gráfico N°49), con  $R-Sq = 39.6\%$ , se obtiene:

$$f_3(x) = 2.0828 \times 10^{-6}x^3 + 5.106 \times 10^{-6}x^2 - 4.6800x + 13.862$$

De manera similar, podemos obtener un buen ajuste de los datos con un polinomio de grado 4, con

$$R-Sq = 53.21\%.$$

$$f_4(x) = 1.7689x^4 - 10.374x^3 + 19.871x^2 - 18.349x + 16.27$$

Se presenta la gráfica con el diagrama de dispersión de las marcas de clase con sus respectivas frecuencias, y el ajuste con los polinomios de grado 3, 4 y 5. Observe el Gráfico N° 46.

Gráficamente se observa que el mejor ajuste lo realiza el polinomio de grado 5.

De manera similar podemos obtener un buen ajuste de los datos con un polinomio de grado 7:

$$f_7(x) = 15.588x^7 - 154.92x^6 + 609.6x^5 - 1204.0x^4 + 1246.5x^3 - 641.89x^2 + 135.85x + 4.4296$$

Con  $R-Sq = 65.53\%$ , observe el Gráfico N° 47.

TABLA N° 3

	Marcas de clase (variable independiente x)	Frecuencias absolutas (variable dependiente y)	Frecuencias relativas acumuladas	Tiempos generados manualmente con el polinomio de grado $n=5$	Tiempos generados manualmente con el polinomio de grado $n=7$
	---	0	0	0.0416	0.0416
1	0.1841	14	0.2	0.3105	0.3502
2	0.4690	11	0.3571	0.6341	0.6816
3	0.7539	5	0.4286	0.7806	0.8328
4	1.0388	18	0.6857	1.2975	1.3109
5	1.3237	4	0.7429	1.4331	1.4460
6	1.6086	5	0.8143	1.6415	1.6982
7	1.8935	7	0.9143	2.1218	2.1717
8	2.1784	2	0.9429	2.3036	2.2897
9	2.4633	3	0.9857	2.5687	2.8471
10	2.7482	1	1	2.6774	2.8716

Gráfico N° 46

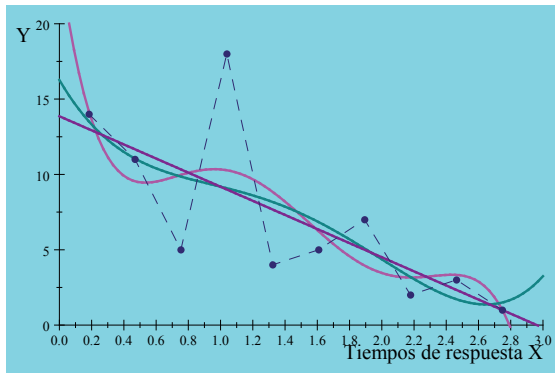
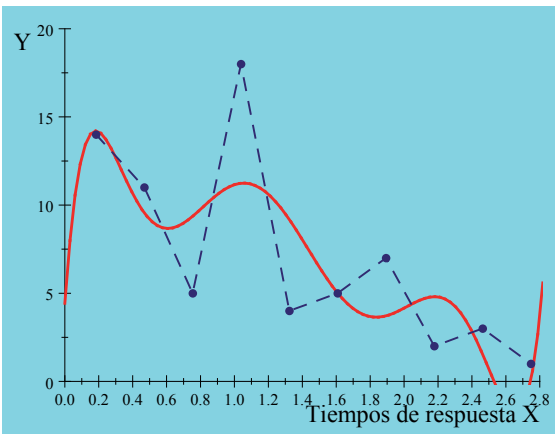


Gráfico N° 47



Con función de densidad:

$$f_d(x) = 0.7797x^7 - 7.7491x^6 + 30.492x^5 - 60.224x^4 + 62.350x^3 - 32.107x^2 + 6.7952x + 0.2216$$

Y función de densidad acumulada:

$$F(x) = 0.0975x^8 - 1.107x^7 + 5.082x^6 - 12.045x^5 + 15.587x^4 - 10.702x^3 + 3.3976x^2 + 0.2216x - 0.0144$$

En este caso el polinomio de grado 7, realiza un mejor ajuste que los polinomios anteriores, para valores positivos de  $f(x)$ .

Los polinomios que mejor ajusta a la muestra se hallaron con el apoyo del *software* Scientific Workplace Vs. 5.0, así como también las gráficas y cálculos manuales, a fin de comprobar que la generación de la variable aleatoria -tiempo que tarda un empleado en dar respuesta a una consulta (en minutos)-, es correcta.

La desventaja del método de la transformada inversa es que se puede calcular la función de densidad  $f_d(x)$  pero es muy difícil calcular  $F(x)$  o encontrar la inversa de  $F$  (como ocurre en este caso), o ésta puede no existir, por lo que se procede a usar el método de aceptación y rechazo.

Simularemos un generador de la variable aleatoria para el polinomio de grado 7, el valor máximo de la función es  $M = 0.7104$  (en 0.1841) con  $a = \text{mínimo} = 0.0416$  y  $b = \text{máximo} = 2.8907$ . Y la probabilidad de que se genere un valor de la variable aleatoria es de  $\frac{1}{M(b-a)} = 0.4941$ , es decir que de los números generados el 49.41% serán aceptados y el resto serán rechazados.

Se ha realizado el algoritmo en código del PROMODEL Vs 7.0 y considerando la función de densidad  $f_d(x)$  se ha corrido la simulación generando tiempos de demora de respuesta (en minutos), por cada vez que llega un cliente a hacer su consulta a la ventanilla de un banco. Se ha extraído una muestra de 70 datos en un *array* para validar el algoritmo, luego se comparan el histograma de frecuencias de la muestra real con el histograma de frecuencias de los valores simulados y se observa que las frecuencias no difieren significativamente, con lo cual se comprueba que el algoritmo del método de aceptación y rechazo está funcionando correctamente. Observe los Gráficos N°48 y N°49.

Gráfico N° 48

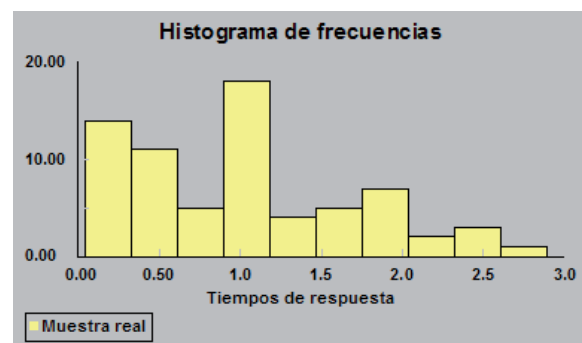
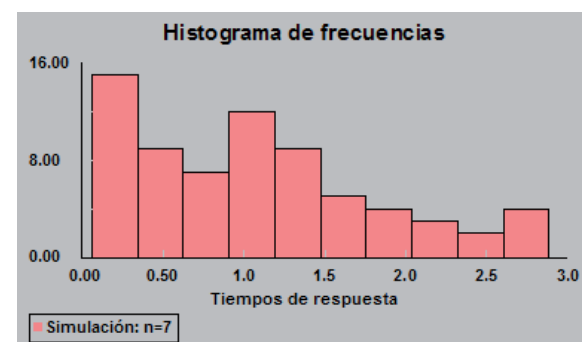


Gráfico N° 49



Finalmente, si no se pueden generar datos porque la función de densidad no se puede determinar, entonces lo hacemos empíricamente, considerando las marcas de clase e intervalos de números aleatorios.

TABLA N°4

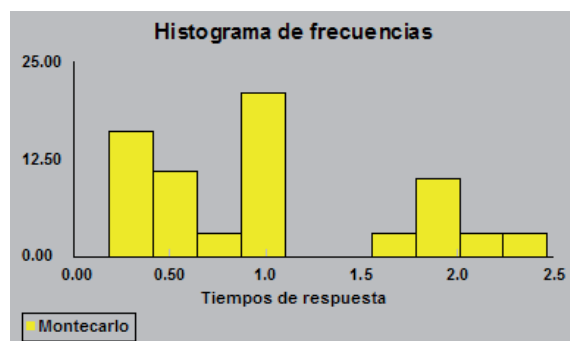
	Marcas de clase	Frecuencias	Frecuencias relativas acumuladas	Intervalos de números aleatorios	Conteo de números aleatorios en cada intervalo
1	0.1841	14	0.2	(0, 0.2]	16
2	0.4690	11	0.3571	(0.2, 0.357]	11
3	0.7539	5	0.4285	(0.357, 0.429]	3
4	1.0388	18	0.6857	(0.429, 0.686]	21
5	1.3237	4	0.7429	(0.686, 0.743]	0
6	1.6086	5	0.8143	(0.743, 0.814]	3
7	1.8935	7	0.9143	(0.814, 0.914]	10
8	2.1784	2	0.9429	(0.914, 0.943]	3
9	2.4633	3	0.9857	(0.943, 0.986]	3
10	2.7482	1	1	(0.986, 1]	0

Usando simulación de Montecarlo, se generan números aleatorios  $U(0,1)$  y se toma una muestra de  $n=70$  valores de la variable aleatoria (que en realidad son las marcas de clase) y se hacen los conteos correspondientes (última columna de la Tabla N° 4) y el histograma de frecuencias (ver Gráfico N° 50).

También se ha corrido la simulación en PRO-MODEL y se ha extraído una muestra de  $n=70$  datos, considerando la siguiente información: extremos de cada intervalo y frecuencias relativas acumuladas porcentuales. Como se puede observar el histograma de valores simulados (ver Gráfico N°51) difieren significativamente del histograma de frecuencias de la muestra real (ver Gráfico N°42); esto ocurre por la aplicación del método empírico.

Se puede observar gráficamente que los datos generados usando la distribución Logistic y el método de aceptación y rechazo son más confiables que los obtenidos por el método empírico.

Gráfico N° 50



En la Tabla N°5 se presentan las medidas descriptivas de la muestra real y de las simulaciones por los distintos métodos

2. Para el caso 2

Se toma una muestra  $n=70$  de los tiempos entre arribos de las llamadas de requerimientos por parte de los usuarios en Mesa de Ayuda, en un día normal, y se obtienen las medidas descriptivas (ver Tabla N°7) y su histograma de frecuencias (ver Gráfico N°52).

Como se trata de una variable aleatoria continua usamos la prueba de bondad de ajuste de Kolmogorov-Smirnov, con un nivel de significancia  $\alpha=0.05$ , y se obtiene que las distribuciones de probabilidad con sus parámetros y  $p$ -value correspondientes (ver Tabla N°6):

- Lognormal(0.025, 1.818, 0.916)
- Weibull(1, 0.900, 7.699)
- Gamma(1, 1, 8.05714)
- Exponencial(1, 8.057)
- Erlang(1, 1, 8.057)

Gráfico N° 51

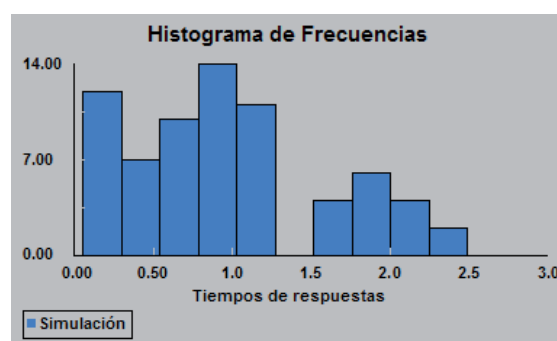


TABLA N° 5

Tamaño de Muestra n = 70	Muestra Real	Simulación Logistic	Simulación de Montecarlo (Manual)	Simulación Empírica (Usando PROMODEL)	Simulación Método de Aceptación y Rechazo (Polinomio de grado 7)
Mínimo	0.0416	0.0203	0.1841	0.0524	0.0595
Máximo	2.8906	2.5150	2.4633	2.4880	2.8833
Media	1.0290	1.0436	0.9981	0.9651	1.0666
Mediana	0.9876	1.0605	1.0388	0.9490	0.9801
Moda	1.0445	0.8390	1.0388	1.0103	0.4057
Desviación estándar	0.6909	0.5630	0.6984	0.6410	0.7560
Sesgo	0.6520	0.1709	0.5226	0.5603	0.7514
Curtosis	-0.2693	-0.2580	-0.8980	-0.5890	-0.1693

TABLA N°6

	Lognormal	Weibull	Gamma	Exponencial	Erlang
p-value	0.620	0.652	0.729	0.729	0.729

Gráfico N°52

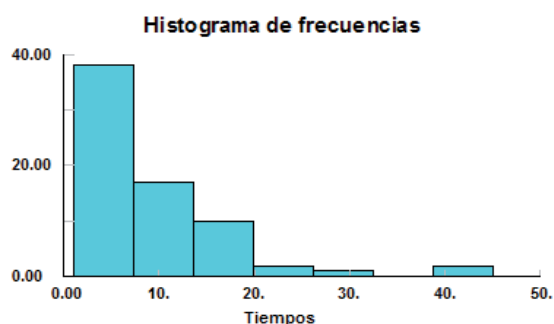


Gráfico N°53

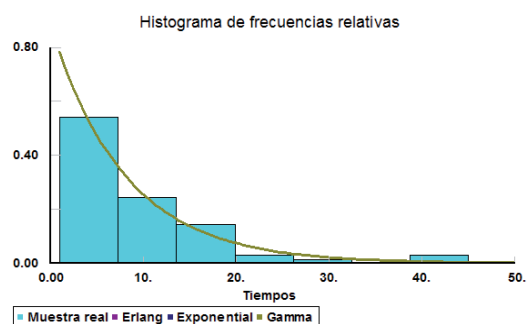


Gráfico N° 54

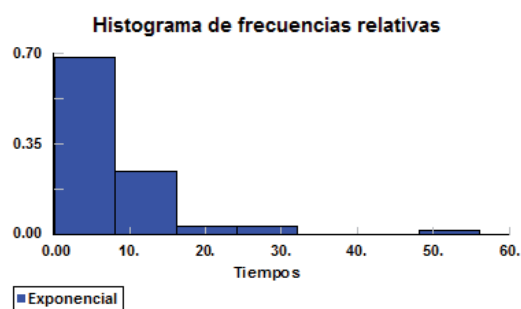
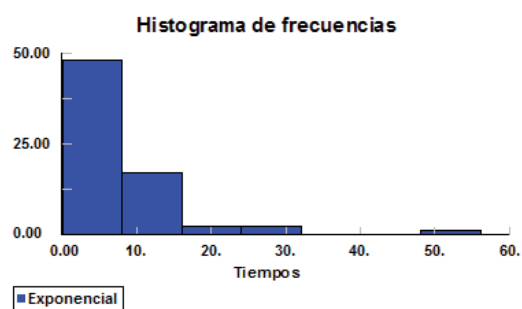


Gráfico N° 55



La distribución Gamma, Erlang y Exponencial tienen el más alto *p-value* (se observa el ajuste que realizan las tres distribuciones en el Gráfico N° 53), eligiendo a la distribución Exponencial como la curva que mejor representa a la muestra real, se pueden generar nuevos tiempos entre arribos.

Se corrió la simulación en PROMODEL versión 7, se usó la orden  $E(8.05714)$  y se extrajo una muestra de  $n=70$  datos en un *array*, obviamente los datos están distribuidos exponencialmente (ver Gráficos N° 54 y N° 55). También se hallaron las medidas descriptivas según se muestra en la Tabla N°7.

TABLA N° 7

Medidas descriptivas	Muestra real	Muestra de Simulación
Mínimo	1.0000	0.0623
Máximo	45.0000	56.0935
Media	9.0571	7.2449
Mediana	7.0000	5.1050
Moda	3.0000	0.9839
Desviación estándar	8.3612	8.4455
Sesgo	2.1949	3.3837
Curtosis	6.0127	15.3812

## CONCLUSIONES

1. La simulación es una técnica poderosa para observar comportamientos en los procesos y cómo éstos cambiarán con la aleatoriedad de las variables y modificación de sus parámetros, además que es barato en relación a la experimentación con el sistema real. Por ello se recomienda encontrar un simulador de las variables aleatorias que interactúan en el sistema con la finalidad de que las medidas de eficacia por ejemplo, tiempo promedio que permanece una entidad en el sistema sean más confiables.
2. La simulación de Montecarlo es una simulación estática pero es fundamental para trabajar con los métodos de la transformada inversa, de composición y empírico.
3. Cuando se ajusta una muestra por una función polinómica ya sea mediante regresión o por técnicas numéricas y se puede calcular su inversa, lógicamente ese polinomio debe hacer un buen ajuste de los datos, entonces se aplica el método de la transformada inversa y se obtiene un excelente simulador de la variable aleatoria.
4. La Prueba de Bondad de Ajuste Chi-Cuadrada se usa para distribuciones discretas y la prueba de Kolmogorov-Smirnov para distribuciones continuas, pero cuando se aplica la segunda prueba a distribuciones discretas y rechaza a alguna de las distribuciones propuestas que podrían representar a la muestra de datos observados, podemos tener la plena seguridad

de que esa distribución no representa a la muestra. La prueba de Kolmogorov-Smirnov puede ser en todos los casos más poderosa que su alternativa, la prueba Chi-cuadrada, desde que no requiere que sus datos se agrupen de alguna manera.

5. Si de inicio se encuentra que una distribución de probabilidad tiene un *p-value* o *R-Sq* (coeficiente de determinación) muy cercano al 100%, entonces esa distribución o polinomio realmente representa a la muestra, el *p-value* = 0.729 para el Caso 2 es alto. Si el *p-value* o *R-Sq* es menor que el 50% tiene la alternativa de trabajar con el método empírico.
6. De todas maneras se hizo la generación de variables aleatorias en el potente *software* de simulación de procesos PROMODEL versión 7, y se extrajeron las muestras correspondientes en un *array* para comprobar que los generadores obtenidos estaban funcionando correctamente la simulación manual se hizo solo para comparar resultados. En el caso del método de aceptación y rechazo se corrió la simulación para 200, 1,000 y 5,000 horas obteniéndose en todos los casos que la probabilidad de que un valor generado y aceptado sea muy próximo a 49.41%, como se menciona en este artículo. También se trabajó con el *software* STAT::FIT versión 2 para realizar los histogramas, determinar las medidas descriptivas y para hallar y graficar la distribución de probabilidad que mejor ajuste realiza a la muestra real. También se utilizó el *software* Scientific WorkPlace versión 5.5 que se usó para la regresión no lineal, que también fue de gran ayuda en este artículo para la elaboración de las gráficas con polinomios y cálculos manuales, aunque es cierto que se podrían haber usado otras técnicas numéricas para estimar los polinomios. Asimismo, he tomado como referencia algunas aplicaciones que se presentan de las distintas distribuciones de probabilidad del manual de PROMODEL así como de la experiencia adquirida en la enseñanza de Simulación de Procesos con Eventos Discretos, y otras más de la bibliografía que se adjunta. El lector puede usar los diferentes paquetes estadísticos y de simulación de procesos que existen en el mercado como SPSS, MINITAB, ARENA y otros *softwares* diseñados para este propósito.



## BIBLIOGRAFÍA

### Libros:

- ARANDA Y GÓMEZ, (1992). *Fundamentos de estadística para economía y administración de empresas*. España, Editorial P.P.U.
- COSS BU. (1999). *Simulación un Enfoque Práctico*. México, Editorial LIMUSA SA.
- EPPEN, GOULD, SCHMIDT. (1992). *Investigación de Operaciones en la Ciencia Administrativa*. México, Ed. Prentice-Hall.
- LEVIN, RICHARD y RUBIN, DAVID. (1996). *Estadística para Administradores*. México, Editorial Prentice Hall Hispanoamericana SA.
- ROSS, SHELDON. (1999). *Simulación*. California, Editorial Prentice Hall.
- SCHMIDT Y TAYLOR. (1979). *Simulación de Sistemas Industriales*. México, Ed. Trillas SA. (Edición en español).
- SHANNON. (1988). *Simulación de Sistemas*. México, Ed. Trillas SA.

WINSTON, WAYNE L. (2005). *Investigación de Operaciones, Aplicaciones y Algoritmos*. México, International Thomson Editores, S. A.

### Internet:

- PALMER, JIMÉNEZ Y RUBÍ. (1999). *Uso de tablas estadísticas en Internet II: Cálculo de probabilidades para un conjunto de distribuciones utilizadas en Estadística*. España, Universitat de les Illes Balears. En: [http://www.intersalud.net/paginas/Num3/TablasII/tablas\\_2.htm#EXPO](http://www.intersalud.net/paginas/Num3/TablasII/tablas_2.htm#EXPO)
- EMILIO GÓMEZ, JOSÉ SARABIA Y FAUSTINO PRIETO. (2009). *La distribución Poisson-Beta: Aplicaciones y propiedades en la teoría del riesgo colectivo*. En <http://www.actuarios.org/esp/anales/2009/Pag%20141-160.pdf>
- JOSÉ TAMBORERO Y ANTONIO CEJALVO. *NTP418: Fiabilidad: La distribución Lognormal*. España, Centro Nacional de Condiciones de Trabajo. En [http://www.jmcprl.net/ntps/@datos/ntp\\_418.htm](http://www.jmcprl.net/ntps/@datos/ntp_418.htm)
- WIKIPEDIA. *Distribución Logística*. En [http://es.wikipedia.org/wiki/Distribucion\\_logistica](http://es.wikipedia.org/wiki/Distribucion_logistica)