EMPLEO DE LA PROGRAMACIÓN EN LOS MÉTODOS PROBABILÍSTICOS PARA LA GENERACIÓN DE NÚMEROS ALEATORIOS Y SUS APLICACIONES EN LA SIMULACIÓN

José Antonio Molinet Berenguer, Yoan Martínez López, Laura Casas Fuentes
Universidad de Camagüey "Ignacio Agramonte y Loynaz". Cuba
jose.molinet@reduc.edu.cu, yoan.martinez@reduc.edu.cu, laura.casas@reduc.edu.cu

Resumen: En el presente artículo se propone la utilización de un lenguaje de programación para el aprendizaje, en las carreras técnicas, de los métodos probabilísticas. De manera específica, consideramos que para el estudio de los métodos probabilísticas que son utilizados en la generación de los números aleatorios, que se imparten en la asignatura de Probabilidades, el estudiante al usar la programación como herramienta, puede visualizar mejor, a través de ejemplos, las aplicaciones de las distribuciones: Normal, Poisson, Binomial, Exponencial y Muller, en la simulación de procesos y fenómenos de acuerdo a su perfil profesional.

Palabras clave: Programación, números aleatorios, simulación.

Abstract: This article proposes the use of a programming language for learning of probabilistic methods in technical careers. Specifically, we believe that the study of probabilistic methods are used in generating random numbers, which are taught in the course of probabilities, for students to use programming as a tool, which they can see better through examples, distributions applications: Normal, Poisson, Binomial, Exponential, and Muller, in processes and phenomena simulation according to their profile.

Key words: Programming, random numbers, simulation.

Introducción

El número de eventos que se pueden simular empleando los números aleatorios generados por las distintas distribuciones ocupan diversas esferas (Bodt & Taylor, 1982). La distribución de Poisson por ejemplo modela eventos tales como el número de autos que pasan a través de un cierto punto en una ruta durante un periodo definido de tiempo, el número de llamadas telefónicas en una central telefónica por minuto, el número de servidores web accedidos por minuto, el número de mutaciones de determinada cadena de ADN después de cierta cantidad de radiación.

La programación de estas distribuciones, empleando para ello cualquier lenguaje de programación, permite al estudiante simular el comportamiento de sistemas que se ajusten a ellas (Navidi, 2007). Lo cual facilita la comprensión de procesos y fenómenos que serán cotidianos en su labor profesional y a su vez el estudio de las distintas condiciones límites a las que podrían verse sujetos dichos procesos, que abarcan desde lo industrial hasta lo social y biológico. La automatización de los cálculos permite que el estudiante resuelva en un tiempo razonable ejemplos de procesos reales, sin importar cuan extenso y complejos sean los cálculos, además de dotarlo de gráficas que faciliten la toma de decisiones.

Desarrollo

En los cursos de probabilidades impartidos en las carreras técnicas, se hace énfasis en el comportamiento estocástico de muchos sistemas reales que no pueden ser siempre caracterizados por una distribución uniforme (Tausworthe, 1965). Aparecen mucho más frecuentemente otras distribuciones tales como la Normal, Exponencial, Poisson y Binomial en fenómenos tales como el número de autos que pasan a través de un cierto punto en una ruta durante un periodo definido de tiempo, el número de servidores web accedidos por minuto o las mediciones físicas en áreas como los experimentos meteorológicos (Pitman, 1993). A lo largo de su vida profesional, el estudiante tendrá que enfrentarse a problemas relacionados con sistemas como los mencionados anteriormente, para solucionar estos problemas tendrá que auxiliarse de técnicas de simulación que exigen en particular conocer de métodos que permitan generar números aleatorios con cualquier distribución así como técnicas de prueba de bondad de ajuste a una distribución arbitraria, entre las teóricas clásicas o una empírica (Press, Teukolsky, Vetterling, & Flannery, 1992).

Hoy en día la Simulación es ampliamente empleada por tres razones principales:

- *Complementar la teoría:* Aunque se entiendan los procesos y las leyes que lo rigen, a veces no es posible obtener una solución analítica (o numérica). Además, en ocasiones se desconocen las condiciones iniciales u otra información referente al fenómeno.
- *Complementar los experimentos:* Los experimentos pueden ser muy costosos, peligrosos o resulta difícil realizar mediciones directamente.
- *La potencia de cómputo aumenta mientras que el costo disminuye:* En 1965 Gordon Moore enunciaba la ley, según la cual, el número de componentes en un chip se duplicaba cada año. Luego en 1975 se modificó a: duplicarse cada dos años.

Un correcto acercamiento de la simulación al fenómeno real, depende directamente de la generación de los números aleatorios según la distribución correspondiente. Generar un conjunto de números aleatorios consiste en obtener los valores x pertenecientes al dominio $[x_{mín}, x_{máx}]$, de forma tal que su frecuencia de ocurrencia coincida con el valor que toma la función de densidad de probabilidad $f(x)$.

Ejemplos de distribuciones y su programación

Distribución Normal

En estadística y probabilidades se llama distribución normal, distribución de Gauss o distribución gaussiana, a una de las distribuciones de probabilidad de variable continua que con más frecuencia aparece en fenómenos reales.

La importancia de esta distribución radica en que permite modelar numerosos fenómenos naturales, sociales y psicológicos. Mientras que los mecanismos que subyacen a gran parte de este tipo de fenómenos son desconocidos, por la enorme cantidad de variables incontrolables que en ellos intervienen, el uso del modelo normal puede justificarse asumiendo que cada observación se obtiene como la suma de unas pocas causas independientes (Walpole, Myers, & Myers, 1992).

La función de densidad de esta distribución está dada por la siguiente ecuación:

$$f(x) = \frac{1}{\sqrt{2\pi}\,\sigma} \exp\left[\frac{-(x-\mu)^2}{2\sigma^2}\right] \qquad (1)$$

La Figura 1 muestra el gráfico de la función densidad de la distribución normal.

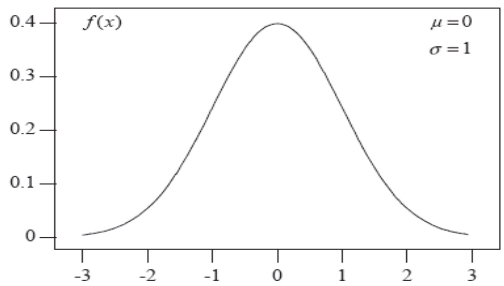

Figura 1 Gráfica de la función normal.

A continuación se muestra el algoritmo que nos permite programar esta distribución (en este caso se ha empleado el lenguaje C++) de forma tal que al pasarle los parámetros necesarios se genere un conjunto de valores que sigan esta distribución y nos permita aplicarlos a un ejemplo real.

Algoritmo:

(1) Generar a=U(-1,1) y b=U(-1,1) donde U es la distribución uniforme en (-1,1)

(2) $U=a^2+b^2$

(3) Si U<1, return ; en caso contrario ir a (1)

Código en C++:

```
double normal(double mu, double sigma)
{
  assert(sigma>0);
```

```
  double p,p1,p2;
  do{
    p1 = uniform(-1.,1.);
    p2 = uniform(-1.,1.);
    p = p1*p2+p2*p2;
 }while(p>=1.);
 return  mu + sigma*p1*sqrt(-2. * log(p) / p);
}
```

Algunas aplicaciones de variables asociadas a fenómenos que siguen esta distribución:

- Caracteres morfológicos de individuos (la estatura).
- Nivel de ruido en telecomunicaciones.
- Caracteres fisiológicos como el efecto de un fármaco.
- Errores cometidos al medir ciertas magnitudes.
- Caracteres sociológicos como el consumo de cierto producto por un mismo grupo de individuos.

Distribución Binomial

En estadísticas, la distribución binomial es una distribución de probabilidad discreta del número de sucesos en una secuencia de n experimentos independientes, los cuales tienen probabilidad θ de ocurrir.

La variable aleatoria binomial y su distribución están basadas en un experimento que satisface las siguientes condiciones:

- El experimento consiste en una secuencia de *n* intentos, donde *n* se fija antes del experimento.
- Los intentos son idénticos, y cada uno de ellos puede resultar en dos posibles resultados, que denotamos por éxito *(S)* o fracaso *(F)* (p(S)+p(F)=1).
- Los intentos son independientes, por lo que el resultado de cualquier intento en particular no influye sobre el resultado de cualquier otro intento.
- La probabilidad de éxito es constante de un intento a otro.

Siguiendo estas premisas la variable aleatoria binomial *X* está definida como el número de éxitos entre los *N* intentos.

La ecuación (2) muestra la fórmula de la función de densidad de la distribución binomial:

$$f(k) = \begin{cases} \binom{n}{k} p^k (1-p)^{n-k} & k \in \{0, 1, \cdots, n\} \\ 0 & \text{otherwise} \end{cases} \qquad (2)$$

La Figura 2 muestra el gráfico de la función binomial:

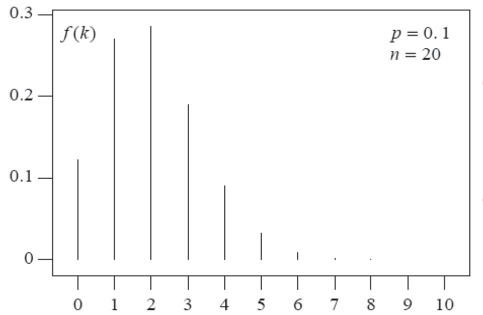

Figura 2 Gráfica de la función binomial.

Para generar un conjunto de números aleatorios empleando la distribución binomial se puede emplear el siguiente algoritmo y su correspondiente implementación en el lenguaje C++.

Algoritmo

(1) Generar a=U(0,1) donde U es la distribución uniforme en (0,1)

(2) Generar n valores de *X* independientes, donde *X*=1 si U<p, sino *X*=0

(3) Return $X = X_1 + \ldots + X_n$

Código en C++:

```
int binomial (int n, double p)
{
   assert(0. <= p && p <= 1. && n>=1);
   int sum = 0;
   for( int i = 0; i < n; i++) sum += bernoulli(p);
   return sum;
}
```

Algunos fenómenos que siguen una distribución binomial:

- ❖ Al nacer un/a bebé puede ser varón o hembra.

- En el deporte un equipo puede ganar o perder.
- En pruebas de cierto o falso sólo hay dos alternativas.
- Un tratamiento médico puede ser efectivo o inefectivo.
- La meta de producción o ventas del mes se pueden o no lograr.
- En pruebas de selección múltiple, aunque hay cuatro o cinco alternativas, se pueden clasificar como correcta o incorrecta.

Distribución Poisson

En teoría de probabilidades y estadísticas, la distribución de Poisson es una distribución de probabilidad discreta que expresa, a partir de una frecuencia de ocurrencia media, la probabilidad que ocurra un determinado número de eventos durante cierto periodo de tiempo.

La ecuación (3) muestra la fórmula de la función de densidad de la distribución de Poisson:

$$f(k) = \begin{cases} \frac{\mu^k}{k!} e^{-\mu} & k \in \{0, 1, \cdots\} \\ 0 & \text{otherwise} \end{cases} \qquad (3)$$

La Figura 3 muestra el gráfico de la función Poisson:

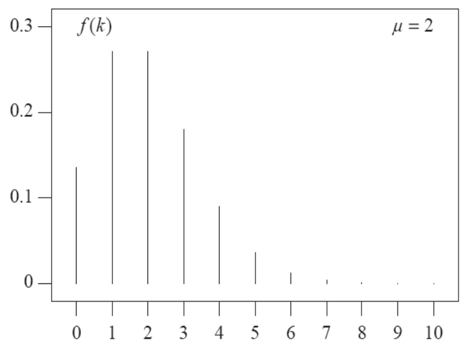

Figura 3 Gráfica de la función Poisson.

Para obtener de forma automática un conjunto de valores que se rijan por esta distribución se puede emplear el siguiente algoritmo y escribirlo en un determinado lenguaje de programación.

Algoritmo:

(1) $a=e^{-\mu}$, $b=1$ y $i=0$

(2) Generar U_{i+1} = U(0,1) y remplazar b por $b* U_{i+1}$, donde U es la distribución uniforme en (0,1)

(3) Si $b<a$, return X=i; en caso contrario i=i+1 e ir a (2)

Código en C++:

```
int poisson(double mu)
 {
  assert( mu > 0.);
  double b = 1.;
  int i;
  for( i = 0; b>=exp(-mu); i++) b* = uniform( 0., 1. );
  return i - 1;
```

La distribución de Poisson se aplica a varios fenómenos discretos de la naturaleza (esto es, aquellos fenómenos que ocurren 0, 1, 2, 3,... veces durante un periodo definido de tiempo o en un área determinada) cuando la probabilidad de ocurrencia del fenómeno es constante en el tiempo o el espacio (Walpole, Myers, & Myers, 1992). Ejemplos de estos eventos que pueden ser modelados por la distribución de Poisson incluyen:

- El número de autos que pasan a través de un cierto punto en una ruta (suficientemente distantes de los semáforos) durante un periodo definido de tiempo.
- El número de errores de ortografía que uno comete al escribir una única página.
- El número de llamadas telefónicas en una central telefónica por minuto.
- El número de servidores web accedidos por minuto.
- El número de animales muertos encontrados por unidad de longitud de ruta.
- El número de núcleos atómicos inestables que decayeron en un determinado período en una porción de sustancia radiactiva. La radiactividad de la sustancia se debilitará con el tiempo, por lo tanto el tiempo total del intervalo usado en el modelo debe ser significativamente menor que la vida media de la sustancia.
- El número de estrellas en un determinado volumen de espacio.
- La distribución de receptores visuales en la retina del ojo humano.
- El número de llamadas telefónicas en una central telefónica por minuto.
- El número de autos que pasan a través de un cierto punto en una ruta (suficientemente distantes de los semáforos) durante un periodo definido de tiempo.

- El número de mutaciones de determinada cadena de ADN después de cierta cantidad de radiación.

A partir de todas estas distribuciones estudiadas y de los algoritmos anteriormente vistos, se pueden obtener códigos en diferentes lenguajes de programación y así crear programas que permitan el cálculo estadístico. Haciendo uso de estos programas el estudiante puede aplicar las distintas distribuciones en la solución de problemas reales como los que fueron expuestos anteriormente, sin que el gran volumen de información que estos generan pueda impedir su tratamiento.

En nuestra experiencia encontramos dos formas principales de aplicar los programas de computación obtenidos. La primera consiste en que el estudiante para simular determinado fenómeno utilice estos códigos y genere un conjunto de números aleatorios y así analizar determinados aspectos del mismo, sin tener que esperar a medir realmente estos datos. La segunda forma reside en calcular los parámetros de entrada de las distribuciones utilizando un conjunto de valores reales, para posteriormente generar los valores a través de la fórmula ya programada y comparar la distribución obtenida con los valores reales.

Conclusiones

En el presente trabajo se realizó un análisis de las diferentes distribuciones que los estudiantes deben dominar, para la aplicación de los mismos en su perfil técnico. También se evidenció como el empleo de la programación representa una importante herramienta en los métodos probabilísticos para la generación de números aleatorios y sus aplicaciones en la simulación.

La generación automática de un conjunto de números que sigan una determinada distribución permite al estudiante realizar un análisis detallado de fenómenos reales que no serían posibles reproducir. De esta manera, el uso de la computación y el conocimiento de cualquier lenguaje de programación facilitan la comprensión mediante gráficas y aplicaciones reales de un número de distribuciones probabilísticas tratadas en los cursos de Probabilidades y Estadísticas.

Referencias bibliográficas

Bodt, B. A., & Taylor, M. S. (1982). *A Data Based Random Number Generator for a Multivariate Distribution.* Aberdeen: Defense Technical Information Center.

Navidi, W. (2007). *Estadística para ingenieros y científicos.* Mexico: McGraw-Hill.

Pitman, J. (1993). *Probability.* New York: Springer-Verlag.

Press, W., Teukolsky, S., Vetterling, W., & Flannery, B. (1992). *Numerical Recipes in C: The Art of Scientific Computing.* New York: Cambridge University Press.

Tausworthe, R. C. (1965). Random Numbers Generated by Linear Recurrence Modulo Two. *Mathematics of Computation*, 201-209.

Walpole, R. E., Myers, R. H., & Myers, S. L. (1992). *Probabilidad y Estadística para Ingenieros.* México: McGraw-Hill