Generación de Números y Variables Aleatorios

Generación de Números Aleatorios

La generación de números aleatorios son un ingrediente fundamental en la simulación de la mayoría de sistemas discretos.

La mayoría de lenguajes de computadora tienen subrutinas, objetos o funciones que generan números aleatorios.

Propiedades de los números aleatorios

Una secuencia de números aleatorios $R_1, R_2, ...$, debe tener dos propiedades estadísticas importantes:

1. **Uniformidad**, es decir, son igualmente probables en cualquier parte.
2. **Independencia**, es decir, el valor actual de una variable aleatorio no tiene relación con el valor anterior.

Cada numero aleatorio R_1 es una muestra independiente extraída de una distribución uniforme continua entre cero y uno.

Función de Distribución de Probabilidad

$$f(x) = \begin{cases} 1, & 0 \leqslant x \leqslant 1 \\ 0, & \text{de lo contrario} \end{cases}$$

Valor esperado

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

Varianza

$$V(R) = \int_0^1 x^2 dx - [E(R)]^2 = \left.\frac{x^3}{3}\right|_0^1 - \left(\frac{1}{2}\right)^2 = \frac{1}{3} - \frac{1}{4} = \frac{1}{12}$$

Algunas consecuencias de las propiedades de uniformidad e independencia:

1. Si el intervalo $(0, 1)$ es dividido en n sub-intervalos de igual longitud, el numero esperado de observaciones de cada intervalo es N/n donde N es el número total de observaciones. Nótese que N tiene que ser lo suficientemente largo para mostrar esta tendencia.
2. La probabilidad de observar un valor en un intervalo particular es independiente del valor previo dibujados.

Generación de Números Pseudo-aleatorios

El mero acto de utilizar un *método conocido* para generar números aleatorios remueve el potencial de una aleatoriedad verdadera.

Si el método es conocido, el conjunto de números aleatorios puede ser repetido, por lo que se puede decir que los números no son verdaderamente aleatorios.

Generar números verdaderamente aleatorios requiere acceso a un proceso físico aleatorio, lo que puede ser poco practico para muchos usos computacionales. Los algoritmos para generar números pseudo-aleatorios, por otro lado, pueden producir números a una velocidad mucho mayor.

El objetivo de cualquier esquema de generación es, sin embargo, producir una secuencia de números entre 0 y 1 que simule o imite las propiedades ideales de distribución uniforme e independencia lo más cercano posible.

En una simulación por computadora, a menudo nos interesa tener números pseudo-aleatorios, porque nos permiten tener el control de los números aleatorios para poder **repetir** el experimento.

En general, se utiliza una forma sistemática de generar números pseudo-aleatorios, estos números son completamente predecibles si se conoce el algoritmo y el valor inicial (semilla).

Primero generamos números aleatorios uniformemente distribuidos, luego los utilizamos para generar números aleatorios en otra distribución.

Los métodos utilizados por una computadora para generar números pseudo-aleatorios, deben tener en cuenta las siguientes consideraciones:

- El método debe ser rápido.
- El método debe ser portable a través de plataformas de hardware y lenguajes de programación.
- El método debe tener un ciclo suficientemente largo.
  - La longitud de un ciclo representa la longitud de la secuencia de números aleatorios antes de que los números anteriores empiecen a repetirse en un orden anterior. Por ejemplo $4, 9, 5, 6, 9, 3, 8, 4, 9, 5, 6, 9, 3, 8, 4, 9, 5, 6, 9, 3, 8, ...$ parece tener una longitud de ciclo de 7 (esto es sólo un ejemplo de ciclo, ¡un número aleatorio de ciclo 7 es completamente inaceptable!)
  - Un caso especial de ciclo es la degeneración, en la que los mismos números aleatorios aparecen repetidamente.
  - Como utilizamos un algoritmo para generar números aleatorios, los ciclos no se pueden evitar. Pero los ciclos largos (por ejemplo, algunos millones o algunos miles de millones) sirven al propósito de las simulaciones generales.
- Los números aleatorios deben poder replicarse.
- Y lo que es más importante, los números aleatorios generados deben aproximarse mucho a las propiedades estadísticas ideales de uniformidad e

independencia.

Técnicas para generar números aleatorios

Método de Generación Congruencial Lineal (GCL) El método congruencial lineal produce una secuencia de enteros $X_1, X_2, X_3, ...$ entre 0 y $m-1$ de acuerdo a la siguiente relación de recurrencia:

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

- Al valor inicial de X_0 se le llama semilla.
- a es el multiplicador.
- c es el incremento.
- m es el módulo.

La selección de a,c,m y X_0 afecta drásticamente las propiedades estadísticas como al media y la varianza, así como la longitud de ciclo.

Cuando $c \neq 0$, entonces se le llama *método congruencial mixto*, cuando $= 0$, se conoce como *método congruencial multiplicativo*

Para la generación de números aleatorios (no enteros) entre 0 y 1 podemos utilizar la siguiente ecuación:

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

Ejemplo 1 Utilice el método congruencial lineal para generar una secuencia de números aleatorios con $X_0 = 27$, $a = 17$, $c = 43$ y $m = 100$.

$$X_0 = 27$$
$$X_1 = (17 \cdot 27 + 43) \bmod 100 = 502 \bmod 100 = 2$$
$$R_1 = \frac{2}{100} = 0.02$$
$$X_2 = (17 \cdot 2 + 43) \bmod 100 = 77 \bmod 100 = 77$$
$$R_2 = \frac{77}{100} = 0.77$$
$$X_3 = (17 \cdot 77 + 43) \bmod 100 = 1352 \bmod 100 = 52$$
$$R_3 = \frac{52}{100} = 0.52$$

Aspectos a Considerar

- Los números $(R1, R2, ...)$ generados en el ejemplo unicamente pueden asumir valores del conjunto $I = 0, 1/m, 2/m, ..., (m-1)/m$.
- Es decir cada Ri es discreto en I, en lugar de ser continuo en el intervalo $[0, 1]$.
- Si m esto no es mayor problema. Valores de $m = 2^{31} - 1$ y $m = 2^{48}$ son utilizados comúnmente.
- La *densidad máxima* es una medida de que tan "densamente" los valores asumidos por $Ri, i = 1, 2...$ llenan los espacios entre $[0, 1]$. El *período máximo* es la longitud de la secuencia antes de que los números comiencen a repetirse.
- Para lograr una *densidad máxima* para un rango determinado, una elección apropiada de a, c, m y X_0 es muy importante. El *período máximo* puede ser alcanzado mediante alguna selección comprobada de estos valores.
  - Para m una potencia de 2, i.e. $m = 2^b$, y $c \neq 0$, el período más largo posible es $P = m = 2^b$, cuando c es un primo relativo a m y $a = 1 + 4k$, donde k es un entero.
  - Para m una potencia de 2, i.e. $m = 2^b$, y $c = 0$, el período más largo posible es $P = m/4 = 2^{b-2}$, el cual es alcanzado si la semilla X_0 es impar y si el multiplicador a esta dado por $a = 3 + 8k$ o $a = 5 + 8k$, donde k es un entero.
  - Para m un número primo y $c = 0$, el mayor periodo posible es $P = m - 1$ donde a satisface la propiedad que el menor valor de k para el cual $a^k - 1$ es divisible por m es $k = m - 1$.

Ejercicio 1

- A partir del ejemplo anterior ($X_0 = 27$, $a = 17$, $c = 43$ y $m = 100$), genere 100 números pseudo-aleatorios mediante el método congruencial lineal. Puede utilizar una hoja de calculo o cualquier lenguaje de computadora.

- ¿Cuál es el período de dicha secuencia de números pseudo-aleatorios? ¿Cual es la media y la varianza? ¿que pasa si genera 1000? ¿Qué pasa si cambia el valor semilla? Comente sus respuestas.

- Genere gráficos de tipo histograma y de dispersion (scatter-plot) con los valores obtenidos.

- Comente sus observaciones con respecto a las propiedades de Uniformidad y Densidad de la secuencia de números generados. ¿Cree que es adecuado para aplicarlo a un modelo se simulación?

Ejercicio 2 Repita el ejercicio anterior para el generador congruencial multiplicativo con $a = 13$, $m = 2^6 = 64$ y $X_0 = 1, 2, 3, 4$

Ejercicio 3 Investigue valores comúnmente utilizados para generadores congruencial lineales, repita el ejercicio para estos valores investigados.

Ejercicio 4 Utilice el siguiente generador congruencial lineal de números aleatorios:

$d_i = 16807 d_{i-1} mod(2^3 1 - 1)$

- Cuántos bits son necesarios para la multiplicación mas larga posible?
- Cuales son las implicaciones de este generador en una computadora con aritmética de 32 bits?

Generadores congruentes lineales combinados

Combinando dos o más generadores congruentes multiplicativos se puede aumentar la longitud del período y se obtienen otras estadísticas mejores. Véase el Ejemplo 8.5 en la página 297.

Generación de Variables Aleatorias

Ahora que hemos aprendido a generar una variable aleatoria uniformemente distribuida, estudiaremos cómo producir variables aleatorias de otra distribución utilizando la variable aleatoria uniformemente distribuida.

Las técnicas analizadas incluyen la transformación inversa y la convolución. También se discute la técnica de aceptación-rechazo.

Todo el trabajo aquí supone la existencia de una fuente de números aleatorios uniformes (0,1), $R_1, R_2, ...$

Función de Densidad de Probabilidad

$$f_R(x) = \begin{cases} 1, & 0 \leqslant x \leqslant 1 \\ 0, & \text{de lo contrario} \end{cases}$$

Función de Probabilidad Acumulada

$$F_R(x) \begin{cases} 0, & x < 0 \\ x, & 0 \leqslant x \leqslant 1 \\ 1, & x > 1 \end{cases}$$

Técnica de la transformación inversa

- La técnica de la transformación inversa puede utilizarse para tomar muestras de las distribuciones exponenciales, uniformes, Weibulll y triangulares.
- El principio básico es encontrar la función inversa de F, F^{-1} tal que $FF^{-1} = F^{-1}F = 1$
- F^{-1} denota la solución a la ecuación $r = F(x)$ en términos de r, no $1/F$.
- Por ejemplo, la inversa de $y = x$ es $x = y$, la inversa de $y = 2x + 1$ es $x = (y - 1)/2$, la inversa de $y = x^2 - 1$ es $x = \sqrt{y + 1}$

Generación de una Distribución Exponencial

Recordemos la distribución exponencial

- Función de densidad de probabilidad:

$$f(x) = \begin{cases} \lambda e^{-\lambda x} & , x \geqslant 0 \\ 0 & , x < 0 \end{cases}$$

- Función de probabilidad acumulada

$$F(x) = \int_{-\infty}^{x} f(t)dt = \begin{cases} 1 - e^{-\lambda x} & , x \geqslant 0 \\ 0 & , x < 0 \end{cases}$$

Este método puede ser utilizado para cual distribución en teoría. Pero es particularmente util para variables aleatorias cuyo función inversa puede ser resuelta con facilidad.

Los pasos son los siguientes:

Paso 1:

- Obtenga la función de probabilidad acumulada para la variable aleatoria X.
- Para la distribución exponencial, la función de probabilidad acumulada es $F(X) = 1 - e^{-\lambda x}$

Paso 2

- Definir $R = F(X)$ en el rango de X
- Para la distribución exponencial, $R = 1 - e^{-\lambda x}$ en el rango de $x \geqslant 0$.

Paso 3

- Resolver la ecuación $F(X) = R$ para X en términos de R.
- Para la distribución exponencial, la solución es la siguiente:

$$\begin{aligned} 1 - e^{-\lambda X} &= R \\ e^{-\lambda X} &= 1 - R \\ -\lambda X &= \ln(1 - R) \\ X &= -\frac{1}{\lambda} \ln(1 - R) \end{aligned}$$

A la ultima ecuación se le llama generador de la variable aleatoria para la distribución exponencial. En general se escribe como $X = F^{-1}(R)$.

Paso 4

Generar (según sea necesario) los números aleatorios $R_1, R_2, ...$ y calcular las variables aleatorias deseadas mediante.

$$X_i = F^{-1}(R_i)$$

En el caso de la distribución exponencial

$$X_i = \frac{-1}{\lambda} \ln{(1 - R_i)}$$

para $i = 1, 2, 3...$ donde R_i es un numero aleatorio uniformemente distribuido entre $(0,1)

En la práctica, dado que tanto R_i como $1 - R_i$, son números aleatorios uniformemente distribuidos, el calculo puede simplificarse según:

$$X_i = \frac{-1}{\lambda} \ln{(R_i)}$$

Una vez tenemos este procedimiento establecido, podemos proceder a resolver otras distribuciones similares para las cuales la función inversa es relativamente fácil de obtener y tiene una formula cerrada.

Generación de una Distribución Uniforme

Si X es una variable aleatoria uniformemente distribuida entre el intervalo $[a, b]$, la función para generar X esta dada por $X = a + (b - a)R$.Veamos el paso a paso:

La función de densidad de probabilidad de X esta dada por:

$$f(x) = \begin{cases} \frac{1}{b-a} & a \leq x \leq b \\ 0 & \text{de lo contrario} \end{cases}$$

Paso 1: La función de densidad acumulada es:

$$F(x) = \begin{cases} 0 & x < a \\ \frac{x-a}{b-a} & a \leq x \leq b \\ 1 & x > b \end{cases}$$

Paso 2: Definimos $F(X) = (X - a)/(b - a) = R$

Paso 3: Resolvemos X en términos de R obtenemos

$$X = a + (b-a)R$$

Paso 4: Generamos R_i según sea necesario, calculamos X_i utilizando la función obtenida.

Generación de una Distribución Weibull

La función de densidad de probabilidad es:

$$f(x) = \begin{cases} \frac{\beta}{\alpha^\beta} x^{\beta-1} e^{-(x/\alpha)^\beta}, & x \geq 0 \\ 0, & \text{otherwise} \end{cases}$$

Paso 1

Obtenemos la función de probabilidad acumulada:

$$F(x) = 1 - e^{-(x/\alpha)^\beta}$$

para $x \geqslant 0$

Paso 2

Sea $F(x) = 1 - e^{-(x/\alpha)^\beta} = R$

Paso 3

Resolvemos para X en terminos de R, obtenemos:

$$X = \alpha[-\ln(1-R)]^{1/\beta}$$

Paso 4

Generar valores de R_i uniformemente distribuidos entre $(0,1)$, para calcular valores de X, según la función del paso 3.

Generación de una Distribución Triangular

Función de densidad de probabilidad

$$f(x) = \begin{cases} x & 0 \leq x \leq 1 \\ 2-x & 1 < x \leq 2 \\ 0 & \text{otherwise} \end{cases}$$

Paso 1

Función de probabilidad acumulada

$$F(x) = \begin{cases} 0 & x \leq 0 \\ \frac{x^2}{2} & 0 < x \leq 1 \\ 1 - \frac{(2-x)^2}{2} & 1 < x \leq 2 \\ 1 & x > 2 \end{cases}$$

Paso 2

sea $R = \frac{X^2}{2}$ para $0 < X \leq 1$ y $R = 1 - \frac{(2-X)^2}{2}$ para $1 < x \leq 2$

Paso 3

Resolver X en términos de R

$$X = \begin{cases} \sqrt{2R} & 0 < R \leq \frac{1}{2} \\ 2 - \sqrt{2(1-R)} & \frac{1}{2} < R \leq 1 \end{cases}$$

Paso 4

Generar valores de R_i uniformemente distribuidos entre $(0, 1)$, para calcular valores de X, según la función del paso 3.

Ejercicio 5

- Mediante una hoja de calculo o cualquier lenguaje de programación, genere una secuencia de 1000 números pseudo-aleatorios mediante un generador congruencial lineal con los siguientes parámetros $m = 2^{32}$, $a = 1664525$, $c = 1013904223$, $X0 = 123456789$.
- A partir de dicha secuencia, genere variables aleatorias con una distribución exponencial, con un valor de $\lambda = 1$
- Genere gráficos tipo histograma tanto de la secuencia de número aleatorios como de la variable aleatoria calculada.
- Calcule la media y desviación estándar de los valores calculados para la variable aleatoria.
- Compare los histogramas empíricos y los valores estadísticos con los valores teóricos para una distribución exponencial.

Ejercicio 6

- Mediante una hoja de calculo o cualquier lenguaje de programación, genere una secuencia de 1000 números pseudo-aleatorios mediante un generador congruencial lineal con los siguientes parámetros $m = 2^{32}$, $a = 1664525$, $c = 1013904223$, $X0 = 123456789$.
- A partir de dicha secuencia, genere variables aleatorias con una distribución uniforme entre 0 y 5.

- Genere gráficos tipo histograma tanto de la secuencia de número aleatorios como de la variable aleatoria calculada.
- Calcule la media y desviación estándar de los valores calculados para la variable aleatoria.
- Compare los histogramas empíricos y los valores estadísticos con los valores teóricos para una distribución exponencial.