Generación de Variables Aleatorias

UCR – ECCI
CI-1453 Investigación de Operaciones
Prof. M.Sc. Kryscia Daviana Ramírez Benavides

Introducción

- Las variables aleatorias se representan por medio de distribuciones de probabilidad.
- El procedimiento para generar variables aleatorias a partir de distribuciones de probabilidad se conoce como **generación de variables aleatorias** o **muestreo de Monte Carlo**.
- El principio del muestreo se basa en la interpretación de frecuencia de la probabilidad y requiere un flujo permanente de números aleatorios.

Tipos de Generadores

- Generador congruencial lineal (más utilizado).
- Generador multiplicativo.
- Generador mixto.

Generador Congruencial Lineal

- Produce una secuencia de enteros x_1, x_2, …; entre 0 y m-1 de acuerdo a la siguiente relación recursiva:

$$x_{i+1} = (ax_i + c) \bmod m$$
$$x_0 = \text{semilla} \qquad (i = 0,1,2,\ldots)$$

- Esto da el residuo de la división de (ax_i + c) entre m.
- Los números aleatorios se entregan por medio de la relación:

$$r_i = \frac{x_i}{m}$$
$$(i = 1,2,3,\ldots)$$

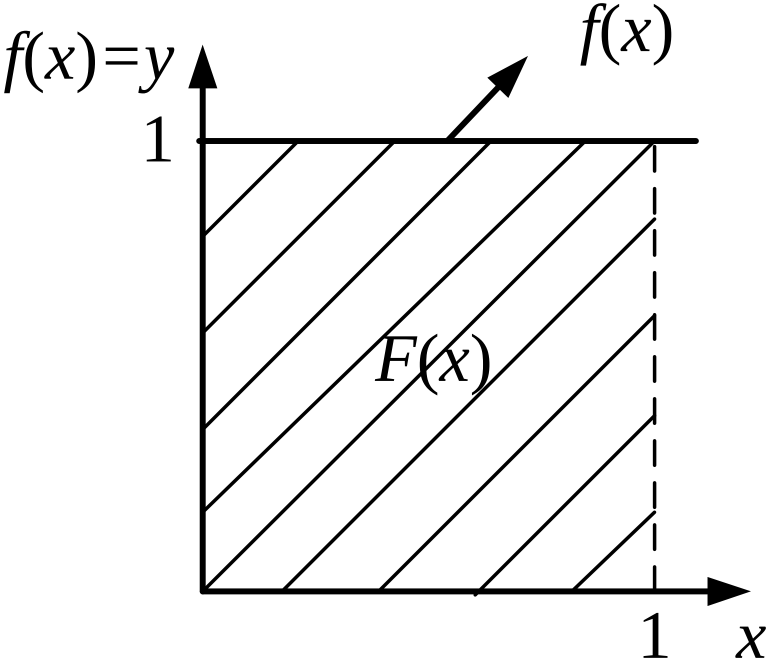
Generador Congruencial Lineal (cont.)

- Técnicamente, un número aleatorio, r_i, se define como una muestra aleatoria independiente extraída de una distribución uniforme continua, cuya función de densidad de probabilidad (**fdp**) está dada por:

$$f(x) = \begin{cases} 1 & 0 \le x \le 1 \\ 0 & \text{en caso contrario} \end{cases}$$

- Así cada número aleatorio estará distribuido de manera uniforme en el intervalo entre 0 y 1.
- Debido a esto, a estos números aleatorios se les conoce como números aleatorios $U(0,1)$, o simplemente como **números aleatorios uniformes**.

Distribución Uniforme



Generación de Valores para Variables Aleatorias Discretas

- El muestreo de Monte Carlo se logra al asignar intervalos de números aleatorios de acuerdo a las probabilidades en la distribución especificada.
- Este método consiste en los siguientes pasos:
  1. Se realiza una segmentación, para cada probabilidad de la distribución se le asigna un rango de valores según su valor.
  2. Se genera un número aleatorio entero *r* entre 00 y 99.
     - Cada número aleatorio en una secuencia (en este caso de 00 a 99) tiene una probabilidad igual (en este caso 0.01) de aparecer, y cada uno es independiente de los números antes y después de él.
  3. Se devuelve la variable aleatoria discreta de la distribución que corresponda con el rango donde pertenece el número aleatorio generado.

Generación de Valores para Variables Aleatorias Discretas (cont.)

Distribución de Tiempo de Servicio		Generación de Valores
Tiempo de Servicio (minutos)	**Probabilidad**	**Rango**
1	0.35	0 – 34
2	0.40	35 – 74
3	0.25	75 – 99

- Al generar un número aleatorio salió el 62.
- El número aleatorio 62 pertenece al rango 35 – 74, por lo que, la variable aleatoria discreta de la distribución de tiempo de servicio que se escoge es 2 minutos.

Generación de Valores para Variables Aleatorias Continuas

- El muestreo de Monte Carlo se logra al transformar los números aleatorios en una variable aleatoria continua a partir de la distribución especificada.
- Hay muchos métodos diferentes para generar variables aleatorias continuas.
- La selección de un algoritmo particular dependerá de la distribución a partir de la cual se quiere generar, tomando en cuenta factores como la exactitud de las v.a., las eficiencias de cómputo y almacenaje, y la complejidad del algoritmo.

Generación de Valores para Variables Aleatorias Continuas (cont.)

- Los dos algoritmos utilizados con más frecuencia son:
  - **Método de transformación inversa (ITM)**.
  - **Método de aceptación-rechazo (ARM)**.
- Entre estos dos métodos, es posible generar variables aleatorias de casi todas las distribuciones utilizadas con más frecuencia.
- Además, se cuenta con dos métodos para generar variables aleatorias a partir de la distribución normal:
  - **Algoritmo de convolución**.
  - **Método directo**.

Método de Transformación Inversa (ITM)

- Este método requiere una función de distribución acumulada (**fda**) $F(x)$ se puede obtener en forma cerrada (tiene fórmula y es posible calcular $F^{-1}(x)$).
- Como ejemplos se tienen las distribuciones exponencial, uniforme, triangular y la de Weibull.

Método de Transformación Inversa (ITM) (cont.)

- Este método consiste en los siguientes pasos:
  1. Dada la función de densidad de probabilidad $f(x)$, se elabora la función de distribución acumulada como:

$$F(x) = \int_{-\infty}^{x} f(t)dt$$

  2. Se genera un número aleatorio $r \in [0,1]$.
  3. Se establece $F(x) = r$ y se determina el valor de x. La variable x es entonces una variable aleatoria continua de la distribución cuya fdp está dada por $f(x)$.

Método de Transformación Inversa (ITM) Función de Rampa

- Se tiene la siguiente fdp:

$$f(x)=\begin{cases}\dfrac{x}{2} & 0\le x\le 2\\ 0 & \text{en caso contrario}\end{cases}$$

- Se calcula fda:

$$F(x)=\int_0^x \frac{t}{2}dt=\frac{x^2}{4} \qquad F(x)=\begin{cases}0 & x<0\\ \dfrac{x^2}{4} & 0\le x\le 2\\ 1 & x\ge 2\end{cases}$$

Método de Transformación Inversa (ITM) Función de Rampa (cont.)

- Se genera un número aleatorio r, por ejemplo, $r = 0.64$.
- Se establece $F(x) = r$ y se determina el valor de x:

$$F(x) = r \qquad x = \pm 2\sqrt{r}$$

$$\frac{x^2}{4} = r \qquad x = 2\sqrt{0.64} = 1.6$$

- Puesto que los tiempos de servicio se definen sólo para valores positivos de x, no es factible un tiempo de servicio $x = -2\sqrt{r}$; entonces se queda como solución $x = 2\sqrt{r}$. Esta ecuación se llama **generador de variables aleatorias** o un **generador de proceso**.

Método de Transformación Inversa (ITM) Distribución Exponencial

- Se tiene la siguiente fdp:

$$f(x)=\begin{cases}\lambda e^{-\lambda x} & x\geq 0, \lambda > 0\\ 0 & \text{en caso contrario}\end{cases}$$

- Se calcula fda:

$$F(x)=\int_0^x \lambda e^{-\lambda t}dt = 1-e^{-\lambda x} \quad F(x)=\begin{cases} 0 & x<0\\ 1-e^{-\lambda x} & x\geq 0\end{cases}$$

Método de Transformación Inversa (ITM) Distribución Exponencial (cont.)

- Se genera un número aleatorio r, por ejemplo, r = 0.
- Se establece $F(x) = r$ y se determina el valor de x:

$$F(x) = r$$

$$1 - e^{-\lambda x} = r$$

$$e^{-\lambda x} = 1 - r$$

$$-\lambda x = \ln(1 - r)$$

$$x = -\frac{1}{\lambda}\ln(1 - r)$$

$$x = -\frac{1}{\lambda}\ln(1 - 0) = 0$$

Método de Transformación Inversa (ITM) Distribución Uniforme Continua

- Se tiene la siguiente fdp:

$$f(x)=\begin{cases}\dfrac{1}{b-a} & a\le 0\le b\\ 0 & \text{en caso contrario}\end{cases}$$

- Se calcula fda:

$$F(x)=\int_a^x \frac{1}{b-a}dt=\frac{x-a}{b-a} \qquad F(x)=\begin{cases}0 & x<a\\ \dfrac{x-a}{b-a} & a\le x\le b\\ 1 & x>b\end{cases}$$

Método de Transformación Inversa (ITM) Distribución Uniforme Continua (cont.)

- Se genera un número aleatorio r, por ejemplo, $r = 0.5$.
- Se establece $F(x) = r$ y se determina el valor de x:

$$F(x) = r$$

$$\frac{x-a}{b-a} = r$$

$$x - a = r(b-a)$$

$$x = a + r(b-a)$$

$$x = a + 0.5(b-a) = 0.5a + 0.5b$$

Método de Transformación Inversa (ITM) Distribución Triangular

- Se tiene la siguiente fdp:

$$f(x)=\begin{cases}\frac{1}{2}(x-2) & 2\le x\le 3\\ \frac{1}{2}\left(2-\frac{x}{3}\right) & 3\le x\le 6\\ 0 & \text{en caso contrario}\end{cases}$$

- Se calcula fda:

$$F_1(x)=\int_2^x \frac{1}{2}(t-2)dt=\frac{1}{4}(x-2)^2$$

$$F_2(x)=\int_3^x \frac{1}{2}\left(2-\frac{t}{3}\right)dt=-\frac{1}{12}\left(x^2-12x+24\right)$$

$$F(x)=\begin{cases}0 & x<2\\ \frac{1}{4}(x-2)^2 & 2\le x\le 3\\ -\frac{1}{12}\left(x^2-12x+24\right) & 3\le x\le 6\\ 1 & \text{en caso contrario}\end{cases}$$

Método de Transformación Inversa (ITM) Distribución Triangular (cont.)

- La función $F_1(x)$ cubre el primer 25% del intervalo fda y la función $F_2(x)$ cubre el restante 75% del intervalo.
- Debido a que tenemos dos funciones, si $r < 0.25$ se usa la función $F_1(x)$; en caso contrario ($r \geq 0.25$) se utiliza la función $F_2(x)$.

$$F_1(x) = r \qquad 0 \leq r < 0.25$$

$$\frac{1}{4}(x-2)^2 = r$$

$$F_2(x) = r \qquad 0.25 \leq r \leq 1$$

$$-\frac{1}{12}\left(x^2 - 12x + 24\right) = r$$

Método de Transformación Inversa (ITM) Distribución Triangular (cont.)

- Se genera un número aleatorio *r*, por ejemplo, *r* = 0 y *r* = 2/3.
- Se establece *F*(*x*) = *r* y se determina el valor de *x*:

$$F_1(x) = r$$
$$\frac{1}{4}(x-2)^2 = r$$
$$x - 2 = \pm\sqrt{4r}$$

$$F_2(x) = r$$
$$-\frac{1}{12}\left(x^2 - 12x + 24\right) = r$$
$$x^2 - 12x = -24 - 12r$$
$$x^2 - 12x + 36 = 36 - 24 - 12r$$
$$(x-6)^2 = 12 - 12r$$
$$x - 6 = \pm\sqrt{12 - 12r}$$

$$x = 2 + \sqrt{4r} \qquad x = 6 - 2\sqrt{3 - 3r}$$

$$x = 2 + \sqrt{4 \times 0} = 2 \qquad x = 6 - 2\sqrt{3 - 3 \times 2/3} = 4$$

Método de Aceptación-Rechazo (ARM)

- Este método requiere una función de distribución acumulada (**fda**) $F(x)$ esté definida en un intervalo finito.
- Como ejemplos se tienen la función rampa, las distribuciones triangular, beta y la de Erlang.

Método de Aceptación-Rechazo (ARM) (cont.)

- Este método consiste en los siguientes pasos:
  1. Se selecciona una constante M, tal que M es el valor más grande de $f(x)$ en el intervalo $[a, b]$.
  2. Se genera dos números aleatorios r_1 y r_2, $r_1, r_2 \in [0,1]$.
  3. Se calcula $x^* = a + (b - a)r_1$. (Esto asegura que cada miembro de $[a, b]$ tiene una probabilidad igual de ser elegido como x^*).
  4. Se evalúa la función $f(x)$ en el punto x^*; sea $f(x^*)$.
  5. Si $r_2 \leq f(x^*) / M$, entonces se acepta x^* como una variable aleatoria continua. De lo contrario, se rechaza x^* y se vuelve al paso 2.

Método de Aceptación-Rechazo (ARM) Función de Rampa

- Se tiene la siguiente fdp:

$$f(x)=\begin{cases}2x & 0\le x\le 1\\ 0 & \text{en caso contrario}\end{cases}$$

- El valor más grande de $f(x)$ ocurre $x = 1$ y es igual a 2. Entonces, $M = 2$.
- Se genera dos números aleatorios r_1 y r_2, por ejemplo, $r_1 = 0$ y $r_2 = 1$.
- Se calcula $x^* = a + (b - a)r_1 = 0 + (1 - 0)0 = 0$
- Se evalúa $f(x^*) = 2x^* = 2*0 = 0$
- Se comprueba $r_2 \le f(x^*) / M$; $1 \ge 0/2$, por lo que se rechaza y se vuelve al paso 2.

Método de Aceptación-Rechazo (ARM) Función de Rampa (cont.)

- Se genera dos números aleatorios r_1 y r_2, por ejemplo, $r_1 = 0.5$ y $r_2 = 0.2$.
- Se calcula $x^* = a + (b - a)r_1 = 0 + (1 - 0)0.5 = 0.5$
- Se evalúa $f(x^*) = 2x^* = 2*0.5 = 1$
- Se comprueba $r_2 \leq f(x^*) / M$; $0.2 \leq 1/2$, por lo que se acepta x^* como una variable aleatoria continua, $x^* = 0.5$.

Método de Aceptación-Rechazo (ARM) Distribución Triangular

- Se tiene la siguiente fdp:

$$f(x) = \begin{cases} -\frac{1}{6} + \frac{x}{12} & 2 \leq x \leq 6 \\ \frac{4}{3} - \frac{x}{6} & 6 \leq x \leq 8 \\ 0 & \text{en caso contrario} \end{cases}$$

- El valor más grande de $f(x)$ ocurre $x = 6$ y es igual a 1/3. Entonces, $M = 1/3$.
- Se genera dos números aleatorios r_1 y r_2, por ejemplo, $r_1 = 0$ y $r_2 = 1$.
- Se calcula $x^* = a + (b - a)r_1 = 2 + (8 - 2)0 = 2$

Método de Aceptación-Rechazo (ARM) Distribución Triangular (cont.)

- La función $f_1(x^*)$ cubre 2/3 del intervalo fdp y la función $f_2(x^*)$ cubre el restante 1/3 del intervalo.
- Debido a que tenemos dos funciones, si $2 \leq x^* \leq 6$ se usa la función $f_1(x^*)$; en caso contrario ($6 \leq x^* \leq 8$) se utiliza la función $f_2(x^*)$.

$$f(x^*) = f_1(x^*) = -\frac{1}{6} + \frac{x^*}{12} \qquad 2 \leq x^* \leq 6$$

$$f(x^*) = f_2(x^*) = \frac{4}{3} - \frac{x^*}{6} \qquad 6 \leq x^* \leq 8$$

Método de Aceptación-Rechazo (ARM) Distribución Triangular (cont.)

- Se evalúa $f_1(x^*)$ porque $x^* = 2$:

$$f(x^*) = f_1(x^*) = -\frac{1}{6} + \frac{x^*}{12} = -\frac{1}{6} + \frac{2}{12} = 0$$

- Se comprueba $r_2 \leq f(x^*) / M$; $1 \geq 0/(1/3)$, por lo que se rechaza y se vuelve al paso 2.

Método de Aceptación-Rechazo (ARM) Distribución Triangular (cont.)

- Se genera dos números aleatorios r_1 y r_2, por ejemplo, $r_1 = 0.75$ y $r_2 = 0.5$.
- Se calcula $x^* = a + (b - a)r_1 = 2 + (8 - 2)0.75 = 6.5$
- Se evalúa $f_2(x^*)$ porque $x^* = 6.5$:

$$f\left(x^*\right) = f_2\left(x^*\right) = \frac{4}{3} - \frac{x^*}{6} = \frac{4}{3} - \frac{6.5}{6} = \frac{1}{4}$$

- Se comprueba $r_2 \leq f(x^*) / M$; $0.5 \leq (1/4)/(1/3)$, por lo que se acepta x^* como una variable aleatoria continua, $x^* = 6.5$.

Distribución Normal

- Esta distribución es muy importante, por lo cual se ha creado varios algoritmos para la generación de variables aleatorias.
- Tanto el método ITM como el método ARM, son inapropiados para esta distribución porque:
  - La fdp no existe en forma cerrada.
  - La distribución no está definida en un intervalo finito.
- Aunque es posible usar métodos numéricos en el ITM y truncar la distribución para el método ARM; otros métodos tienden a ser más eficaces:
  - Algoritmo de convolución.
  - Método directo.

Distribución Normal
Algoritmo de Convolución

- En este algoritmo se hace uso del teorema del límite central.
- **Teorema del Límite Central:** La suma X de n variables aleatorias independientes y distribuidas de manera idéntica (por ejemplo, x_1, x_2, …, x_n, cada una con media μ ($E(x_i) = \mu$) y varianza finita σ^2 ($Var(x_i) = \sigma^2$)) tiene una distribución aproximadamente normal con media $n\mu$ y varianza $n\sigma^2$.

$$\sum_{i=1}^{n} x_i \Rightarrow \mathrm{n}\left(n\mu, n\sigma^2\right)$$

$$\overline{X} = \frac{\sum_{i=1}^{n} x_i}{n}$$

Distribución Normal
Algoritmo de Convolución (cont.)

- Sea *X* con distribución n(μ, σ^2), se va a generar un valor *Z* con distribución n(0, 1).
- Si esto se aplica a *R*(0, 1) variables aleatorias; r_1, r_2, …, r_n , con media μ y varianza finita σ^2, se deduce que:

$$z = \frac{\sum_{i=1}^{n} r_i - n\mu}{\left(n\sigma^2\right)^{1/2}} \Rightarrow \mathrm{n}(0,1)$$

Distribución Normal
Algoritmo de Convolución (cont.)

- Este método consiste en los siguientes pasos:
  - Se generan *n* variables aleatorias con distribución *U*(0, 1).
  - Se calcula la variable *z*:

$$z = \frac{\sum_{i=1}^{n} r_i - n\mu}{\left(n\sigma^2\right)^{1/2}} \Rightarrow \mathrm{n}(0,1)$$

  - Se calcula la variable aleatoria normal *x* de la siguiente forma:

$$x = z\sigma + \mu$$

Distribución Normal
Algoritmo de Convolución (cont.)

- Se utiliza casi siempre los siguientes valores:
  - $n = 12$.
  - $\mu = 0.5$
  - $\sigma^2 = 1/12$
- Se calcula la variable z:

$$z = \sum_{i=1}^{12} r_i - 6 \Rightarrow \mathrm{n}(0,1)$$

Distribución Normal
Método Directo

- Este método fue creado por Box y Muller (1958).
- No es tan eficaz como algunas nuevas técnicas, pero es fácil aplicarlo y ejecutarlo.
- El algoritmo genera dos números aleatorios $R(0, 1)$, r_1 y r_2, y luego los transforma en dos v.a. normales, cada una con media 0 y varianza 1.

Distribución Normal
Método Directo (cont.)

- Este método consiste en los siguientes pasos:
  - Se generan dos números aleatorios r_1 y r_2, $U(0, 1)$.
  - Se transforman en dos variables aleatorias normales, cada una con media 0 y varianza 1, usando las transformaciones directas:

$$z_1 = \left(-2\ln r_1\right)^{1/2} \operatorname{sen}\left(2\pi r_2\right)$$

$$z_2 = \left(-2\ln r_1\right)^{1/2} \cos\left(2\pi r_2\right)$$

  - Se calculan las variables aleatorias normales x_1 y x_2 de la siguiente forma:

$$x_1 = z_1\sigma + \mu$$

$$x_2 = z_2\sigma + \mu$$

Referencias Bibliográficas

- Winston, Wayne L. “Investigación de Operaciones”. Grupo Editorial Iberoamérica. México, 2005.