

Unidad III: Generación de variables aleatorias

3.1 Conceptos básicos

Un generador de números aleatorios es un dispositivo informático o físico diseñado para producir secuencias de números sin un orden aparente.

Algoritmos

Los algoritmos para la generación de valores uniformemente distribuidos están presentes en todas las calculadoras y lenguajes de programación, y suelen estar basados en congruencias numéricas del tipo:

$$x_{n+1} \equiv (ax_n + c) \pmod{m}$$

El éxito de este tipo de generadores de valores de una variable aleatoria depende de la elección de los cuatro parámetros que intervienen inicialmente en la expresión anterior:

El valor inicial o semilla: x_0

La constante multiplicativa: a

La constante aditiva: c

El número m respecto al cual se calculan los restos

Estos cuatro valores deben ser números enteros no negativos y que cumplan la siguiente condición: $x_0, a, c < m$.

La mayor parte de los generadores de números aleatorios son, en realidad, pseudoaleatorios; se calcula (o introduce internamente) un valor x_0 , que llamaremos semilla, y, a partir de él, se van generando $x_1, x_2, x_3, \dots$

Siempre que se parta de la misma semilla, se obtendrá la misma secuencia de valores.

Por la condición anterior, es evidente que todos los valores generados por este procedimiento son números enteros entre 0 y $m-1$. El número máximo de cifras distintas que pueden obtenerse con el procedimiento descrito es m , así que llegará un momento en que el primer número generado se repetirá produciéndose un ciclo.

El ciclo dónde inevitablemente caerá el generador interesa que sea de la mayor longitud posible (como máximo m), para evitar que se repitan pronto los valores aleatorios. Por ejemplo, para los valores $a = 3, c = 5, x_0 = 2$ y $m = 32$ se obtiene la siguiente secuencia de valores:

2-11-6-23-10-3-14-15-18-27-22-7-26-19-30-31-2-11-6

La secuencia generada tiene como longitud 16 números (el número generado en la decimoséptima posición es el 2 inicial, por lo que toda la secuencia se repite a partir de ahí), muy inferior a la longitud máxima que podría tener ($m=32$). Determinadas elecciones de parámetros del generador (x_0, a, c y m) conducen a ciclos de amplitud máxima.

Si $c \neq 0$:

$$m.c.m.(c, m) = 1$$

Rectificación $m.c.d.(c, m) = 1$ Máximo común divisor

$$a \equiv 1 \pmod{p} \text{ para cada primo } p \text{ de } m$$

$$a \equiv 1 \pmod{4} \text{ si } 4 \text{ es divisor de } m$$

Si $c=0$:

m es primo

$a^{m-1/p} \equiv 1 \pmod{m}$ para cada factor primo p de $m-1$ (Rectificación: la condición es que NO SEA congruente para cada factor primo p de $m-1$).

Por ejemplo, tomando como valores $m = 2^5 = 32$, $a = 5$, $x_0 = 1$ y $c = 3$ se obtiene la siguiente secuencia de números, que tiene longitud máxima:

1-8-11-26-5-28-15-14-9-16-19-2-13-4-23-22-17-24-27-10-21-12-31-30-25-0-3-18-29-20-7-6-1

3.2 Variables aleatorias discretas

Variable aleatoria discreta: una V.A. es discreta si su recorrido es un conjunto discreto.

En gran cantidad de experimentos aleatorios es necesario cuantificar los resultados, es decir, asignar a cada resultado del experimento un número, con el fin de poder realizar un estudio matemático.

3.3 Variables aleatorias continuas

Es aquella que puede tomar infinitos valores dentro de un intervalo de la recta real.

En el caso de variables aleatorias continuas no tiene sentido plantearse probabilidades de resultados aislados. La probabilidad de valores puntuales es cero.

El interés de estas probabilidades está en conocer la probabilidad correspondiente a un intervalo.

Dicha probabilidad se conoce mediante una curva llamada función de densidad y suponiendo que bajo dicha curva hay un área de una unidad.

Conociendo esta curva, basta calcular el área correspondiente para conocer la probabilidad de un intervalo cualquiera.

3.4 Métodos para generar variables aleatorias

Existen varios métodos que nos permiten generar variables aleatorias. Lo normal es que existan varias opciones para generar una misma variable aleatoria. La elección del método adecuado se puede basar en una serie de factores como:

Exactitud. se prefiere un método exacto frente a métodos aproximados, como soluciones numéricas.

Velocidad. Uno de los datos que se toma en consideración es el tiempo de generación de la variable.

Espacio. Necesidades de memoria del método utilizado. En general, los métodos no consumen mucha memoria.

3.4.1 Método de la transformada inversa

El método de la transformada (o transformación) inversa, también conocido como método de la inversa de la transformada, es un método para la generación de números aleatorios de cualquier distribución de probabilidad continua cuando se conoce la inversa de su función de distribución (cdf). Este método es en general aplicable, pero puede resultar muy complicado obtener una expresión analítica de la inversa para algunas distribuciones de probabilidad. El método de Box-Muller es un ejemplo de algoritmo que aunque menos general, es más eficiente desde el punto de vista computacional.

El método se utiliza para simular valores de las distribuciones exponencial, Cauchy, triangular, de Pareto y Weibull.

El problema que resuelve el método de la transformada inversa es el siguiente:

- Sea X una variable aleatoria cuya distribución puede ser descrita por la cdf F .
- Se desea generar valores de X que están distribuidos según dicha distribución.

Numerosos lenguajes de programación poseen la capacidad de generar números pseudo-aleatorios que se encuentran distribuidos de acuerdo con una distribución uniforme standard. Si una variable aleatoria posee ese tipo de distribución, entonces la probabilidad de que el número caiga dentro de cualquier subintervalo (a, b) del intervalo entre 0 a 1 es la longitud del subintervalo, o sea $b - a$.

El método de la transformada inversa funciona de la siguiente manera:

1. Se genera un número aleatorio a partir de la distribución uniforme standard; se lo llama u .
2. Se calcula el valor x tal que $F(x) = u$; y se lo llama x_{elegido} .
3. Se toma x_{elegido} como el número aleatorio extraído de la distribución caracterizada por F .

3.4.2 Método de convolución

Convolución es un operador matemático que transforma dos funciones f y g en una tercera función que en cierto sentido representa la magnitud en la que se superponen f y una versión trasladada e invertida de g . Una convolución es un tipo muy general de media móvil, como se puede observar si una de las funciones se toma como la función característica de un intervalo.

Convolución discreta

Cuando se trata de hacer un procesamiento digital de señal no tiene sentido hablar de convoluciones aplicando estrictamente la definición ya que solo disponemos de valores en instantes discretos de tiempo. Es necesario, pues, una aproximación numérica. Para realizar la convolución entre dos señales, se evaluará el área de la función $x(\tau) * h(t - \tau)$. Para ello, disponemos de

muestreos de ambas señales en los instantes de tiempo nt , que llamaremos $x[k]$ y $h[n - k]$ (donde n y k son enteros). El área es, por tanto,

$$y[n] = \sum_{k=-\infty}^{\infty} t \cdot x[k] \cdot h[n - k] = t \cdot \left[\sum_{k=-\infty}^{\infty} x[k] \cdot h[n - k] \right]$$

3.4.3 Método de composición

Determinar cuál era el **método de composición de Wolfgang Amadeus Mozart** ha sido una cuestión muy estudiada. La visión decimonónica de este tema se basaba a menudo en una concepción romántica e idealizada del proceso de composición; los estudios más recientes han tratado de abordar el asunto a través del examen sistemático de las cartas y documentos que han sobrevivido, llegando a diversas conclusiones.

3.5 Procedimientos especiales

Existen diferentes tipos de métodos para generar variables aleatorias, pero también existen casos especiales para generar estas los cuales son:

La distribución de Poisson parte de la distribución binomial:

Cuando en una distribución binomial se realiza el experimento un número " n " muy elevado de veces y la probabilidad de éxito " p " en cada ensayo es reducida, entonces se aplica el modelo de distribución de Poisson:

Se tiene que cumplir que:

- " p " < 0,10
- " $p \cdot n$ " < 10

La distribución de Poisson sigue el siguiente modelo:

Vamos a explicarla:

El número "e" es 2,71828

"l" = $n * p$ (es decir, el número de veces "n" que se realiza el experimento multiplicado por la probabilidad "p" de éxito en cada ensayo)

"k" es el número de éxito cuya probabilidad se está calculando

La distribución binomial parte de la distribución de Bernouilli:

La distribución de Bernouilli se aplica cuando se realiza una sola vez un experimento que tiene únicamente dos posibles resultados (éxito o fracaso), por lo que la variable sólo puede tomar dos valores: el 1 y el 0

La distribución binomial se aplica cuando se realizan un número "n" de veces el experimento de Bernouilli, siendo cada ensayo independiente del anterior. La variable puede tomar valores entre:

0: si todos los experimentos han sido fracaso

n: si todos los experimentos han sido éxitos

3.6 Pruebas estadística. (Pruebas de bondad de ajuste)

En la construcción del modelo de simulación es importante decidir si un conjunto de datos se ajusta apropiadamente a una distribución específica de probabilidad. Al probar la bondad del ajuste de un conjunto de datos, se comparan las frecuencias observadas FO realmente en cada categoría o intervalo de clase con las frecuencias esperadas teóricamente FE.

Prueba Ji cuadrada

La prueba Ji cuadrada hace uso de la distribución del mismo nombre para probar la bondad del ajuste al comparar el estadístico de prueba χ^2 con el valor en

tablas de la mencionada distribución Ji cuadrada con v grados de libertad y un nivel de significancia α .