TP 2.2 - GENERADORES DE NÚMEROS PSEUDOALEATORIOS DE DISTINTAS DISTRIBUCIONES DE PROBABILIDAD

Galarza Karen
Ingeniería en Sistemas de Información
Universidad Tecnológica Nacional
karengalarza94.kg@gmail.com

Gonzalez Aquino Yoana
Ingeniería en Sistemas de Información
Universidad Tecnológica Nacional
yoanakim1@gmail.com

Grancelli Eliseo
Ingeniería en Sistemas de Información
Universidad Tecnológica Nacional
eliseograncelli@gmail.com

Soto Matias Francisco
Ingeniería en Sistemas de Información
Universidad Tecnológica Nacional
matiasfranciscosoto123@gmail.com

Petrelli Juan Franco
Ingeniería en Sistemas de Información
Universidad Tecnológica Nacional
jfpetrelli@gmail.com

Agustina Monti
Ingeniería en Sistemas de Información
Universidad Tecnológica Nacional
agusmonti10@gmail.com

13 de julio de 2023

ABSTRACT

En el siguiente documento haremos un estudio de las distintas distribuciones de probabilidad, construyendo generadores de números pseudoaleatorios con disintos tipos de probabilidades. Veremos que las distribuciones de probabilidad teóricas son útiles ya que si la distribución real de un conjunto de datos dado es razonablemente cercana a la de una distribución de probabilidad teórica, muchos de los cálculos se pueden realizar en los datos reales utilizando hipótesis extraídas de la distribución teórica.

1. Introducción

Para realizar una simulación, puede ser necesario generar valores de una variable aleatoria que pertenezca a una distribución diferente de la uniforme. Ahora que tenemos acceso a una fuente de generadores de números aleatorios de distribución uniforme, podemos aplicar ciertas técnicas para convertirlos en las distribuciones deseadas. Haremos esto mediante el método de la transformada inversa, aunque existen otros conocidos para hacerlo (entre ellos el métodos de rechazo y el de composición)

2. Distribucion Continua

Una distribución continua describe las probabilidades de los posibles valores de una variable aleatoria continua. Una variable aleatoria continua es una variable aleatoria con un conjunto de valores posibles (conocido como el rango) que es infinito y no se puede contar (1.8, 2.5, 135.36).

2.1. Función de Densidad

Sea X una variable aleatoria continua. Decimos que su función de densidad $f_x(x)$ de probabilidad asociada a X es cuando se cumple:

- $f_x(x) \leq 0$ para todo xϵ R
- $\int_{-\infty}^{+\infty} f_x(x)dx = 1$
- $P(a \leq X \leq b)$ $\int_a^b f_x(x)dx$

2.2. Función de distribución acumulada

Sea X una variable aleatoria continua y $f_x(x)$ su densidad de probabilidad. Se denomina Función de distribución acumulada de la variable X a:

$F_x : R \rightarrow [0, 1]$
$X \rightarrow F_x(x) = \int_{-\infty}^{x} f_t dt$

Propiedades:

- $F_x = P(X \leq x)$
- F_x es no decreciente
- $F_x^{'} = f_x(x)$en los puntos de continuidad de f_x

2.3. Esperanza matemática de una variable continua

Sea X una variable aleatoria continua y $f_x(x)$ su función de densidad de probabilidad .La esperanza matemática o media poblacional de X, que se nota E(X) o µX es siempre que la integral sea finita:

- $E(X) = \mu X = \int_{-\infty}^{+\infty} f_x(x)dx$

2.4. Varianza de una variable continua

Sea X una variable aleatoria continua y $f_x(x)$ su función de densidad de probabilidad. La varianza de X que se nota V(X) y es el numero no negativo de:

- $V(X) = \int_{-\infty}^{+\infty} (x - \mu x)^2 f_x(x)dx$

3. Las distribuciones que tienen este tipo de variable son:

- Distribución Uniforme
- Distribución Exponencial
- Distribución Gamma
- Distribución Normal

3.1. Distribución uniforme

Es el modelo continuo más simple. Corresponde al caso de una variable aleatoria que sólo puede tomar valores comprendidos entre dos extremos a y b, de manera que todos los intervalos de una misma longitud (dentro de (a, b)) tienen la misma probabilidad de ocurrir. También puede expresarse como el modelo probabilístico correspondiente a tomar un número al azar dentro de un intervalo (a, b). De la anterior definición se desprende que la función de densidad debe tomar el mismo valor para todos los puntos dentro del intervalo (a, b) (y cero fuera del intervalo). Matemáticamente, la función de densidad uniforme se define así:

$$f(x) = \begin{cases} \frac{1}{b-a} & \text{si } a < x < b \\ 0, & \text{fuera del intervalo (a,b)} \end{cases} \tag{1}$$

El valor esperado y la varianza de una variable aleatoria uniformemente distribuida están dadas por

$$EX = \frac{b+a}{2}$$

$$VX = \frac{(b-a)^2}{12}$$

Si la variable aleatoria X sigue distribución uniforme con dominio (a,b), podemos escribir

$$X \sim U(a, b)$$

donde a y b son los extremos mínimo y máximo del dominio.

Transformada inversa(Distribucion Uniforme):

$$f\left(x\right) = \frac{1}{b-a} a \leq x \leq b \tag{2}$$

$$F\left(x\right) = \int_0^x \frac{1}{b-a} a \leq x \leq b \tag{3}$$

$$F\left(x\right) = \frac{x-a}{b-a} \tag{4}$$

$$x = F\left(r\right)^{-1} \tag{5}$$

$$r = \frac{x-a}{b-a} \tag{6}$$

$$r\left(b-a\right) = x-a \tag{7}$$

$$x = r\left(b-a\right) + a \tag{8}$$

3.2. Distribución exponencial

Describe procesos en los que interesa saber el tiempo hasta que ocurre determinado evento, sabiendo que el tiempo que pueda ocurrir desde cualquier instante dado t, hasta que ello ocurra en un instante tf, no depende del tiempo transcurrido anteriormente en el que no ha pasado nada. El uso de la distribución exponencial supone que los tiempos de servicio son aleatorios, es decir, que un tiempo de servicio determinado no depende de otro servicio realizado anteriormente ni de la posible cola que pueda estar formándose. Se puede expresar mediante la función

$$f(x) = \alpha e^{-\alpha x}, \quad \text{si } a > 0 \quad \text{y} \quad x \geq 0$$

La media y la varianza se pueden expresar como

$$EX = \frac{1}{\alpha}$$

$$VX = EX^2$$

Si la variable aleatoria X sigue distribución exponencial podemos decir que

$$X \sim \text{Exp}(\lambda)$$

donde a y b son los extremos mínimo y máximo del dominio.

Transformada inversa(Distribucion Exponencial):

$$f\left(x\right)=\alpha e^{-\alpha x},\alpha>0,x\geq0 \tag{9}$$

$$F\left(x\right)=\int_{0}^{x}\alpha e^{-\alpha x} \tag{10}$$

$$F\left(x\right)=1-e^{-\alpha x} \tag{11}$$

$$x=F\left(r\right)^{-1} \tag{12}$$

$$r=1-e^{-\alpha x} \tag{13}$$

$$1-r=e^{-\alpha x} \tag{14}$$

$$ln\left(1-r\right)=ln\left(e^{-\alpha x}\right) \tag{15}$$

$$ln\left(1-r\right)=-\alpha x \tag{16}$$

$$x=\frac{-ln\left(1-r\right)}{\alpha} \tag{17}$$

3.3. Distribución Gamma

Si un determinado proceso consiste de k eventos sucesivos y si el total del tiempo transcurrido para dicho proceso se puede considerar igual a la suma de k valores independientes de la variable aleatoria con distribución exponencial, cada uno de los cuales tiene un parámetro definido a, la distribución de esta suma coincidirá con una distribución exponencial con parámetros a y k.

Se utiliza para modelar variables que describen el tiempo hasta que se produce p veces un determinado suceso. Su función de densidad es de la forma:

$$f(x)=\frac{\alpha^k x^{k-1}e^{-\alpha x}}{(k-1)!}$$

Su media y varianza se formulan como sigue:

$$EX=\frac{k}{\alpha}$$

$$VX=\frac{k}{\alpha^2}$$

Si una variable aleatoria continua X tiene distribución gamma con parámetros $\alpha>0$ y $\lambda>0$, entonces escribiremos

$$X\sim\Gamma(\alpha,\lambda)$$

3.4. Distribución Normal

Es la más conocida y ampliamente utilizada distribución de probabilidad (entre otras cosas por el Teorema Central del Límite y porque a partir de ella se pueden derivar muchas otras de las distribuciones existentes). Como curiosidad podemos mencionar que el nombre de Normal proviene del hecho de que durante un tiempo,médicos y biólogos creían que todas las variables naturales de interés seguían este modelo. Su función de densidad es

$$f(x)=\frac{1}{\sigma_x\sqrt{2\pi}}e^{-\frac{1}{2}(\frac{x-\mu_x}{\sigma_x})^2}$$

Su valor esperado y varianza están dados por

$$EX=\mu_x$$

$$VX=(\sigma_x)^2$$

3.4.1. Distribuciones discretas de probabilidad

$$\text{Distribuciones discretas}\begin{cases}\text{Binomial}\\ \text{Binomial negativa}\\ \text{Hipergeométrica}\\ \text{Poisson}\\ \text{Empírica discreta}\end{cases}$$

Distribución binomial
Es una distribución de probabilidad discreta que cuenta el número de éxitos en una secuencia de n ensayos de Bernoulli independientes entre sí con una probabilidad fija p de ocurrencia de éxito entre los ensayos. Un experimento de Bernoulli se caracteriza por ser dicotómico, esto es, solo dos resultados son posibles, a uno de estos se le denomina "éxito" y tiene una probabilidad de ocurrencia p y al otro se le denomina "fracaso" y tiene una probabilidad $q = 1 - p$.

Si la variable aleatoria X sigue distribución binomial con parámetros n $\in \mathbb{N}$ y $0 < p < 1$, podemos escribir

$$X \sim \text{Bin}(n, p)$$

Distribución binomial negativa (Pascal)
Es una ampliación de la distribución geométrica, utilizada en procesos en los cuales se ve necesaria la repetición de ensayos hasta conseguir un número de casos favorables (primer éxito). Mide el número de éxitos en una secuencia de n ensayos de Bernoulli independientes entre sí, con una probabilidad p de ocurrencia de éxitos en los ensayos.Un experimento de Bernoulli se caracteriza por presentar sólo son posibles dos resultados(A y no A). La variable aleatoria es el número de ensayos Bernoulli necesarios para obtener el primer éxito. Si deseamos conocer el número de estos para conseguir n éxitos, la variable aleatoria es binomial negativa.El número de experimentos de Bernoulli de parámetro p independientes realizados hasta la consecución del k-ésimo éxito es una variable aleatoria que tiene una distribución binomial negativa con parámetros k y p. La distribución geométrica es el caso concreto de la binomial negativa cuando k = 1.Su función de probabilidad es:

$$f(x) = \binom{k+x-1}{x} p^k (1-p)^x \text{ con x = 0,1,2,...}$$

donde k es el número total de éxitos en una sucesión de k+x ensayos, con x el número de fallas que ocurren antes de obtener k éxitos. El valor esperado y la variación de X se representa con

$$EX = \frac{k(1-p)}{p}$$
$$VX = \frac{k(1-p)}{p^2}$$

Si X es una variable aleatoria continua con distribución binomial negativa con parámetros r y p entonces escribiremos

$$X \sim \text{BN(r,p)}$$

Distribución hipergeométrica
Es una distribución de probabilidad discreta relacionada con muestreos aleatorios y sin reemplazo. Considérese una población que consta de N elementos tales que cada uno de ellos pertenece a la clase I o a la II. Sean N_p el número de elementos que pertenecen a la clase I y N_q el número de elementos miembros de la clase II, donde $p + q = 1$. Si en una población de N elementos se toma una muestra aleatoria que conste de n elementos (con $n < N$) sin que tenga lugar algùn reemplazo, entonces el número de elementos x de la clase I en la muestra de n elementos, tendrá una distribución hipergeométrica.

Esta distribución está descrita por la siguiente función de probabilidad:

$$f(x) = \frac{\binom{N_p}{x}\binom{N_q}{n-x}}{\binom{N}{n}} \quad \text{donde} \quad 0 \leq x \leq N_p \quad \text{y} \quad 0 \leq n - x \leq N_q$$

La esperanza y varianza se describen como sigue:

$$EX = np$$

$$VX = npq \binom{N-n}{N-1}$$

Si una variable aleatoria discreta X sigue una distribución hipergeométrica entonces escribiremos

$$X \sim HG(N, K, n)$$

donde $k = qN$

Distribución Poisson
Expresa, a partir de una frecuencia de ocurrencia media, la probabilidad de que ocurra un determinado número de eventos durante cierto período de tiempo. Concretamente, se especializa en la probabilidad de ocurrencia de sucesos con probabilidades muy pequeñas, o sucesos raros. Si tomamos una serie de n ensayos independientes de Bernoulli, en cada uno de los cuales se tenga una probabilidad p muy pequeña relativa a la ocurrencia de un cierto evento, a medida que n tiene al infinito, la probabilidad de x ocurrencias está dada por la distribución de Poisson

$$f(x) = e^{-\lambda} \frac{\lambda^x}{x!} \quad \text{con} \quad x = 0, 1, 2, ...$$

Si la variable aleatoria discreta X tiene una distibución de Poisson con parámetro λ, podemos escribir

$$X \sim \text{Poisson}(\lambda)$$

Distribución empírica discreta
Se utiliza cuando no se conoce la distribución de la variable aleatoria, hay que identificarla. Los valores son generados a partir de una tabla de distribución de frecuencias.

3.4.2. El método de la transformada inversa

Supongamos que existe una variable aleatoria X cuya distribución puede ser descrita por su función de distribución acumulada. También supongamos que queremos generar valores de X acordes a esta distribución. El método de la transformada inversa nos ofrece una solución. Funciona de la siguiente manera:

1. Generar un número aleatorio u de distribución uniforme en el intervalo $(0, 1)$, es decir, genera números a partir de $U \sim U(0, 1)$.
2. Encontrar la inversa de la función de distribución acumulada, es decir, $F_x^{-1}(x)$.
3. Calcular $X = F_x^{-1}(u)$.

Dicho en otras palabras, dada una variable uniforme continua U en $(0, 1)$ y una función de distribución acumulada invertible F_X, la variable aleatoria $X = F_x^{-1}(U)$ tiene distribución F_X.

3.5.

3.6.

3.7.

4. Resultados

Mediante la aplicación de los distintas pruebas a los distintos generadores obtuvimos los siguientes resultados.

4.1. Método de la parte media del cuadrado

Se define la semilla = 1991 y se generan 1500 números aleatorios. En la Figura 1 graficamos los números generados por el generador ordenados en forma ascendente (Naranja) y lo comparamos con el valor esperado (Azul) lo que nos indicaría que la distribución no es uniforme.

Imagenes/1.png

En esta figura podemos observar que tampoco parecen tener una distribución uniforme.

Resultados de tests generador parte media del cuadrado
Test chicuadrados: Rechazado
Test de corridas: Rechazado
Test prueba de series: Rechazado

Imagenes/4.png

En este diagrama de dispers

4.2. Generador congruencial lineal

Se define la semilla = 1991, módulo = 32768, multiplicador = 26765, incremento = 21001 y se generan 1500 números aleatorios. En la Figura 3 graficamos los números generados por el generador ordenados en forma ascendente (Naranja) y lo comparamos con el valor esperado (Azul) lo que nos indicaría que la distribución podría ser uniforme

Imagenes/2.png

Mediante la figura 4 podemos observar que los números parecen tener una distribución uniforme.
Resultados de tests generador congruencial lineal
Test chicuadrados: Aceptado
Test de corridas: Aceptado
Test prueba de series: Aceptado

Imagenes/5.png

4.3. Generador python

En la figura 5 graficamos los números generados por el generador ordenados en forma ascendente (Naranja) y lo comparamos con el valor esperado (Azul) lo que nos indicaría que la distribución podría ser uniforme.

Imagenes/3.png

Mediante la figura 6 podemos observar que los números parecen tener una distribución uniforme.
Resultados de tests generador python

Test chicuadrados: Aceptado
Test de corridas: Aceptado
Test prueba de series: Aceptado

Imagenes/6.png

5. Conclusión

Considerando las pruebas realizadas, se puede observar que ninguno de los generadores pseudo-aleatorios han logrado superar todas las pruebas. Esto evidencia que ningún generador produce números aleatorios en realidad, si no que emulan este comportamiento mediante algoritmos informáticos que no se predicen con facilidad y permiten generar estos números. Es realmente imposible modelar con precisión las caracteristicas de la distribución uniforme mediante un buen generador. Esto quiere decir que algunas propiedades no se cumplirán, las cuales pueden no influir mucho en los resultados de determinado estudio, dando lugar a que el criterio de aceptación de un generador dado debe basarse en la aplicación que se le vaya a dar. Es precisamente responsabilidad del analista realizar las pruebas pertinentes.

6. Anexo

Bibliografía

[1] En
https://es.overleaf.com/latex/templates/style-and-template-for-preprints-arxiv-bio-arxiv/pkzcrhzcdxmc.

[2] En
https://www.random.org/analysis/

[3] En
https://tereom.github.io/est-computacional-2018/numeros-pseudoaleatorios.html/

[4] En
https://https://latex.codecogs.com/eqneditor/editor.php?lang=es-es/