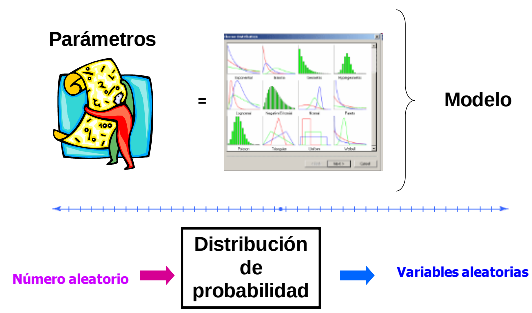
UNIDAD 3

GENERACION DE VARIABLES ALEATORIAS

Uso de parámetros en modelos

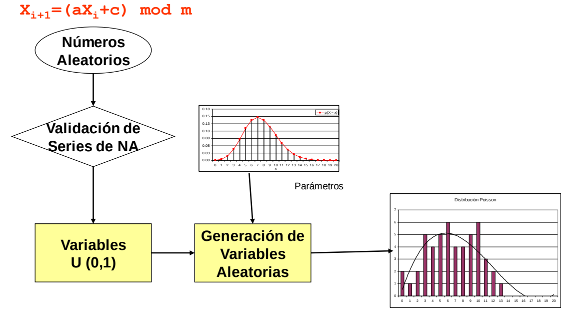


La generación de variables aleatorias

- Significa la obtención de variables que siguen una distribución de probabilidad determinada.
- La generación de variables estocásticas o aleatorias requiere de dos etapas:
  1. Generar números aleatorios distribuidos uniformemente (R).
  2. Generar con R y con las distribuciones de probabilidad las variables aleatorias o estocásticas.

Procedimiento

$X_{i+1}=(aX_i+c) \bmod m$



UTN SANTA FE

Propiedades de números aleatorios

- **Independencia:** la probabilidad de observar un valor en un intervalo en particular es independiente de los valores previamente observados.
- **Uniformidad:** cada número aleatorio Ri es una muestra independiente tomada de una distribución continua uniforme entre cero y uno:

f(x)=1/(b-a) para a=0 y b=1 positivos

UTN SANTA FE

Propiedades de los números aleatorios

- Si el intervalo (0, 1) es dividido en n clases, o sub intervalos de longitudes iguales, el número esperado de observaciones en cada intervalo es N/n, donde N es el número total de observaciones.

UTN SANTA FE

Obtención de números aleatorios

- Algunos métodos usan dispositivos físicos para generar números con igual probabilidad de ocurrencia (tirada de una moneda, de un dado, giro de un disco, etc.).
- Ciertos dispositivos físicos sofisticados asocian señales que se generan por algún fenómeno con números que tienen la particularidad de formar una serie de largo ciclo en donde ninguno de los números generados se repite.

UTN SANTA FE

Generación de Números Aleatorios

- Generación con dispositivos de Hardware o físicos:
  - Cuenta con fuentes externas como por ejemplo los tiempos de desgaste de un material radiactivo (Carbono 14) o los ruidos electrónicos (blanco, gaussiano) de dispositivos para generar números aleatorios, etc.

UTN SANTA FE

Generación de Números Aleatorios

- Métodos físicos $\Rightarrow$ costoso y poco práctico.
- Alternativa: técnicas numéricas
- Desventaja: no son números aleatorios puros por lo que se los conoce como números pseudoaleatorios.
- **Técnicas numéricas:** generadores de números aleatorios, que no reproducen rigurosamente los números tal como lo hacen las máquinas, pero son fáciles de implementar y programar.

UTN SANTA FE

Generación de Números Aleatorios

- **Definición intuitiva:** Una sucesión de números aleatorios puros, se caracteriza por que no existe ninguna regla o plan que nos permita conocer sus valores.
- Los números aleatorios obtenidos a través de algoritmos recursivos se llaman pseudo-aleatorios (pero que en el uso los llamaremos aleatorios).

UTN SANTA FE

Características de los generadores

- Los números deben estar uniformemente distribuidos entre 0 y 1.
- Deben ser independientes uno de otro (no debe existir correlación).
- Deben tener ciclos largos antes que la secuencia se repita. La longitud de un ciclo es el número de números aleatorios en una secuencia antes de la repetición de la próxima secuencia similar.
- El generador debe ser rápido y no requerir excesiva memoria de cálculo.

UTN SANTA FE

Técnicas de generación de números aleatorios

- Método de cuadrados medios
- Método congruencial lineal
- Método de generadores congruenciales lineales combinados.

Método de cuadrados medios

1. Seleccionar un entero de n dígitos denominado semilla (seed) S0.
2. Encontrar el cuadrado del número: si el número de dígitos del resultado es menor que 2n, se agregan ceros a la izquierda del número hasta completar 2n dígitos.
3. Tomar los n dígitos centrales del número calculado en 2.

Método de cuadrados medios

4. Situar un punto o coma decimal antes del primer dígito del número calculado en 2. Así se obtiene un número aleatorio.
5. Tomar el número encontrado como nueva semilla y repetir el proceso a partir de 2.
   Ejemplo: sea la semilla S0= 5625
   S1 = $(5625)^2$ = 31640625 R1= 0,6406
   S2 = $(6406)^2$ = 41036836 R2= 0,0368
   S3 = $(0368)^2$ = 00135425 R3= 0,1354
   Etc.

Método congruencial lineal

- Utiliza la ecuación recursiva con números aleatorios generados en cada iteración:

$$Z_i = (aZ_{i-1} + c) \bmod m \qquad R_i = \frac{Z_i}{m}$$

- Z_0 : semilla
  a, c y m son enteros no negativos y deben satisfacer que m > 0, a < m, c < m y Z_0 < m
  Si la constante c es cero el generador es congruencial lineal multiplicativo.
- Si tiene la fórmula completa es congruencial lineal mixto.

Método congruencial lineal

1. La primera parte de la ecuación es la operación de división modular, o sea que lo que se calcula en el paréntesis se divide por m y el resto es el valor buscado de Z_i.
2. Posteriormente se divide este resto por el módulo para obtener el número aleatorio R

Método congruencial lineal

- Ejemplo: sean la semilla Z0 = 1; y las constantes a = 6; c = 1; m = 25

- Z1 = (6x1+1)mod 25 → Z1=7 → R1 = 7/25=0,28
- Z2 = (6x7+1)mod 25 → Z2=18 → R2 = 18/25=0,72
- Z3 = (6x18+1)mod 25 →Z3=9 → R3 = 9/25=0,36
  Etc.

Método congruencial lineal

- Si en lugar de a=6 se toma a=5
- Z1 = (5x1+1)mod 25→ Z1=6 → R1 = 6/25=0,24
- Z2 = (5x6+1)mod 25 → Z2=6 → R2 = 6/25=0,24
- Z3 = (5x6+1)mod 25 → Z3=6 → R3 = 6/25=0,24
- Etc.

  Esto demuestra que los parámetros deben ser cuidadosamente seleccionados

Pruebas uniformidad e independencia

- Prueba de Uniformidad para ajustar la distribución obtenida con la distribución uniforme:
  - Chi-cuadrado
  - Kolmogorov- Smirnov)
- Pruebas de Independencia:
  - Prueba de Autocorrelación
  - Prueba de Gap: cuenta el número de dígitos que aparece entre la repetición de un dígito en particular, usa el test de Kolmogorov-Smirnov para comparar con el tamaño esperado de los gaps.
  - Prueba del póker: Trabaja con los números agrupados como en una mano de poker. Las manos obtenidas se comparan a lo esperado usando un test de chi-cuadrado.

Generación de variables aleatorias continuas y discretas

- Método de la transformada inversa.
- Método del rechazo.
- Método de caracterización.
- Métodos de composición y descomposición.

UTN SANTA FE

Transformada Inversa

- Sea f(x) la distribución de probabilidad para generar las variables aleatorias, y su distribución acumulada F(x)

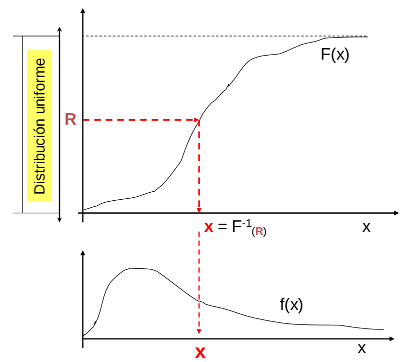
$$F(x) \in (0\text{-}1)$$

$$F(x) = R$$

$$x = F^{-1}(R)$$

UTN SANTA FE

Transformada Inversa



UTN SANTA FE

Distribución exponencial

$$f(x) = \begin{cases} \lambda e^{-\lambda x} & \text{si } x \geq 0 \\ 0 & \text{si } x \geq 0 \end{cases}$$

$F(x) = \int \lambda e^{-\lambda x}\, dx = 1 - e^{-\lambda x}$ entre 0 y x

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

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

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

R y 1 – R tienen una distribución uniforme
Por lo que es indistinto usarlos

$x = -1/\lambda \ln R$

UTN SANTA FE

Distribución uniforme continua

$$f(x) = \begin{cases} \dfrac{1}{b-a} & \text{si } a \leq x \leq b \\ 0 & \text{Para } x < 0 \text{ y } x > b \end{cases}$$

$$F(x) = \int \frac{1}{b-a}\, dx = \frac{x-a}{b-a} = R$$

$$x = a + (b - a)\, R$$

UTN SANTA FE

Otras distribuciones empíricas

$$f(x) = 2x \quad 0 \le x \le 1$$

$$F(x) = \int_0^x 2x\,dx = 2\frac{x^2}{2}\Big]_0^x = x^2 = R$$

$$x = \pm\sqrt{R}$$

$$\boxed{x = \sqrt{R}} \qquad 0 \le x \le 1$$

UTN SANTA FE

Distribuciones discretas

P(xi) = ri   i=1,2,3……,n

$$F(x) = P(X \le x_i) = \sum_{\forall x_i \le x} p(x_i) \quad i = \mathbf{1}, \mathbf{2}, \ldots n$$

$$\boxed{F(x_{i-1}) = r_{i-1} \le R \le r_i = F(x_i)}$$

1. Se genera R y se lo compara en el intervalo
2. Se busca el entero positivo tal que U ≤ F(x), y se retorna el valor de X=xi

UTN SANTA FE

Ejemplo

x	p(x)	F(x)
0	0,5	0,5
1	0,3	0,8
2	0,2	1

$$F(x) = \begin{cases} 0 & x < 0 \\ 0{,}5 & 0 \le x < 1 \\ 0{,}8 & 1 \le x < 2 \\ 1 & 2 \le x \end{cases}$$

$$x = F^{-1}(r) = \begin{cases} 0 & 0 \le R < \le 0{,}5 \\ 1 & 0{,}5 < R < 0{,}8 \\ 2 & 0{,}8 < R \le 1 \end{cases}$$

UTN SANTA FE

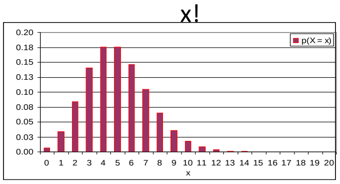
Ventajas y desventajas de la T.I.

- La desventaja de este método es que se necesita tener la expresión de x=F-1 (R), la que no es siempre fácil de encontrar.
- De todas maneras puede trabajarse con tablas de números e interpolar el método de interpolación depende de que la función sea continua o discontinua, lo que también hay que especificarle.

Ejemplo 3

- Diseñar un generador de variables aleatorias para una distribución Poisson :

$$f(x)=\frac{e^{-5}5^{x}}{x!}$$



Método de caracterización

- Las distribuciones con características especiales se generan mediante algoritmos especializados.
- Son los casos de las distribuciones normal y de Poisson, entre otros.
- Las expresiones de las funciones de probabilidad son complejas. Por eso se resuelven con algoritmos.

Distribución normal

- Para generar variables aleatorias a partir de la función de distribución normal, se considera la ventaja del teorema central del límite que asegura:
  – la suma de k variables aleatorias independientes y distribuidas idénticamente con una media μ y una varianza σ^2 es aproximadamente distribuida normalmente con media nμ y variancia $n\sigma^2$.

Distribución normal

- Si se toman k números aleatorios entre 0 y 1 con media 0,5 y varianza 1/12 (porque la media de una distribución uniforme continua es (a+b)/2 y la variancia es (b-a)2/12,

  Entonces para a = 0 y b = 1 se tienen los valores de media 0,5 y varianza 1/12
- La variable aleatoria de una distribución normal con media 0,5 k y varianza k/12 se calcula como:

$$x = \sum_{i=1}^{k} R$$

Distribución normal

- Para transformar la variable x en la variable normal Z estándar se tiene:

$$Z = .\frac{\left(\sum_{i=1}^{k} R_i - 0{,}5k\right)}{\sqrt{k/12}}$$

- Cuando k aumenta mejora la aproximación, pero aumenta el tiempo de cálculo, por lo que un valor razonable de k es 12:

$$Z = .\sum_{i=1}^{12} R_i - 6$$

Distribución normal

Y como se quiere una variable normal X con media μ y desviación normal σ, se calcula con la ecuación:

$$x = \sigma.\left(\sum_{i=1}^{12} R_i - 6\right) + \mu \qquad 0 \le R \le 1$$

Distribución de Poisson

Para una distribución de Poisson, con tiempos entre arribos distribuidos exponencialmente con media $1/\lambda$, el número de arribos n en un período dado T tiene una distribución de Poisson con parámetro λ (llegadas/unidad de tiempo).

Distribución de Poisson

- Para generar una variable con distribución de Poisson, se suman las variables generadas exponencialmente hasta que la suma exceda el valor de T y se retorna el número de variables generadas como variable de Poisson.

Demostración:

- La función de Poisson es

$$f(x)=\frac{e^{-\lambda}\lambda^{x}}{x!}\qquad x=1,2,3....\infty$$

- Los intervalos entre arribos de las llegadas de Poisson (discreta) tienen una distribución exponencial (continua).

Demostración

- Podemos decir que si los tiempos entre arribos antes del arribo n son variables exponenciales T_1, T_2, T_x, la siguiente ecuación indica que hay exactamente N=n arribos durante Δt:

$$\sum_{j=0}^{x} T_j \leq \Delta t \leq \sum_{j=0}^{x+1} T_j \qquad x = 0,1,2...$$

- Siendo Tj = -1/λ ln Rj, se reemplaza en la ec. anterior:

$$-\sum_{j=0}^{x} \mathbf{1}/\lambda \ln R_j \leq \Delta t \leq -\sum_{j+0}^{x+1} \mathbf{1}/\lambda \ln R_j$$

- Por propiedades de los logaritmos (suma de logaritmos = logaritmo del producto de los términos) y aplicando el antilogaritmo:

$$\prod_{j=0}^{x+1} R_j \leq e^{-\lambda\Delta t} \leq \prod_{j=0}^{x} R_j$$

- Cuando se cumple con la condición de la desigualdad, se considera que x es la variable de Poisson.

Algoritmo de generación de variables discretas de Poisson.

- Se tiene un contador en 1 (s = 1), se calcula A = $e^{-\lambda\Delta t}$ y se hace x = 0
- Se genera R
- Se hace S=S.R
- Si S > A, x=x+1 y se repite desde 2.
- Si S<A, x es la variable de Poisson.

Ejemplo:

- Se quiere generar una variable aleatoria de Poisson cuyo parámetro es la velocidad de arribos λ=2
- Primera iteración
  - λ=2, Δt =1; A= e^{-2} = 0,13533528 ; S =1; x = 0
  1. R = 0,453
  2. S = 1 x 0,453 =0,453
  3. S > A (0,453 > 0,13533528), x = 0 +1 = 1
- Segunda iteración
  1. R = 0,721
  2. S = 0,453 x 0,721 = 0,326613
  3. S > A (0,326613 > 0,13533528), x = 1 +1 = 2

- Tercer iteración
  1. R = 0,294
  2. S = 0,326613 x 0,294 = 0,09602422
  3. S < A (0,09602422 < 0,13533528), x = 2 es la variable de Poisson

- Entonces, para un intervalo de tiempo con distribución exponencial λ = 2 arribos/seg (Δt = 1 seg), se producen 2 arribos.

Método del rechazo

- El método sirve para generar variables aleatorias cuando x está acotada en un intervalo [a,b], a ≤ x ≤b.
- Se utiliza la función distribución de probabilidad f(x) y el concepto de moda:

  M = máximo valor de f(x) ⇒ f(x) acotada.
- Es un método iterativo para generar cada valor de x.

Método del rechazo

- Como paso previo hay que dividir todos los valores de la función por la moda, implica escalonarla entre 0 y 1, de manera de asociarla con los números aleatorios.

Método del rechazo

- El procedimiento es:

1. Generar 2 números aleatorios R_1 y R_2 ambos entre [0,1]
2. Determinar x de acuerdo a: $x=a+(b-a).R_1$
3. Evaluar $\frac{f(x)}{M} = \frac{f[a+(b-a)R_1]}{M} = \frac{f(x)}{M}$

1. Determinar si se cumple que:

$$R_2 \leq \frac{f(x)}{M} = \frac{f[a+(b-a).R_1}{M}$$

2. Si la respuesta es sí, entonces x es el valor de la variable aleatoria, y x es aceptado.
3. Si la respuesta es no, entonces x es rechazado, y hay que volver a comenzar desde 1.

Método del rechazo

- La teoría sobre la que se basa el método es que la probabilidad de que $R_2 \leq f(x)/M$ es f(x)/M.
- Entonces si esto se cumple y x fue elegido de acuerdo a la expresión acumulativa de la distribución uniforme (x=a+(b-a).R), entonces x pertenece a f(x).
- En el caso particular en que todos los valores de x fueran aceptados, entonces se tendría distribución uniforme.

Desventajas del método

- El número esperado de intentos para que x sea aceptada es M, entonces si M es grande, aumenta la ineficiencia.

Ejemplo:

- Sea la distribución $f(x) = 2.x$; $0 \leq x \leq 1$;
- $a = 0$; $b=1$; $f(x) = 0 \ \forall \ x < 0$
- Se desea estimar una variable aleatoria con esa distribución.
- $M=2$
- $R1= 0,21$ y $R2=0,87$
- $x=a+R(b-a)=0,21$
- $f(x) = 2 * 0,21 = 0,42$
- $f(x)/M = 0,42/2 = 0,21 < R2 \Rightarrow x$ se acepta

¿Qué método conviene usar para generar variables aleatorias?

- Si la función es invertible, usar transformada inversa.
- Si la función no es invertible, el método de rechazo o el método de caracterización para distribuciones normales o de Poisson.
- Los métodos de convolución y composición se usan para casos especiales.