CAPÍTULO II

2. MODELOS ESTOCÁSTICOS A UTILIZARSE PARA IMPUTACIÓN DE DATOS

2.1 Introducción

Este segundo capítulo aborda el tema de las técnicas y principios científicos que permiten la generación de números aleatorios, los mismos que son necesarios para la simulación de sistemas que se explican estocásticamente. Para esto, la sección 2.2 trata acerca de la Distribución Uniforme; la siguiente sección detalla los Métodos para la Generación de números aleatorios; después se describe un Método de Generación de Variables Aleatorias no Uniformes y por último se muestra las pruebas de hipótesis sobre los parámetros o la distribución de una población.

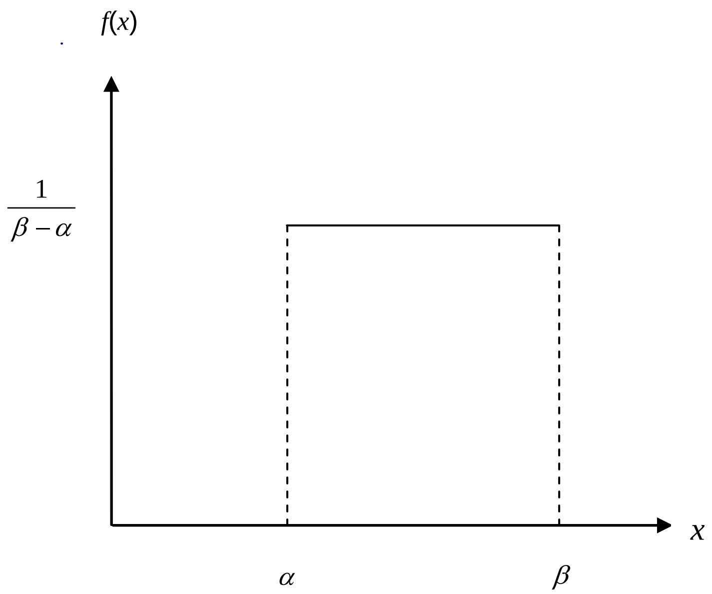
2.2 Distribución Uniforme

Una variable aleatoria X tiene una *distribución uniforme continua* con parámetros α y β si y solo sí su función de densidad f, está dada por

$$f(x) = \begin{cases} \dfrac{1}{\beta - \alpha}, & X \in (\alpha, \beta) \\ 0 \quad , & para\, el\; resto\; de\; X \end{cases} \tag{2.1}$$

Los parámetros α y β de esta variable son constantes reales con $\alpha < \beta$, véase Gráfico 2.1.

GRÀFICO 2.1
Efectos de la Imputación en el análisis de datos multivariados
Densidad de la Distribución Uniforme



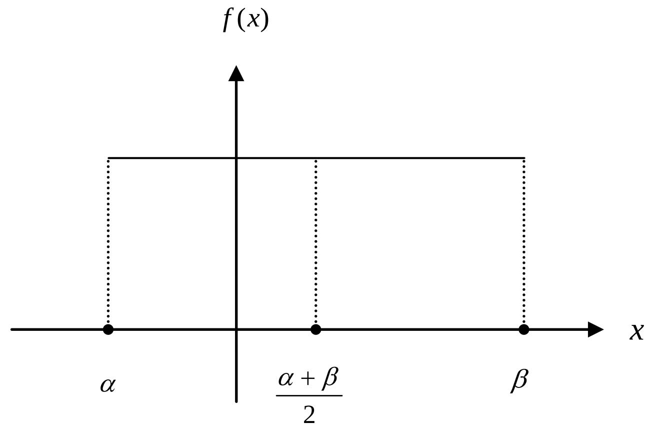
Elaborado por: G. Cuenca

La media y la varianza de la distribución uniforme están dadas por:

$$E(X) = \int_{-\infty}^{\infty} xf(x)dx = \int_{\alpha}^{\beta} x\left(\frac{1}{\beta-\alpha}\right)dx = \left(\frac{x^2}{2(\beta-\alpha)}\right)_{\alpha}^{\beta} = \frac{\beta^2}{2(\beta-\alpha)} - \frac{\alpha^2}{2(\beta-\alpha)}$$

$$= \frac{\beta^2-\alpha^2}{2(\beta-\alpha)} = \frac{(\beta+\alpha)(\beta-\alpha)}{2(\beta-\alpha)} = \frac{\beta+\alpha}{2} = \mu \qquad (2.2)$$

GRÁFICO 2.2
Efectos de la Imputación en análisis de datos multivariados
Media de la distribución uniforme



Elaborado por: G. Cuenca

Por lo tanto, μ se ubica en el punto medio entre α y β, véase Gráfico 2.2

Además si $X \sim U(\alpha, \beta)$,

$$Var(X) = E(X-\mu)^2 = E(X^2) - \mu^2$$

$$E(X^2) = \int_{-\infty}^{\infty} x^2 f(x)dx = \int_{-\infty}^{\infty} x^2 \left(\frac{1}{\beta-\alpha} \right) dx = \frac{1}{\beta-\alpha} \int_{\alpha}^{\beta} x^2 dx = \frac{1}{\beta-\alpha} \left(\frac{x^3}{3} \right)_{\alpha}^{\beta}$$

$$= \frac{1}{\beta-\alpha} \left(\frac{\beta^3}{3} - \frac{\alpha^3}{3} \right) = \frac{1}{\beta-\alpha} \left(\frac{\beta^3-\alpha^3}{3} \right) = \frac{\cancel{(\beta-\alpha)}(\beta^2+\alpha\beta+\alpha^2)}{3\cancel{(\beta-\alpha)}}$$

$$Var(X) = \frac{\beta^2+\alpha\beta+\alpha^2}{3} - \left(\frac{\alpha+\beta}{2} \right)^2 = \frac{\beta^2+\alpha\beta+\alpha^2}{3} - \frac{\alpha^2+2\alpha\beta+\beta^2}{4}$$

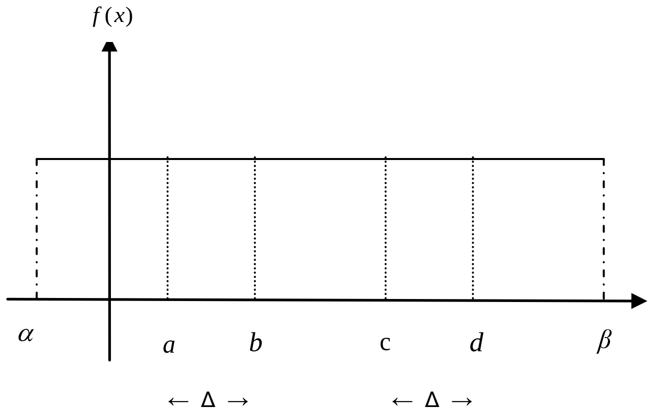
$$= 4(\beta^2+\alpha\beta+\alpha^2) - 3(\alpha^2+2\alpha\beta+\beta^2)$$

$$= \frac{4\beta^2+4\alpha\beta+4\alpha^2-3\alpha^2-6\alpha\beta-3\beta^2}{12}$$

$$= \frac{\beta^2-2\alpha\beta-\alpha^2}{12} = \frac{(\beta-\alpha)^2}{12} \tag{2.3}$$

Para comprobar que la probabilidad de que un valor ocurra en un intervalo, solo depende de la longitud de dicho intervalo, efectuamos lo siguiente:

GRÁFICO 2.3
Efectos de la Imputación en análisis de datos multivariados
Números en el intervalo $X \in (\alpha, \beta)$



Elaborado por: G. Cuenca

Sea $X \sim U(\alpha, \beta)$, tal que:

$$f(x) = \begin{cases} \dfrac{1}{\beta - \alpha}, & X \in (\alpha, \beta) \\ 0 \quad , & \text{para el resto de } X \end{cases}$$

Además, el intervalo (a, b) está incluido en (α, β) al igual que (c, d), esto es:

$$Si\ (a,b) \subseteq (\alpha, \beta) \quad y \quad (c,d) \subseteq (\alpha, \beta)$$

Supongamos además que:

$(b - a) = \Delta x$ y que igualmente $(c - d) = \Delta x$, esto significa que:

$$\mathrm{P}(X \in (a,b)) = \int_a^b \frac{1}{\beta - \alpha} dx = \frac{1}{\beta - \alpha} (b - a) = \frac{\Delta x}{\beta - \alpha}$$

$$\mathrm{P}(X \in (c,d)) = \int_c^d \frac{1}{\beta - \alpha} dx = \frac{1}{\beta - \alpha} (d - c) = \frac{\Delta x}{\beta - \alpha}$$

Por tanto $P\left(X \in (a,b)\right)=P\left(X \in (c,d)\right)=\dfrac{\Delta x}{\beta-\alpha}$ (2.4)

Caso particular: Si $\alpha=1$ y $\beta=10$; $X \sim U(1,10)$

En este caso la densidad f es tal que:

$$f(x)=\begin{cases}\dfrac{1}{9} & X \in (1,10)\\ 0 & \text{resto de } X\end{cases}$$

$$\text{si } a=2,\ \text{b}=5$$

$$\text{P}\left(X \in (2,5)\right)=\int_2^5 \frac{1}{9}\,dx=\frac{1}{9}(5-2)=\frac{3}{9}$$

$$\text{si c}=1,\ \text{d}=4$$

$$\text{P}\left(X \in (1,4)\right)=\int_1^4 \frac{1}{9}\,dx=\frac{1}{9}(4-1)=\frac{3}{9} \quad \text{entonces}\ \Delta\text{x}=3$$

En este caso particular se puede comprobar que escogiendo números que se encuentran en intervalos de igual longitud $\Delta x=3$, la probabilidad de que se efectúe una lectura en dicho intervalo es la misma, cuando la variable aleatoria es uniforme.

En cambio si escogemos intervalos de longitud diferente a tres, la probabilidad de que algo ocurra en dicho intervalo, no es $\frac{3}{9}$.

$$\text{si } h=1,\ i=6 \quad \text{entonces}\ \Delta x=6-1=5$$

$$\text{P}\left(X \in (1,6)\right)=\int_1^6 \frac{1}{9}\,\text{d}x=\frac{1}{9}(6-1)=\frac{5}{9}$$

La Distribución Acumulada de $X \sim U(\alpha, \beta)$ está dada por:

$$P(X \leq x) = F(x) = \begin{cases} 0 & si \quad X \leq \alpha \\ \dfrac{x-\alpha}{\beta-\alpha} & si \quad \alpha < X < \beta \\ 1 & si \quad X \geq \beta \end{cases} \tag{2.5}$$

Si $X \sim U(0,1)$; $F(x) = \begin{cases} 0 & \text{si} \quad X \leq 0 \\ x & \text{si} \quad X \in (0,1) \\ 1 & \text{si} \quad X \geq 1 \end{cases}$

esto es $F(x) = x,\ x \in (0,1)$ lo cual significa que:

$$F(0.10) = P(X \leq 0.10) = 0.10$$
$$F(0.15) = P(X \leq 0.15) = 0.15$$
$$F(0.99) = P(X \leq 0.99) = 0.99,\ etc.$$

Por definición la función generadora de momentos de una variable aleatoria continua X es:

$$M_X(t) = E\left[e^{Xt}\right] = \int_{-\infty}^{\infty} e^{Xt} f(x)dx \tag{2.6}$$

la variable independiente es t, y por lo general estamos interesados en los valores de t en una vecindad de cero, por ejemplo $|t| < h$

Ahora se calcula la Función Generadora de Momentos de la Distribución Uniforme

$$
\begin{aligned}
M_X(t) &= E(e^{tX}) = \int_{-\infty}^{\infty} e^{tX} f(x)dx, \\
M_X(t) &= \int_{\alpha}^{\beta} e^{tX}\left(\frac{1}{\beta-\alpha}\right)dx = \frac{1}{\beta-\alpha}\int_{\alpha}^{\beta} e^{tX}dx \\
&= \frac{1}{\beta-\alpha} * \frac{1}{t}\left(e^{tX}\right)_{\alpha}^{\beta} = \frac{1}{\beta-\alpha} * \frac{1}{t}\left(e^{t\beta} - e^{t\alpha}\right) \\
&= \frac{e^{t\beta}-e^{t\alpha}}{t(\beta-\alpha)}, \quad t \neq 0
\end{aligned} \tag{2.7}
$$

Esto es, $M_X(t)$ no está definida en t=0.

2.3 Prueba de Bondad de Ajuste Ji Cuadrada χ^2

La prueba de bondad de ajuste χ^2 se aplica a situaciones en las que queremos determinar si un conjunto de datos se puede considerar como una muestra aleatoria de una población que tiene una distribución dada. El contraste de hipótesis y el estadístico de prueba utilizados para éste análisis, se presentan en el Cuadro 2.1

Cuadro 2.1 *Efectos de la Imputación en el Análisis de Datos Multivariados* **Contraste de Hipótesis de la Prueba de Bondad de Ajuste**
H_0: La distribución de la población donde se obtuvo la muestra es $F_0(x)$ vs. H_1: No es verdad H_0 Estadístico de Prueba es : $\sum_{i=1}^{m}\frac{(f_i - e_i)^2}{e_i}$ que sigue una distribución χ^2 y con (m-p-1) grados de libertad

Elaborado por: G. Cuenca

Donde m es el número de términos en la suma y p es el número de parámetros que se estiman en el modelo con base en los datos muestrales.

Con el propósito de ilustrar esta prueba de hipótesis, considere si los siguientes números aleatorios provienen de una distribución Poisson con parámetro conocido $\lambda = 3$

Tabla 2.1
Efectos de la Imputación en el análisis de datos multivariados
Prueba de Bondad de Ajuste

i	Frecuencia Observada f_i	Probabilidad Poisson $\lambda = 3$	e_i
0	18	0.050	22.000
1	53	0.149	65.700
2	103	0.224	98.600
3	107	0.224	98.600
4	82	0.168	73.900
5	46	0.101	44.440
6	18	0.050	22.200
7	10	0.022	9.680
8	3	0.012	5.280

Elaborado por: G. Cuenca

Cuadro 2.2
Efectos de la Imputación en el análisis de datos multivariados
Prueba de Bondad de Ajuste

H_0: La distribución de la población donde se obtuvo la muestra es Poisson $\lambda = 3$

vs.

H_1: No es verdad H_0

Estadístico de Prueba es : $\sum_{i=1}^{m} \frac{(f_i - e_i)^2}{e_i}$ = 6.828

Valor p = 0,998

Elaborado por: G. Cuenca

De acuerdo a los resultados obtenidos mediante la prueba de bondad de ajuste, el valor p es 0.998, por lo tanto no existe evidencia estadística suficiente para rechazar la hipótesis nula, es decir, la distribución de la población donde se obtuvo la muestra es Poisson $\lambda = 3$.

2.4 Prueba de Kolmogorov-Smirnov

La prueba de bondad de ajuste KS es una alternativa a la prueba χ^2 que permite comprobar si una muestra aleatoria proviene de una población con una distribución dada, pero se prefiere el uso de la prueba KS en el caso de distribuciones continuas ya que esta prueba trabaja directamente sobre las observaciones y en cambio la prueba χ^2 trabaja sobre los datos agrupados.

Recordemos que dada una muestra aleatoria $X_1, X_2, ..., X_n$ y la muestra ordenada $X_{(1)}, X_{(2)}, ..., X_{(n)}$, la distribución empírica de la muestra es:

$$\hat{F}_n(x) = \begin{cases} 0 & si \quad X < X_{(1)} \\ \dfrac{k}{n} & si \quad X_{(k)} \leq X < X_{(k+1)};\ si\ k = 1,2,\ldots,n-1 \\ 1 & si \quad X \geq X_{(n)} \end{cases}$$

La prueba KS consiste en verificar el contraste de hipótesis:

Cuadro 2.3 *Efectos de la Imputación en el Análisis de Datos Multivariados* **Contraste de Hipótesis de la Prueba de Kolmogorov-Smirnov**
H_0: La distribución de la población donde se obtuvo la muestra es $F_0(x)$ vs. H_1: No es verdad $F_0(x)$ Estadístico de Prueba es : $max\left\|\hat{F}_n(x) - F_0(x)\right\|$ que sigue una distribución D y con (n, p) grados de libertad

Elaborado por: G. Cuenca

Con el propósito de ilustrar esta prueba de hipótesis, se tiene una matriz de datos cuyas columnas son muestras tomadas de cinco poblaciones todas ellas Normal, independientes e idénticamente distribuidas, con parámetros μ=0 y σ^2=1, $\mathbf{X} \in \mathrm{M}_{4x5}$, *i= 1,2,3* y *j= 1,2,3,4,5*

Tabla 2.2
Efectos de la Imputación en el análisis de datos multivariados
Matriz de Datos de variables aleatorias independientes con distribución Normal (0, 1)
Tamaño de muestra n=4

0.464	0.137	2.455	-0.323	-0.068
0.906	-0.513	-0.525	0.595	0.881
-0.482	1.678	-0.057	-1.229	-0.486
-1.787	-0.261	1.237	1.046	-0.508

Elaborado por: G. Cuenca

Tabla 2.3 *Efectos de la Imputación en el análisis de datos multivariados* **Prueba de Kolmogorov-Smirnov**			
X_n	$\hat{F}_{20}(x)$	$F_0(x)$	$max\left\|\hat{F}_n(x) - F_0(x)\right\|$
-1.787	1/20	0.037	0.013
-1.229	2/20	0.109	0.009
-0.525	3/20	0.299	0.149
-0.513	4/20	0.305	0.105
-0.508	5/20	0.305	0.056
-0.486	6/20	0.312	0.012
-0.482	7/20	0.316	0.034
-0.323	8/20	0.375	0.026
-0.261	9/20	0.397	0.053
-0.068	10/20	0.472	0.028
-0.057	11/20	0.476	0.074
0.137	12/20	0.556	0.044
0.464	13/20	0.677	0.027
0.595	14/20	0.726	0.026
0.881	15/20	0.811	0.061
0.906	16/20	0.819	0.014
1.046	17/20	0.853	0.003
1.237	18/20	0.893	0.008
1.678	19/20	0.954	0.004
2.455	20/20	0.993	0.006

Elaborado por: G. Cuenca

Cuadro 2.4 *Efectos de la Imputación en el análisis de datos multivariados* **Prueba de Kolmogorov-Smirnov**
H_0: La distribución de la población donde se obtuvo la muestra es Normal(0,1) vs. H_1: No es verdad H_0 Estadístico de Prueba es : $max\left\|\hat{F}_n(x) - F_0(x)\right\|$ = 0.149 Valor p = 0.766

Elaborado por: G. Cuenca

De acuerdo a los resultados obtenidos mediante la prueba de kolmogorov-Smirnov, el valor p es 0.766, por lo tanto no existe evidencia estadística suficiente para rechazar la hipótesis nula, es decir, la distribución de la población donde se obtuvo la muestra es Normal (0, 1).

2.5 Generación de Números Pseudo Aleatorios *U*(0,1)

Los números "pseudo aleatorios" son la base en la construcción de los modelos de simulación donde hay presencia de variables estocásticas, ya que estos permiten el funcionamiento de las abstracciones con los que un fenómeno que no se puede construir físicamente, sea numéricamente construìdo o recreado.

Existe un gran número de métodos que permiten la generación de números aleatorios entre 0 y 1, la importancia del método a utilizar radica en los números que genera, ya que estos números deben cumplir ciertas características para que sean válidos. Dichas características son:

- Ser uniformemente distribuidos.
- Ser estocàsticamente independientes lo cual significa que si X_1 y X_2 son dos variables aleatorias, X_1 y X_2, son independientes si y sólo si $f_{12}(X_1, X_2) = f_1(X_1) f_2(X_2)$; siendo f_{12} la distribución conjunta de X_1 y X_2 y además f_1 y f_2 las marginales de X_1 y X_2 respectivamente.
- Además es recomendable que los períodos del generador sean "largos" es decir sin repetición dentro de una longitud determinada de la sucesión de valores generados. [2]

2.5.1 Generadores Congruenciales Lineales

La generación de números pseudos aleatorios se realiza a través de una "relación de recurrencia", es decir para una sucesión $X_0, X_1, ..., X_n$, es una expresión que define a cada término X_n, en función de uno o más de los términos que le preceden. Los valores de los términos necesarios para empezar a calcular se llaman condiciones iniciales. Se han propuesto varios esquemas como los *métodos congruenciales*: *congruencial mixto* y *congruencial multiplicativo.* [2]

2.5.1.1 Método Congruencial Mixto

El *Método Congruencial Mixto* genera una sucesión de números pseudo aleatorios en la cual el sucesor X_{n+1} del número pseudo aleatorio X_n es determinado justo a partir de X_n. Particularmente para el caso del generador congruencial mixto la relación de recurrencia es la siguiente:

$$X_{n+1} = (aX_n + c) \bmod m \, , \qquad (2.8)$$

Donde

X_0 > 0: representa la semilla y es un valor que elige el investigador;

a > 0 : se denomina multiplicador;

c > 0 : es una constante aditiva la que se denomina incremento;

m es el “módulo”, siendo; $m > X_0$, $m > a$ y además $m > c$

Esta “relación de recurrencia” nos dice que X_{n+1} es el residuo de dividir $aX_n + c$ para el módulo. Es decir que los valores posibles de X_{n+1} son $0, 1, 2, 3, \ldots, m-1$, tal que, m representa el número posible de valores diferentes que pueden ser generados. [2]

Para ilustrar la generación de números pseudoaleatorios por medio de este método, suponga que se tiene un generador en el cual los valores de sus parámetros son: $a = 5$, $c = 7$, $X_0 = 4$ y $m = 8$.

Como se puede apreciar en la Tabla 2.4 el “período del generador” es ocho, esto es la sucesión se repite una vez que se obtuvo el octavo número generado

TABLA 2.4
Efectos de la Imputación en el análisis de datos multivariados
Método Congruencial Mixto
Números pseudos aleatorios del generador $X_{n+1} = (5X_n + 7) \bmod 8$

n	X_n	(5X_n+7)/8	X_{n+1}	Números Uniformes
0	4	3+3/8	3	0.375
1	3	2+6/8	6	0.750
2	6	4+5/8	5	0.625
3	5	4+0/8	0	0.000
4	0	0+7/8	7	0.875
5	7	5+2/8	2	0.250
6	2	2+1/8	1	0.125
7	**1**	**1+4/8**	**4**	**0.500**
8	4	3+3/8	3	0.375
9	3	2+6/8	6	0.750
10	6	4+5/8	5	0.625
11	5	4+0/8	0	0.000
12	0	0+7/8	7	0.875

Elaborado por: G. Cuenca

El procedimiento para obtener los números pseudo aleatorios se realiza de la siguiente forma, donde la semilla es 4:

$$X_{n+1} = (5X_n + 7) \bmod 8$$

si $n = 0$

$$\begin{aligned} X_1 &= (5X_0 + 7) \bmod 8 \\ &= \frac{5(4) + 7}{8} = \frac{27}{8} = 3 + \frac{3}{8} = 3.375 \end{aligned}$$

donde $\frac{3}{8}$ es el residuo y al dividir 3 para 8, el resultado es el número uniforme 0.375.

si $n=1$

$$X_2 = (5X_1 + 7)\bmod 8$$
$$= \frac{5(3)+7}{8} = \frac{22}{8} = 2 + \frac{6}{8} = 2.75$$

donde el residuo es $\frac{6}{8}$ y al dividir 6 para 8, el resultado es el número uniforme 0.750.

si $n=2$

$$X_3 = (5X_2 + 7)\bmod 8$$
$$= \frac{5(6)+7}{8} = \frac{37}{8} = 4 + \frac{5}{8} = 4.62$$

donde el residuo es $\frac{5}{8}$ y al dividir 5 para 8, el resultado es el número uniforme 0.625

y así sucesivamente se calculan los restantes cinco números uniformes de la sucesión, los cuales son: 0.000, 0.875, 0.250, 0.125 y 0.500.

Al analizar este ejemplo se podría pensar que el período de todo generador es siempre igual a m. Sin embargo, esto no es verdad ya que el período del generador depende de los valores asignados a los parámetros a, c, X_0 y m, es decir, se requiere seleccionar valores adecuados para estos parámetros con el fin de que el generador tenga "período largo".

Con el fin de ilustrar el caso que se presenta cuando el *período del generador* es menor que m, suponga que se tiene un caso en el cual los valores de los parámetros son: $a = X_0 = c = 7$ y $m = 10$. Para estos valores, la sucesión de números pseudo aleatorios y uniformes son mostrados en la Tabla 2.5. Se puede apreciar que el período del generador es cuatro, lo cual deja claro que una selección inadecuada de los valores de los parámetros del generador, puede conducirnos a obtener períodos indeseables.

TABLA 2.5

Efectos de la Imputación en el análisis de datos multivariados

Método Congruencial Mixto

Números pseudoaleatorios del generador $X_{n+1} = (7X_n + 7) \mathrm{mod}\, 10$

n	X_n	(7X_n+7)/10	X_{n+1}	Números Uniformes
0	7	5+6/10	6	0.600
1	6	4+9/10	9	0.900
2	9	7+0/10	0	0.000
3	**0**	**0+7/10**	**7**	**0.700**
4	7	5+6/10	6	0.600
5	6	4+9/10	9	0.900
6	9	7+0/10	0	0.000

Elaborado por: G. Cuenca

De la misma forma que el ejemplo anterior los números pseudo aleatorios se obtienen de la siguiente manera, donde la semilla es 7:

$$X_{n+1} = (7X_n + 7) \mathrm{mod}\, 10$$

si $n = 0$

$$\begin{aligned} X_1 &= (7X_0 + 7) \mathrm{mod}\, 10 \\ &= \frac{7(7)+7}{10} = \frac{56}{10} = 5 + \frac{6}{10} = 5.600 \end{aligned}$$

donde $\frac{6}{10}$ es el residuo y al dividir 6 para 10, el resultado es el número uniforme 0.600

si $n = 1$

$$\begin{aligned} X_2 &= (7X_1 + 7) \mathrm{mod} 10 \\ &= \frac{7(6)+7}{10} = \frac{49}{10} = 4 + \frac{9}{10} = 4.900 \end{aligned}$$

donde $\frac{9}{10}$ es el residuo y al dividir 9 para 10, el resultado es el número uniforme 0.900.

si $n = 2$

$$\begin{aligned} X_3 &= (7X_2 + 7) \mathrm{mod}\, 10 \\ &= \frac{7(9)+7}{10} = \frac{70}{10} = 7 + \frac{0}{10} = 7.000 \end{aligned}$$

donde $\frac{0}{10}$ es el residuo y al dividir 0 para 10, el resultado es el número uniforme 0.000

si $n = 3$

$$X_4 = (7X_3 + 7) \bmod 10$$
$$= \frac{7(0)+7}{10} = \frac{7}{10} = 0 + \frac{7}{10} = 0.700$$

donde $\frac{7}{10}$ es el residuo y al dividir 7 para 10, el resultado es el número uniforme 0.700.

Azarang y García expresan:

"Se advierte la necesidad de establecer algunas reglas que pueden ser utilizadas en la selección de los valores de los parámetros, para que el generador resultante tenga período completo". [1]

El valor apropiado del módulo m debe ser el número entero más grande que la computadora acepte, el multiplicador a debe ser un entero impar no divisible para 3 ó 5, la constante aditiva c, puede ser cualquier constante y el valor de la semilla X_0, es irrelevante, para el generador congruencial mixto, es decir, el valor de este parámetro resulta tener poca o ninguna influencia sobre las propiedades estadísticas de las sucesiones.

2.5.1.2 Método Congruencial Multiplicativo

El *Método Congruencial Multiplicativo* al igual que el congruencial mixto genera una sucesión de números pseudos aleatorios en la cual el sucesor X_{n+1} del número pseudo aleatorio X_n es determinado justo a partir de X_n, de acuerdo a la siguiente relación de recurrencia:

$$X_{n+1} = aX_n \bmod m \,, \tag{2.9}$$

Al igual que el generador anterior, en éste también se debe seleccionar adecuadamente los valores de los parámetros a, X_0 y m, con el fin de asegurar un período largo para las sucesiones generadas por este método.

Para ilustrar la obtención del período de un generador utilizando el *Método Congruencial Multiplicativo*, suponga que se tiene un generador con los siguientes parámetros: $a = 5$, $X_0 = 5$ y m=32. Estos valores se muestran en la Tabla 2.6.

TABLA 2.6
Efectos de la imputación en el análisis de datos multivariados
Método Congruencial Multiplicativo
Números pseudoaleatorios del generador $X_{n+1} = 5X_n \bmod 32$

n	X_n	$5X_n$ / 32	X_{n+1}	Números Uniformes
0	5	0+25/32	25	0.781
1	25	3+29/32	29	0.906
2	29	4+17/32	17	0.531
3	17	2+21/32	21	0.656
4	21	3+9/32	9	0.281
5	9	1+13/32	13	0.406
6	13	2+1/32	1	0.031
7	**1**	0+5/32	**5**	**0.156**
8	5	0+25/32	25	0.781
9	25	3+29/32	29	0.906
10	29	4+17/32	17	0.531
11	17	2+21/32	21	0.656

Elaborado por: G. Cuenca

Se puede apreciar en Tabla 2.6 que el período del generador es ocho, esto es la sucesión se repite una vez que se obtuvo el octavo número generado.

El procedimiento para obtener los números pseudo aleatorios se realiza de la siguiente forma, donde la semilla es 5.

$$X_{n+1} = 5X_n \bmod 32$$

si $n = 0$

$$X_1 = 5X_0 \bmod 32$$

$$= \frac{5(5)}{32} = \frac{25}{32} = 0 + \frac{25}{32} = 0.781$$

donde $\frac{25}{32}$ es el residuo y al dividir 25 para 32, el resultado es el número uniforme 0.781.

si $n=1$

$$X_2 = 5X_1 \bmod 32$$
$$= \frac{5(25)}{32} = \frac{125}{32} = 3 + \frac{29}{32} = 3.906$$

donde el residuo es $\frac{29}{32}$ y al dividir 29 para 32, el resultado es el número uniforme 0.906.

si $n=2$

$$X_3 = 5X_2 \bmod 32$$
$$= \frac{5(29)}{32} = \frac{145}{32} = 4 + \frac{17}{32} = 4.531$$

donde el residuo es $\frac{17}{32}$ y al dividir 17 para 32, el resultado es el número uniforme 0.531.

y así sucesivamente se calculan los restantes cinco números de la uniformes de la sucesión, los cuales son: 0.656, 0.281, 0.406, 0.031 y 0.156.

2.6 Métodos de Generación de Variables Aleatorias No Uniformes

Generalmente en las simulaciones de sistemas estocásticos existen una o varias variables aleatorias interactuando, las cuales siguen distribuciones diferentes a la distribución uniforme. Por consiguiente, para simular este tipo de variables, es necesario contar con un generador de números uniformes y una función que transforme estos números uniformes en valores de la distribución de probabilidad deseada. Para esto, se suele utilizar el método de la transformada inversa. [1]

2.6.1 Método de la Transformada Inversa

El "Método de la Transformada Inversa" utiliza la distribución acumulada $F(x)$ de una variable aleatoria X que se va a simular. Puesto que $F(x)$ está definida en el intervalo (0,1), y que además $F(x)=x$ para $x \in (0,1)$ se puede generar un número aleatorio uniforme y y tratar de determinar el valor de la variable aleatoria para la cual su distribución acumulada es igual a y. Recordemos que F es una función sobreyectiva e inyectiva y por tanto un isomorfismo, además $\lim_{x \to -\infty} F(x) = 0$ y $\lim_{x \to \infty} F(x) = 1$.

Para convertir a un valor x, tomado de una distribución específica, a partir de un valor uniforme, se deberá encontrar y en términos de x, a partir de:

$$
\begin{aligned}
&F(x) = y \\
&F^{-1}(F(x)) = F^{-1}(y) \\
&x = F^{-1}(y)
\end{aligned}
\tag{2.10}
$$

Este método tiene la dificultad principal de que en algunas ocasiones es difícil encontrar la transformada inversa. Sin embargo, si esta función inversa ya ha sido establecida, generando números aleatorios uniformes se podrán obtener valores de la variable aleatoria que sigan la distribución de probabilidad deseada.

2.6.1.1 Distribución exponencial

Utilizando el "Método de la Transformada Inversa" a continuación se desarrolla un generador de variables aleatorias con distribución exponencial. La función de densidad de probabilidad de una variable aleatoria exponencial con parámetro β es:

$$f(x)=\begin{cases} \dfrac{1}{\beta}e^{-x/\beta} & \text{si } x>0 \\ 0 & \text{si } x\leq 0 \end{cases} \tag{2.11}$$

Su función acumulada F es:

$$P(X\leq x)=F(x)=\begin{cases} 1-e^{-x/\beta} & \text{si } x>0 \\ 0 & \text{si } x\leq 0 \end{cases} \tag{2.12}$$

Aplicando el método de la transformada inversa, se tiene que si X es exponencial con parámetro β y y uniforme con parámetro 0 y 1:

$$1-e^{-x/\beta}=y$$
$$e^{-x/\beta}=1-y$$

$$\ln e^{-x/\beta} = \ln(1-y)$$

$$x = -\beta \ln(1-y) \qquad (2.13)$$

Para ilustrar este generador de variables aleatorias con distribución exponencial, sea $X \sim E(1,\beta)$, tal que $\beta = 2$ y $Y \sim U(0,1)$, si por ejemplo y=0.25.

$$x = -2 \ln(1-0.25)$$
$$x = 0.575$$

Este es un valor tomado de una población exponencial.

2.6.2 Procedimientos Especiales

Existen algunas distribuciones como las distribuciones normal, binomial, poisson, etc. cuya simulación a través del método de transformada inversa resulta complicada. Para estas distribuciones es posible utilizar algunas de sus propiedades para facilitar y agilizar el proceso de generación de números aleatorios.

2.6.2.1 Variables que siguen una Distribución Normal

Puesto que no es posible expresar la distribución acumulada de la distribución normal en forma explícita, por ende no se puede utilizar el método de la transformada inversa.

$$f(x) = \frac{1}{\sqrt{2\pi\sigma}} e^{\frac{-1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \qquad \infty < x < \infty \qquad (2.14)$$

Entonces si se desea generar números aleatorios que sigan una Distribución Normal con parámetros conocidos μ y σ^2, se puede hacer uso del Teorema del Límite Central el cual establece que la suma de n variables aleatorias independientes se aproxima a una distribución Normal con media μ y varianza σ^2, a medida que n se aproxima al infinito.

Si por ejemplo, $X_1, X_2, ..., X_n$ es una sucesión de n variables aleatorias independientes, tal que $X_i \sim U(0, 1)$, i=1,2,…n

Para la distribución uniforme con parámetros $\alpha = 0$ y $\beta = 1$ se conoce que:

$$\mu = \frac{1}{2} \quad \text{y} \quad \sigma^2 = \frac{1}{12}$$

Aplicando el Teorema del Límite Central a los X_i tenemos que:

$$Z = \frac{\sum_{i=1}^{n} X_i - n\mu}{\sqrt{n/12}} \tag{2.15}$$

Donde Z es un número aleatorio tomado de una población que sigue una distribución Normal con parámetros $\mu = 0$ y $\sigma^2 = 1$.

Para $n = 12$ se obtienen buenas aproximaciones, entonces se tiene que:

$$Z = \frac{\sum_{i=1}^{12} X_i - 12\mu}{\sqrt{12}.\frac{1}{\sqrt{12}}} = \frac{\sum_{i=1}^{12} X_i - 12(1/2)}{\sqrt{12}.\frac{1}{\sqrt{12}}} = \left(\sum_{i=1}^{12} X_i\right) - 6 \qquad (2.16)$$

Recuérdese que X_i ~ *U*(0, 1)

Si se desea obtener números aleatorios X que sigan una distribución Normal con media μ y varianza σ^2, se parte del siguiente resultado, que se relaciona con la distribución normal estándar esto es, $X \sim \mathrm{N}(\mu, \sigma^2)$, entonces $Z = \frac{X-\mu}{\sigma}$ es normal estándar.

2.6.2.2 Variables que siguen una Distribución Poisson

Una variable aleatoria X tiene una distribución poisson si y sólo si su distribución de probabilidades está dada por:

$$P(X = x) = \frac{\lambda^x e^{-\lambda}}{x!} \quad \text{para } x = 0, 1, 2, \ldots \qquad (2.17)$$

Entonces la generación de números al azar que sigan una distribución poisson, se lo puede hacer aplicando el método de la transformada inversa.

$$p_{i+1} = P(X = i) = \frac{\lambda}{i+1} p_i; \quad i \geq 0 \qquad (2.18)$$