Generación de números aleatorios

Carlos Javier Uribe Martes

Ingeniería Industrial
Universidad de la Costa

Febrero 11, 2020

Contenido

Introducción

- Para desarrollar un modelo de simulación el ingrediente fundamental es la generación de una secuencia de números aleatorios $R_1, R_2, \ldots, R_n$ [1].

Propiedades de los números aleatorios

- Cada número aleatorio R_i debe ser una muestra *independiente* obtenida de una distribución *uniforme* continua entre cero y uno [1].

Propiedades de la distribución uniforme $[0, 1]$

$$f(x) = \begin{cases} 1, & 0 \leq x \leq 1 \\ 0, & \text{de lo contrario} \end{cases} \quad ; \quad E(R) = \frac{1}{2}; \quad V(R) = \frac{1}{12}$$

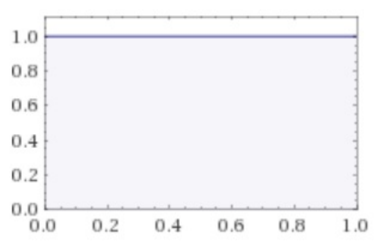

Figura: Función de densidad de probabilidad

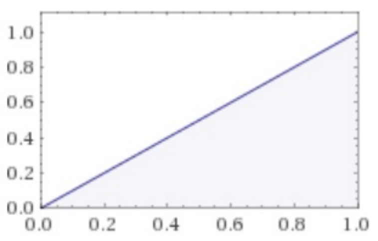

Figura: Función de probabilidad acumulada

Generación de números pseudo-aleatorios

- Generar números aleatorios a través de un algoritmo remueve la verdadera aleatoriedad, toda vez que el patrón puede ser repetido [1].
- Se busca generar una secuencia de números que *imite* las propiedades de los números aleatorios [1].

Técnicas para generación de números aleatorios

- Una técnica adecuada debe tener las siguientes características:
  - Eficiencia [3].
  - Periodo máximo [3].
  - Secuencia reproducible [3].
  - Portabilidad [2].

Cuadrados medios de Von Neumann y Metropolis

1. Seleccione una *semilla* X_0 con D dígitos $(D > 3)$.
2. Sea Y_0 el resultado de elevar X_0 al cuadrado, defina X_1 igual a los D dígitos del centro de Y_0 y sea $R_1 = 0.X_1$.
3. Sea Y_i el resultado de elevar X_i al cuadrado, defina X_{i+1} igual a los D dígitos del centro de Y_i y sea $R_{i+1} = 0.X_{i+1}$.

Generador congruencial lineal

- Utiliza la siguiente relación recursiva

$$X_{i+1} = (aX_i + c) \mod m, \quad i = 0, 1, 2, \ldots$$

- El valor inicial X_0 es llamado *semilla*, a es el *multiplicador*, c es el *incremento* y m el *módulo*, todos enteros no negativos.
- Para obtener R_i emplee:

$$R_i = \frac{X_i}{m}, \quad i = 1, 2, \ldots$$

Generador congruencial lineal

- Los valores de a, c, m y X_0 afectan directamente las propiedades estadísticas y el periodo de la secuencia generada [1].
- Deben satisfacerse las siguientes relaciones:
  - $a < m$
  - $c < m$
  - $m > 0$
  - $X_0 < m$
- La secuencia se repetirá con periodo $p \leq m$, por lo que el generador alcanza el periodo máximo cuando $p = m$

Generador congruencial multiplicativo

- Si el incremento $c = 0$, se denomina *método congruencial multiplicativo*.
- No alcanza el periodo máximo ya que la secuencia no contendrá $X_i = 0$, sin embargo pueden llegar a alcanzar el periodo $m - 1$ si se seleccionan m y a en forma adecuada [3]:
  - m ha de ser un número primo.
  - a ha de ser raíz primitiva de m, es decir,

$$a^n \mod m \neq 1 \quad n = 1, \ldots, m-2$$

Generador congruencial multiplicativo

- La siguiente tabla indica los parámetros evaluados por Fishman y Moore (1986) que tienen buen comportamiento [3].

Parámetros	**Fishman y Moore**
a	48.271
m	$2^{31} - 1$

Generador congrencial mixto

- Si el incremento $c \neq 0$, se denomina *método congruencial mixto*.
- Las siguientes condiciones aseguran que el generador congruencial mixto tendrá periodo máximo [2, 3]:
  1. El único entero positivo que divide a m y a c es 1, es decir, son primos entre sí.
  2. Si q es un número primo que divide a m, entonces q también divide a $a-1$.
  3. Si 4 divide a m, entonces 4 también divide a $a-1$.

Generadores congrenciales mixtos

- La siguiente tabla indica los generadores congruenciales lineales mixtos propuestos por Coveyou y MacPherson (1967) y por Kobayashi (1978) [3].

Parámetros	Kobayashi	Coveyou y MacPherson
a	314.159.269	5^{15}
c	453.806.245	1
m	2^{31}	2^{35}

Métodos congruenciales NO lineales

- **Algoritmo congruencial cuadrático**: Emplea la relación recursiva:

$$X_{i+1} = \left(aX_i^2 + bX_i + c\right) \mod m, \quad i = 0, 1, 2, \ldots$$

- **Algoritmo de Blum, Blum y Shub**: Es similar al algoritmo congruencial cuadrático, con $a = 1, b = 0, c = 0$, entonces la relación recursiva es:

$$X_{i+1} = X_i^2 \mod m, \quad i = 0, 1, 2, \ldots$$

Combinación de generadores congruenciales lineales

- Una manera de conseguir secuencias aleatorias con periodos más largos es combinar dos o más generadores congruenciales multiplicativos.

Generador de Wichmann-Hill

- Consiste en tres generadores congruenciales lineales con módulos primos. Cada uno es utilizado para producir un aleatorio entre 0 y 1.
- Estos tres resultados se suman, módulo 1, para obtener el resultado final.
  1. $x_i = 171x_{i-1} \mod 30269$
  2. $y_i = 172y_{i-1} \mod 30307$
  3. $z_i = 170z_{i-1} \mod 30323$
  4. $r_i = \left(\frac{x_i}{30269} + \frac{y_i}{30307} + \frac{z_i}{30323}\right) \mod 1$

MRG32k3a de L'Ecuyer

- Consiste en dos generadores congruenciales recursivos de tercer orden.
- Estos se combinan para obtener un nuevo aleatorio en cada iteración.
  1. $x_i = (1403580x_{i-2} - 810728x_{i-3}) \mod \left(2^{32} - 209\right)$
  2. $y_i = (527612y_{i-1} - 1370589y_{i-3}) \mod \left(2^{32} - 22853\right)$
  3. $z_i = (x_i - y_i) \mod \left(2^{32} - 209\right)$
  4. $r_i = \dfrac{z_i}{2^{32} - 209}$

Pruebas para números aleatorios

- Para verificar si las propiedades deseadas de un conjunto de números aleatorios diferentes tipos de pruebas pueden desarrollarse.
- Para cada prueba debe definirse un nivel de significancia α, el cual representa la probabilidad de rechazar la hipótesis nula cuando ésta es cierta:

$$\alpha = P(\text{rechazar } H_0 | H_0 \text{ es cierta})$$

Prueba de medias

- Busca comprobar que el valor esperado de los números en la secuencia R_i sea igual a 0.5 mediante las siguientes hipótesis:

$$H_0: \ \mu_{R_i} = 0{,}5$$
$$H_1: \ \mu_{R_i} \neq 0{,}5$$

Prueba de medias

1. Determine el promedio de los n números aletorios de la secuencia:

$$\bar{R} = \frac{1}{n}\sum_{i=1}^{n} R_i$$

2. Calcule los límites de aceptación inferior y superior:

$$LI_{\bar{R}} = \frac{1}{2} - z_{\alpha/2}\left(\frac{1}{\sqrt{12n}}\right)$$
$$LS_{\bar{R}} = \frac{1}{2} + z_{\alpha/2}\left(\frac{1}{\sqrt{12n}}\right)$$

3. Si el valor de $\bar{R}$ está dentro de los límites de aceptación no hay evidencia suficiente para rechazar H_0 con un nivel de confianza 1-α.

Prueba de varianza

- Busca determinar si la varianza de la secuencia aleatoria generada es igual a 1/12 mediante las siguientes hipótesis:

$$H_0 : \ \sigma^2_{R_i} = \frac{1}{12}$$
$$H_1 : \ \sigma^2_{R_i} \neq \frac{1}{12}$$

Prueba de varianza

1. Determine la varianza muestral de la secuencia $R_1, R_2, \ldots, R_n$:

$$V(R) = \frac{\sum_{i=1}^{n} (R_i - \bar{R})^2}{n-1}$$

2. Calcule los límites de aceptación inferior y superior mediante:

$$LI_{V(R)} = \frac{\chi^2_{\frac{\alpha}{2},n-1}}{12(n-1)}$$

$$LS_{V(R)} = \frac{\chi^2_{\frac{(1-\alpha)}{2},n-1}}{12(n-1)}$$

3. Si el valor de $V(R)$ está dentro de los límites de aceptación no hay evidencia suficiente para rechazar H_0 con un nivel de confianza 1-α.

Prueba de frecuencias

- Trata de determinar si el conjunto de números generados se distribuye de acuerdo con la distribución uniforme $[0,1]$ para lo cual formula las siguientes hipótesis:

$$H_0: \; R_i \sim U\left[0,1\right]$$
$$H_1: \; R_i \nsim U\left[0,1\right]$$

Prueba de frecuencias

Prueba chi-cuadrado

- Para la distribución uniforme, la frecuencia esperada en cada clase, E_i está dada por:

$$E_i = \frac{N}{n}$$

para n clases igualmente espaciadas, donde N es el número total de observaciones.

- Utiliza el estadístico de prueba:

$$\chi_0^2 = \sum_{i=1}^{n} \frac{(O_i - E_i)^2}{E_i}$$

donde O_i es la frecuencia observada en la i-ésima clase.

Prueba de frecuencias

Prueba chi-cuadrado

- La distribución muestral de χ_0^2 es aproximadamente chi-cuadrado con $n-1$ grados de libertad.
- Si el estadístico de prueba χ_0^2 es menor que el valor $\chi_{\alpha,n-1}^2$ no hay evidencia suficiente para rechazar H_0 con un nivel de confianza 1-α.

Prueba de frecuencias

Prueba Kolmogorov-Smirnov

- Contrasta la función de densidad acumulada $F(x)$ de la distribución teórica con la función de densidad empírica $S_N(x)$ de la muestra de N obsevaciones.
- Se basa en la mayor desviación absoluta entre $F(x)$ y $S_N(x)$ en el rango de la variable aleatoria, utilizando el estadístico de prueba:

$$D = max|F(x) - S_N(x)|$$

Prueba de frecuencias

Prueba Kolmogorov-Smirnov

1. Ordene los datos de menor a mayor. Sea $R_{(i)}$ la i-ésima menor observación.
2. Determine los valores:

$$D^{+} = max_{1 \leq i \leq N} \left\{ \frac{i}{N} - R_{(i)} \right\}$$

$$D^{-} = max_{1 \leq i \leq N} \left\{ R_{(i)} - \frac{i-1}{N} \right\}$$

3. Calcule $D = max(D^{+}, D^{-})$.
4. Identifique el valor crítico D_{α} correspondiente a α y N.
5. Si $D \leq D_{\alpha}$ se concluye que no hay evidencia suficiente para rechazar H_0 con un nivel de confianza 1-α.

Pruebas de independencia

En su mayoría buscan probar la independencia de los números de un conjunto R_i mediante las hipótesis:

$$\begin{aligned} H_0 &: \text{ los números del conjunto } R_i \text{ son independientes} \\ H_1 &: \text{ los números del conjunto } R_i \text{ no son independientes} \end{aligned}$$

Prueba de corridas arriba y abajo

1. Determine una secuencia de unos y ceros así: si $R_{i+1} \leq R_i$ asigne un cero a la secuencia, de lo contrario asigne un uno.
2. Defina C_0 como el número de corridas en la secuencia (una corrida es cualquier cantidad de unos o ceros consecutivos).
3. Determine el estadístico de prueba mediante las ecuaciones:

$$\mu_{C_0} = \frac{2n-1}{3}$$
$$\sigma^2_{C_0} = \frac{16n-29}{90}$$
$$Z_0 = \left| \frac{C_0 - \mu_{C_0}}{\sigma_{C_0}} \right|$$

4. Si el estadístico Z_0 es mayor que el valor crítico de $Z_{\alpha/2}$, se concluye que los números del conjunto R_i no son independientes.

Prueba de corridas arriba y abajo de la media

1. Determine una secuencia de unos y ceros así: si $R_i \leq 0{,}5$ asigne un cero a la secuencia, de lo contrario asigne un uno.
2. Defina C_0 como el número de corridas en la secuencia, n_0 el número de ceros de ceros y n_1 el número de unos.
3. Determine el estadístico de prueba mediante las ecuaciones:

$$\mu_{C_0} = \frac{2n_0 n_1}{n} + 0{,}5$$

$$\sigma^2_{C_0} = \frac{2n_0 n_1 (2n_0 n_1 - n)}{n^2(n-1)}$$

$$Z_0 = \frac{C_0 - \mu_{C_0}}{\sigma_{C_0}}$$

4. Si el estadístico Z_0 está fuera del intervalo $\left[-Z_{\alpha/2}, Z_{\alpha/2}\right]$ se concluye que los números del conjunto R_i no son independientes.

Prueba de autocorrelación

- Prueba la correlación entre los números generados y compara la correlación muestral con la correlación esperada de cero.
- Requiere el cálculo de la autocorrelación entre cada m números (siendo conocida m como *lag* o retraso), empezando con el i-ésimo número de la secuencia.
- La autocorrelación ρ_{im} entre los siguientes números será de interés R_i, $R_{i+m}, R_{i+2m}, \ldots, R_{i+(M+1)m}$. Donde el valor de M es el entero más grande tal que $i + (M+1)m \leq N$,

Prueba de autocorrelación

- Una autocorrelación diferente de cero implica una falta de independencia en los datos. La siguiente prueba con dos colas es adecuada:

$$H_0 : \ \rho_{im} = 0$$
$$H_1 : \ \rho_{im} \neq 0$$

- Para valores grandes de M, la distribución del estimador de ρ_{im}, denotado $\hat{\rho}_{im}$, es aproximadamente normal si los valores R_i, $R_{i+m}, R_{i+2m}, \ldots, R_{i+(M+1)m}$ no están correlacionados.

Prueba de autocorrelación

- El estadístico:

$$Z_0 = \frac{\hat{\rho}_{im}}{\sigma_{\hat{\rho}_{im}}}$$

  está normalmente distribuido con media 0 y varianza 1, bajo el supuesto de independencia y para valores grandes de M.

- Donde

$$\begin{aligned} \hat{\rho}_{im} &= \frac{1}{M+1}\left[\sum_{k=0}^{M} R_{i+km}R_{i+(k+1)m}\right] - 0{,}25 \\ \sigma_{\hat{\rho}_{im}} &= \frac{\sqrt{13M+7}}{12(M+1)} \end{aligned}$$

- No rechace H_0 si $-z_{\alpha/2} \leq Z_0 \leq z_{\alpha/2}$ donde $z_{\alpha/2}$ se puede obtener de la tabla de probabilidades para la distribución chi-cuadrado.

Referencias

- Banks, J., Carson II, J. S., Nelson, B. L. y Nicol, D. M. *Discrete-Event System Simulation.* Fifth (Pearson, 2014).
- Law, A. M. *Simulation modeling and analysis.* Fifth (McGraw-Hill, 2015).
- Pazos Arias, J. J., Suárez González, A. y Díaz Redondo, R. *Teoría de colas y simulación de eventos discretos.* (Prentice Hall, 2003).