Generación de Números Aleatorios

Propósito

- Métodos de generación de números aleatorios.

- Test de Aleatoriedad:
  - Test de Frecuencia.
  - Test de Autocorrelación.

Generación de Números Pseudo-Aleatorios

- "Pseudo": el hecho de utilizar un método deja de ser aleatorio.
- Objetivo: Producir una secuencia de núm. en [*0,1*] que simulan o imitan las propiedades ideales de distribución uniforme e independiente

- Consideraciones importantes para las generadores:
  - Rápido
  - Portable
  - Tener un *ciclo* suficientemente largo
  - Replicable
  - Aproximado a las propiedades estadísticas ideales de uniformidad e independencia.

Propiedades de los Números Aleatorios

- Propiedades que deben cumplir:
  - Uniformidad.
  - Independencia.
  - Analisis de estas propiedades.
- Cada número aleatorio, R_i, debe ser una extracción muestral independiente de una Distr. Uniforme(0,1) con fdp:

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

$$E(R) = \int_0^1 x\,dx = \left.\frac{x^2}{2}\right|_0^1 = \frac{1}{2}$$

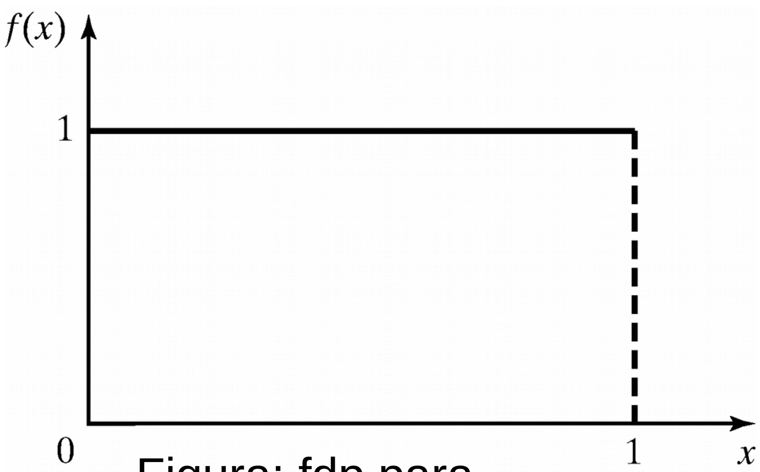


Figura: fdp para núm. aleatorios

Técnicas

- Método de Congruencia Lineal (MCL).
- Generadores Congruentes Lineales Combinados (CLCG).
- Series de Números Aleatorios (Random-Number Streams).

Método de Congruencia Lineal

- Produce una secuencia de enteros, X_1, X_2, … entre *0* y *m-1* a través de la relación recursiva:

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

multiplicador

increment
o

módulo

- La selección de los valores para *a*, *c*, *m*, y X_0 afectan drásticamente las propiedades y la longitud de ciclo.
- Para convertir los enteros en números aleatorios:

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

Ejemplo [MCL]

- Usando $X_0 = 27$, $a = 17$, $c = 43$, and $m = 100$.
- Los valores para X_i y R_i son:

$X_1 = (17*27+43) \bmod 100 = 502 \bmod 100 = 2,\quad R_1 = 0.02;$

$X_2 = (17*2+43) \bmod 100 = 77,\quad R_2 = 0.77;$

$X_3 = (17*77+43) \bmod 100 = 52,\quad R_3 = 0.52;$

…

Otras Características

- Densidad Máxima
  - Los valores R_i, $i = 1,2,\ldots$, no dejan grandes vacíos en $[0,1]$
- Período Máximo
  - Se logra con una apropiada elección de a, c, m, y X_0.
    - *Ejemplos.*
- La mayoría de los lenguajes usan valores para *m* potencia de 2, por eficiencia.

Generadores Congruentes Lineales Combinados (CLCG)

- Razón: generadores de períodos largos son necesitados por el aumento de la complejidad de los sistemas de simulación.
- Método: Combinar 2 o mas GCL Multiplicativos.
- Sea $X_{i,1}$, $X_{i,2}$, …, $X_{i,k}$, las $i^{\text{e-sima}}$ salidas desde k diferentes GCL multiplicativos.
  - Donde el j^{es} generador:
    - Tiene módulo primo m_j, y el multiplicador a_j es elegido t.q. el perído es m_j-1.
    - produce enteros $X_{i,j}$ aprox. ~ Uniforme [1, m-1].
    - $W_{i,j} = X_{i,j}$ -1 es aprox. ~ Uniforme sobre los enteros en [0, m_j-2]

Generadores Congruentes Lineales Combinados (CLCG)

- L'Ecuyer sugiere generadores de la forma:

$$X_i = \left(\sum_{j=1}^{k} (-1)^{j-1} X_{i,j} \right) \bmod m_1 - 1 \qquad \text{Aquí, } R_i = \begin{cases} \dfrac{X_i}{m_1}, & X_i > 0 \\ \dfrac{m_1 - 1}{m_1}, & X_i = 0 \end{cases}$$

- El máximo período posible es:

$$P = \frac{(m_1 - 1)(m_2 - 1)\dots(m_k - 1)}{2^{k-1}}$$

- Se puede obtener facilmente períodos máximos de 2^{62}(incluso hasta 2^{191})

Generadores Congruentes Lineales Combinados (CLCG)

- Ejemplo: para computadoras de 32-bit, L'Ecuyer [1988] sugiere una combinación de $k = 2$ generadores con $m_1 = 2{,}147{,}483{,}563$, $a_1 = 40{,}014$, $m_2 = 2{,}147{,}483{,}399$ and $a_2 = 20{,}692$.
- Algoritmo:

  Step 1: Seleccionar la semilla
  - $X_{1,0}$ en el rango[*1, 2,147,483,562] para el* 1er generador
  - $X_{2,0}$ en el rango [*1, 2,147,483,398]* para el 2do generador.

  Step 2: Por cada generador individual,

  $X_{1,j+1} = 40{,}014\ X_{1,j} \bmod 2{,}147{,}483{,}563$

  $X_{2,j+1} = 20{,}692\ X_{1,j} \bmod 2{,}147{,}483{,}399$.

  Step 3: $X_{j+1} = (X_{1,j+1} - X_{2,j+1}) \bmod 2{,}147{,}483{,}562$.

  Step 4: Return

$$R_{j+1} = \begin{cases} \dfrac{X_{j+1}}{2{,}147{,}483{,}563}, & X_{j+1} > 0 \\[2ex] \dfrac{2{,}147{,}483{,}562}{2{,}147{,}483{,}563}, & X_{j+1} = 0 \end{cases}$$

  Step 5: Setear $j = j+1$, ir a step 2.

  - Generador Combinado tiene período: $(m_1 - 1)(m_2 - 1)/2 \sim 2 \times 10^{18}$

Series de Números Aleatorios (Random-Numbers Streams)

- La semilla GCL:
  - Es el valor inicial X_0 que inicializa la secuencia de núm. aleatorios.
  - Cualquier valor de la secuencia puede ser usado para “sembrar” el generador.
- Una serie de Núm. Aleatorios (NA) es:
  - Referirse a una semilla “inicial” tomada desde la secuencia $X_0, X_1, \ldots, X_P$.
  - Si estas semillas están separadas por b valores, entonces el stream i debería estar definido por la semilla inicial:

$$S_i = X_{b(i-1)}$$

  - Algunos $b = 10^5$ … hasta $b = 10^{37}$.
- Un simple generador con k streams puede actuar como k distintos generadores virtuales.
- Sirve para comparar dos o mas alternativas de un sistema, por ej..

Test para Números Aleatorios

- Dos categorias:
  - Testing para la uniformidad:

    H_0: $R_i \sim U[0,1]$

    H_1: $R_i \not\sim U[0,1]$

    - Fracaso en el rechazo de la hipotesis nula, H_0, significa que no se ha detectado evidencia de no-uniformidad.
  - Testing para la independencia:

    H_0: $R_i \sim$ independiente

    H_1: $R_i \not\sim$ independiente

    - Fracaso en rechazar la hipotesis nula, H_0, sign. Significa que evidencia de dependencia no se ha detectado.
- Establecer un nivel de confianza α, la probabilidad de rechazar H_0 cuando éste es verdadera:

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

Valores frecuentes: $\alpha = 0{,}01$ *o* $0{,}05$

Test para Números Aleatorios

- Cuando deben ser usados:
  - Para generadores de los lenguajes de simulación, no son necesarios.
  - Para generadores de lenguajes de propósito general, planillas de calculo, etc., los tests debe ser aplicados.

- Tipos de tests:
  - Tests teóricos.
  - Tests Empiricos (énfasis en estos).

Tests de Uniformidad

- Testeo de Frecuencia:
  - Dos diferentes métodos:
    - Test Kolmogorov-Smirnov
    - Test Chi-square

Test Kolmogorov-Smirnov

- Compara la fda continua , $F(x)$, de la distr. Uniforme con la fda empírica, $S_N(x)$, de una muestra de N observaciones.
  - Conocemos: $F(x) = x, \quad 0 \leq x \leq 1$
  - Si la muestra es es $R_1, R_2, \ldots, R_N$, entonces la fda empirica, $S_N(x)$ es:

$$S_N(x) = \frac{\#\text{ of } R_1, R_2, \ldots, R_n \text{ which are } \leq x}{N}$$

- El test está basado sobre el estadis.: $D = max| F(x) - S_N(x)|$
  - La distribución muestral de D es conocida (Kolmogorov-Smirnov) y es tabulada en función de N y α.

Test Kolmogorov-Smirnov

- Ejemplo: números generados *0.44, 0.81, 0.14, 0.05, 0.93*.

Step 1:

$R_{(i)}$	0.05	0.14	0.44	0.81	0.93
i/N	0.20	0.40	0.60	0.80	1.00
$i/N - R_{(i)}$	0.15	0.26	0.16	-	0.07
$R_{(i)} - (i-1)/N$	0.05	-	0.04	0.21	0.13

Step 2: (rows $i/N - R_{(i)}$ and $R_{(i)} - (i-1)/N$)

Ordeno $R_{(i)}$ de menor a mayor

$D^+ = max\ \{i/N - R_{(i)}\}$

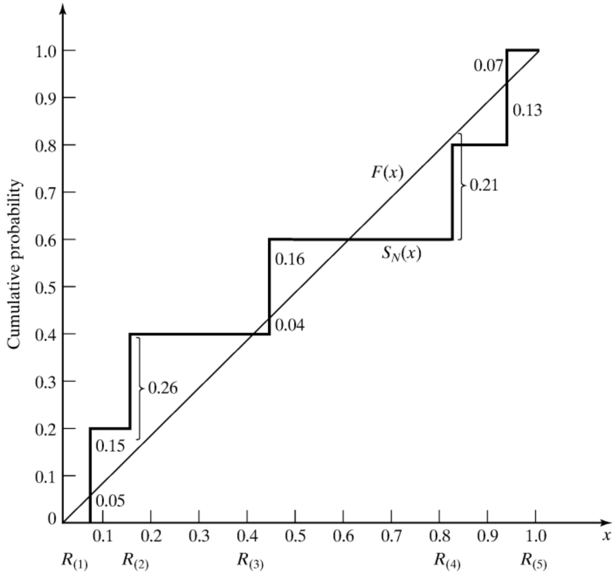
$D^- = max\ \{R_{(i)} - (i-1)/N\}$

Step 3: $D = max(D^+, D^-) = 0.26$

Step 4: For $\alpha = 0.05$,

$D_\alpha = 0.565 > D$

H_0 **no es rechazada.**



Test Chi-square

- Test Chi-square test usa la estadística muestral:

n es el # de clases

$$X_0^2 = \sum_{i=1}^{n} \frac{(O_i - E_i)^2}{E_i}$$

E_i es el # esperado en la i^{th} clase

O_i es el # observaciones en la i^{th} clase

  - Es aproximadamente una distr. chi-square con *n-1* grados de libertad (la cual es tabulada)
  - Para la distr. Uniforme, E_i, el número esperado en cada clase es:

$$E_i = \frac{N}{n}, \quad \text{dende N es el \# total de observaciones}$$

- Valido solo para grande muestras (N >= 50)

Tests para Autocorrelación

- Testea la autocorrelación entre cada m números, comenzando desde un i^{es} número.
  - La autocorrelación ρ_{im} entre numeros: $R_i, R_{i+m}, R_{i+2m}, R_{i+(M+1)m}$
  - M es el mas grande entero t.q.

$$i+(M+1)m \leq N$$

- Hipótesis:

$$H_0: \quad \rho_{im} = 0, \qquad \text{si los numeros son independientes}$$
$$H_1: \quad \rho_{im} \neq 0, \qquad \text{si los numeros son dependientes}$$

- Si los valores son independientes:
  - Para un M grande, la distr. del estimador de ρ_{im}, denotado $\hat{\rho}_{im}$ es aproximadamente una normal.

Tests para Autocorrelación

- El estadístico de la prueba puede ser formado como:

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

  - Z_0 es distr. normalmente con media = *0* y varianza = *1*, y:

$$\hat{\rho}_{im} = \frac{1}{M+1}\left[\sum_{k=0}^{M} R_{i+km} R_{i+(k+1)m}\right] - 0.25$$

$$\hat{\sigma}_{\rho_{im}} = \frac{\sqrt{13M+7}}{12(M+1)}$$

- La hipotesis nula no es rechaza si: $-Z_{\alpha/2} \leq Z \leq Z_{\alpha/2}$.

Ejemplo

- Test entre los *3rd, 8th, 13th*, y asi, para las observaciones de la pag. 291 (libro Jerry Banks).
  - Donde, α = *0.05, i = 3, m = 5, N = 30, y* M = 4

$$\hat{\rho}_{35} = \frac{1}{4+1}\left[\begin{array}{l}(0.23)(0.28) + (0.28)(0.33) + (0.33)(0.27) \\ + (0.27)(0.05) + (0.05)(0.36)\end{array}\right] - 0.25$$

$$= -0.1945$$

$$\hat{\sigma}_{\rho_{35}} = \frac{\sqrt{13(4)+7}}{12(4+1)} = 0.128$$

$$Z_0 = -\frac{0.1945}{0.1280} = -1.516$$

  - Observando la tabulación (Tabla A.3 del libro), $z_{\alpha/2}$ $(z_{0.025})$ = *1.96*. , la hipótesis no es rechazada.

Referencias

- RIPLEY, B.D. [1987], *Stochastic Simulation*, Wiley, New York.
- L'ECUYER, P. [1998], *Random Number Generation*, Chapter 4 in *Handbook of Simulation*, J. banks, ed., pp.93-137. Wiley, New York.
- L'ECUYER, P. [1988], *Efficient and Portable Combines Random Number Generators*, Communication of the ACM, Vol. 31, pp. 742-749, 774.
- **Jerry Banks, John S. Carson, II, Barry L. Nelson, *Discrete-Event System Simulation*, David M. Nicol. Quinta Edición. ISBN-10: 0136062121. Publisher: Prentice Hall.**