Capítulo 3 MÉTODOS PARA PROBAR NUMEROS ALEATORIOS

III.1 Introducción

Existen algunos métodos disponibles para verificar varios aspectos de la calidad de los números pseudoaleatorios. Si no existiera un generador particular de números aleatorios disponible, se le recomienda al analista usar estos métodos cuando se realice una simulación.

Las dos propiedades mas importantes esperadas en los números aleatorios son uniformidad e independencia. La prueba de uniformidad puede ser realizada usando las pruebas de ajuste de bondad disponibles. Por ejemplo, un numero estadístico suficiente de números aleatorios pueden ser usados para verificar la distribución de los números contra la distribución uniforme teórica usando ya sea el método **Chi-Cuadrada** o el método **Kolmogomorov-Smirnov**(**KS)** para números aleatorios. Este tipo de prueba es denominada "Prueba de frecuencia".

Los números pueden estar uniformemente distribuidos y aun no ser independientes uno del otro. Por ejemplo, una secuencia de números monótonamente se incrementa dentro del rango de cero a uno esta uniformemente distribuida si la cantidad incremental es constante para todos. (0, 0.1, 0.2, 0.3,......,0.9).

La prueba de Frecuencias es utilizada para comprobar que los datos estén Uniformemente distribuidos.

La prueba de autocorrelación checa la correlación entre números aleatorios y los compara con la deseable correlación de cero.

La prueba GAP (de huecos o de distancia) es usada para asegurar que la recurrencia de cada dígito particular en un flujo de números suceda con un intervalo aleatorio. La prueba KS es entonces usada para comparar estos intervalos con la longitud esperada de huecos.

La prueba Póquer, prueba grupos de números juntos como una mano de poker y compara cada mano con la mano esperada usando la prueba Chi-cuadrada.

La prueba de corrida arriba abajo es generalmente la prueba principal usada para verificar la dependencia. Esta prueba detecta si un patrón inaceptable estadísticamente que se incrementa o decrece existe entre números adyacentes en un flujo de números.

Prueba de Series. Mide la correlación entre elementos adyacentes en una secuencia de números aleatorios.

Los números aleatorios no deben contener ningún patrón concebible. Obviamente, uno puede concebir un número grande de patrones posibles entre números, y una prueba especial puede ser creada para detectar cada patrón particular. No es recomendable ni practico realizar todas las pruebas para verificar la confiabilidad del generador números aleatorios. Se debe de comparar el costo de realizar estas pruebas con el costo de resultante de la imperfección del generador en un proyecto de simulación actual.

III.2 Prueba de Frecuencias

Una prueba básica que siempre será desarrollada para validar un nuevo generador es la prueba de uniformidad. Dos métodos de pruebas disponibles. Estas son las pruebas Kolmogorov-Smirnov y la prueba Chi-Cuadrada Ambas de estas pruebas miden el grado de ajuste entre la distribución de una muestra de números aleatorios generados y y la distribución uniforme teórica. Ambas de estas pruebas están basadas en la Hipótesis Nula de que no existe diferencia entre la distribución de la muestra y la distribución teórica.

III.2.1 La prueba de Kolmogorov-Smirnov.

Esta prueba compara la pdf (función de densidad de probabilidad), F(x), de la distribución uniforme con el pdf empírico, $S_n(x)$, de una muestra de N observaciones.

Por definición

$$F(x) = x, \qquad 0 \le x \le 1$$

Conforme N crece, $S_N(x)$ deberá tener una mejor aproximación de F(x),, dado que la hipótesis nula sea verdadera.

$$S_N(x) = \frac{numero \;\; de \;\; R_1, R_2, ...R_N \;\; donde \;\; son \;\; \le x}{N} \qquad (1)$$

La prueba Kolmogorov-Smirnov esta basada en la desviación máxima absoluta entre F(x) y $S_N(x)$ sobre el rango de e la variable aleatoria- Esto es, basado en la estadística

$$D = \max \ \left| F(x) - S_n(x) \right| \qquad (2)$$

La distribución de la muestra D es conocida y es tabulada como una función de N en la tabla Kolmogorov-Smirnov. Para probar contra una pdf uniforme, el procedimiento sigue los pasos siguientes:

Paso 1: Ordene los datos en forma ascendente. Sea R_i , el la i^{va} más pequeña observación, tal que

Paso 2: Usando la fdp teórica $R_1 \leq R_2 \leq ...... \leq R_N$ F(x), calcule

$$D^{+} = \max_{1 \leq i \leq N} \left[\frac{i}{N} - R_i \right] \quad (3) \qquad D^{-} = \max_{1 \leq i \leq N} \left[R_i - \frac{i-1}{N} \right] \quad (4)$$

Paso 3: Calcule D = max(D^+, D^-)
Paso 4: Encuentre el valor crítico Dα de la tabla KS para un nivel de significancia y un tamaño de muestra N.
Paso 5: Si D ≤ al valor crítico Dα, acepte la distribución candidato como aquella que tiene un buen ajuste a los datos observados; de otra forma rechace.

Esta prueba esta basada en la desviación absoluta mayor entre las fdp empírica y teórica para todo valor dado de x. Esta desviación es comparada con los valores críticos de KS tabulados para determinar si la desviación puede ser atribuida a los efectos aleatorios y por lo tanto sea una distribución candidato a ser aceptada tener un buen ajuste a los datos observados. Más específicamente, la prueba tiene los pasos siguientes:

Ejemplo:
En este ejemplo se usa la prueba KS para examinar bajo un nivel de significancia de α=0.05 si un conjunto de datos representa números aleatorios (por ejemplo esta la distribución uniforme entre 0 y 1). Suponga que cinco datos son dados: 0.53, 0.35, 0.03, 0.94, y 0.22

Solución. Para la distribución Uniforme la fdp es F(x)= 1/(b-a) a≤x≤b
Para este caso particular a=0 y b=1. Por lo tanto F(x)=x. Ahora se ordenan los valores en forma ascendente y se realizan los cálculos relativos.

La tabla siguiente resume los cálculos realizados:

i	$F(x_i)$	i/n	i/n - $F(x_i)$	$F(x_i)$ – (i-1)/n
1	0.03	0.20	0.17	0.03
2	0.22	0.40	0.18	0.02
3	0.35	0.60	0.25	-0.05
4	0.53	0.80	0.27	-0.07

5	0.94	1.00	0.06	-.14
			D^+= 0.27	D^-=0.14

De acuerdo a los cálculos, D = max(0.27, 0.14) = 0.27. El valor crítico de KS de la tabla en el **apéndice de tablas** para un tamaño de 5 y un nivel de significancia de 0.05 es 0.565. Debido a que D es menor que este valor crítico, la hipótesis de que los datos dados pertenecen a una distribución Uniforme es aceptada.

III.2.2 Prueba Chi-Cuadrada

La prueba Chi-Cuadrada en lugar de medir la diferencia de cada punto entre la muestra y la desviación verdadera, checa la desviación del valor esperado.

$$X^2_{calculada} = \sum_{i=1}^{n} \frac{(O_i - E_i)^2}{E_i} \qquad (5)$$

Donde n es el número de intervalos de clase (ejemplo: O_i es el número observado en la clase i^{va}, y E_i es el número esperado en cada clase i^{va} , y n es el número de clases. Para una distribución uniforme, E_i , el número en cada clase esta dado por;

$$E_i = \frac{N}{n} \qquad (6)$$

Para clases igualmente espaciadas, donde N es el número total de observaciones. Puede ser mostrado que la distribución de la muestra Chi-Cuadrada esta aproximadamente a la distribución Chi-Cuadrada con n-1 grados de libertad.

Ejemplo:

Use la prueba Chi-Cuadrada con α=0.05 para probar si los datos dados a continuación en la **tabla 1** están uniformemente distribuidos.

Tabla 1

0.34	0.9	0.25	0.89	0.87	0.44	0.12	0.21	0.46	0.67
0.83	0.76	0.79	0.64	0.7	0.81	0.94	0.74	0.22	0.74
0.96	0.99	0.77	0.67	0.56	0.41	0.52	0.73	0.99	0.02
0.47	0.3	0.17	0.82	0.56	0.05	0.45	0.31	0.78	0.05
0.79	0.71	0.23	0.19	0.82	0.93	0.65	0.37	0.39	0.42
0.99	0.17	0.99	0.46	0.05	0.66	0.1	0.42	0.18	0.49
0.37	0.51	0.54	0.01	0.81	0.28	0.69	0.34	0.75	0.49
0.72	0.43	0.56	0.97	0.3	0.94	0.96	0.58	0.73	0.05
0.06	0.39	0.84	0.24	0.4	0.64	0.4	0.19	0.79	0.62

0.18 0.26 0.97 0.88 0.64 0.47 0.6 0.11 0.29 0.78

Haciendo 10 intervalos de 0 a 1 con incrementos de .1 (de igual longitud) tenemos la tabla siguiente:

Intervalo i	Ei	Oi	$\frac{(O_i - E_i)^2}{E_i}$
1	10	8	0.4
2	10	8	
3	10	10	
4	10	9	
5	10	12	
6	10	8	
7	10	10	
8	10	14	
9	10	10	
10	10	11	
TOTAL	100		$X^2c = \sum_{i=0}^{n} \frac{(O_i - E_i)^2}{E_i} = 3.4$

El valor de X^2 en el apéndice de tablas es $X^2_{calcualda}$=3.4. Esto comparado con el valor crítico $X^2_{0.05,9}$=16.9. Debido a que $X^2_{calculada}$ < que el valor de $X^2_{0.05,9}$ de la tabla, la hipótesis Nula de que no existe diferencia entre la distribución de la muestra y la distribución uniforme se Acepta.

III.3 Prueba de Autocorrelación

Correlación es la relación reciproca entre dos o mas cosas (elementos). A veces un grupo de números generados pueden parecer aleatorios, pero existe una relación entre cada cierto números de ellos a partir de alguno específico.

Amplitud de autocorrelación: Es la distancia que existe entre los números de la lista que tiene la relación entre sí. Se da cada n-ésimo número aleatorio e inicia en el elemento i.

Esta prueba se aplica con la suposición de los números aleatorios tiene una distribución uniforme e independiente sobre el intervalo de 1 a 0.

Conceptos y parámetros que usamos en autocorrelación

Para analizar la correlación general para todos los pares sucesivos de números aleatorios se utiliza la estadística:

Densidad de probabilidad

$$\rho_{im} = \frac{1}{M+1}\sum_{k=0}^{m} r_{(i+km)} * r_{[i+(k+1)m]} \qquad (7)$$

Donde:

N es el total de números en toda la serie; Tamaño de la muestra.

i es el primer numero donde empieza la amplitud de autocorrelación**.**

m es la amplitud de la autocorrelación **.**

M es el entero mayor tal que i+(M+1)*m<N

Este valor, se obtiene de acuerdo a los valores dados cuidando que se cumpla la condición. Es un parámetro de la formula:

$$M = Truncar\left\{\frac{(N-i)}{m}\right\} - 1 \qquad (8)$$

Cumpliéndose la condición: i + (M + 1) m < M

Desviación estándar de la autocorrelación

$$\sigma_{\rho im} = \frac{\sqrt{13M+7}}{12(M+1)} \qquad (9)$$

$\sigma\rho_{im}$ (Desviación estándar de la densidad de probabilidad.)

La estadística para determinar la significancia de la autocorrelación para la secuencia propuesta de M+1 números es:

$$Z = \frac{\rho_{im} - 0.25}{\sigma_{\rho im}} \qquad (10)$$

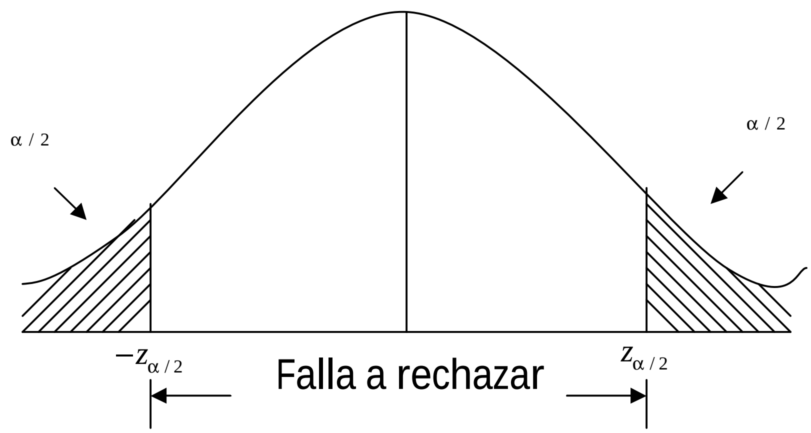
Z significancia de la autocorrelación que tiene una distribución Normal, con media cero y una varianza de uno, bajo la suposición de independencia.

Nivel de significancia

Si se define el nivel de significancia por medio de α y $Z_{1-\alpha/2}$ el valor de Z hace que:

$$\mathbf{P(Z >= Z_{1-\alpha/2}) = \alpha / 2} \qquad (11)$$

Se utiliza (α **/ 2** puesto que se va a tomar en cuenta ambos lados del área bajo la curva)



Para determinar la autocorrelación se establecen las siguientes Hipótesis;

Hipótesis Nula
$H_0=\rho_{im}=0$ Los números aleatorios están correlacionados (No son Aleatorios)

Hipótesis Alternativa
$H_1=\rho_{im}\neq 0$ Los números aleatorios No están correlacionados (Sí son aleatorios)

Criterio de rechazo $|Z_0| > Z_{1-\alpha/2}$

Entonces , si: $|Z| > Z_{1-\alpha/2}$ se rechaza la hipótesis de aleatoriedad..

y si $|Z| \leq Z_{1-\alpha/2}$ Se acepta la hipótesis de aleatoriedad.

Ejemplo 1

Tenemos la Siguiente serie de Números:

0.20,0.96,0.78,0.18,0.09,0.80,0.02,0.53,0.05,0.30,0.70,0.59,0.98,0.03,0.37,0.86,0.73,0.06,0.53,0.25,0.67,0.78,0.33,0.97,0.63,0.25,0.33,0.72,0.91,0.00,0.24,0.64,0.90,0.08,0.33,0.94,0.33,0.16,0.45,0.70,0.18,0.07

A la primer vista, estos números pueden parecer aleatorios. No obstante, al examinar de cerca estos números se ve que existe una relación clara entre cada sexto numero, a partir del segundo. Cada uno de estos números varía en magnitud sucesivamente de muy grande a muy pequeño.

0.20,**0.96**,0.78,0.18,0.09,0.80,**0.02**,0.53,0.05,0.30,0.70,0.59,**0.98,**0.90,0.03,0.37,0.86,0.73,**0.06**,0.53,0.25,0.67,0.78,0.33,**0.97**,0.63,0.25,0.33,0.72,0.91,**0.00,**0.24,0.64,0.90,0.08,0.33,**0.94**,0.33,0.16,0.45,0.70,0.18,**0.07**.

Ejemplo 2
Determínese si el segundo, el séptimo, el doceavo, y el vigésimo segundo de los números aleatorios de la secuencia que sigue están autocorrelacionados.

Sea $\alpha = 0.1$
0.13,0.91,0.11,0.02,0.65,0.33,0.86,0.05,0.25,0.28,0.80,0.82,0.10,0.78,0.88,0.76,0.29,0.20,0.66,0.17,0.71,0.45,0.40,0.35.

Puesto que nos interesa el grado de autocorrelación de cada quinto numero a partir del segundo, i=2,m=5,N=25 y M= 3.

$$\rho_{25} = \frac{1}{4}\sum_{k=0}^{3} r_{(2++5k)} * r_{(2+5*(k+1)}$$

$$\rho_{25} = \frac{1}{4}((0.91)(0.86)+0.86)(0.80)+(0.80)(0.76)+(0.76)(0.71)) = 0.6546$$

$$\sigma\rho_{im} = \frac{\sqrt{13*\mathrm{M}+7}}{12*(\mathrm{M}+1)}$$

$$\rho_{2\,5} = \frac{\sqrt{(139*3+7}}{12(3+1)} = 0.141$$

$$Z=(\rho_{im} - 0.25) \,/\, \frac{\sqrt{13*\mathrm{M}+7}}{12*(\mathrm{M}+1)}$$

Z= (0.6546-0.25)/0.141=2.87

Z= 2.87 > $Z_{.95}$= 1.316 por lo tanto se rechaza la hipótesis nula y por lo tanto se considera que los datos **son aleatorios.**

Ejemplo 3
Dados los siguientes números aleatorios

.12, .01, .23, .28, .89, .31, .64, .28, .83, .93, .99, .15, .33, .35, .91, .41, .6, .27, .75, .88, .68, .49, .05, .43, .95, .58, .19, .36, .69, .87

Determine si el 3°, 8°, 13° y los siguientes números en la secuencia están autocorrelacionados. Use

α=.05, i=3 (iniciando con el 3er. Número), m=5 (cada 5 números), N=30, y M=4 (entero mayor tal que 3+(m+1)5≤30). Entonces;

P_{35}=1/5[(8.23)(.28)+(.28)(.33)+(.33)(.27)+(.27)(.05)+(.05)(.36)=0.05548

$$\sigma_{\rho 35} = \frac{\sqrt{(13)4+7}}{12(4+1)} = .1280$$

$$Z = \frac{0.05548 - 0.25}{0.1280} = \frac{-0.19452}{0.1280} = -1.5196875$$

$Z_{0.975}$= 1.96

Z= -1.5196 < $Z_{.975}$= 1.64 por lo tanto se acepta la hipótesis nula y por lo tanto se considera que los datos No **son aleatorios.**

III.4 Pruebas de Huecos

La prueba de huecos (GAP**)** es usada para asegurar que la recurrencia de cada dígito particular en un flujo de números suceda con un intervalo aleatorio. Se pueden usar dos pruebas para comparar estos intervalos con la longitud esperada de los huecos:
La prueba Chi-Cuadrada (**χ2**) y la prueba Kolmogorov – Smirnov (KS) es entonces usada para comparar

III.4.1 La prueba Kolmogorov – Smirnov (KS)

Para determinar si los números aleatorios generados cumplen con las propiedades especificadas (uniformidad e independencia) se tendrán las hipótesis siguientes :

H_0 si $D_{calculada}$ < $D_{confiabilidad}$; se aprueba que los dígitos están ordenados aleatoriamente.

H_1 si $D_{calculada}$ > $D_{confiabilidad}$; se rechaza que los dígitos están ordenados aleatoriamente.

La prueba de huecos se utiliza para determinar la significancia de los intervalos entre la repetición de cierto dígito. Si el dígito *k* va seguido por *x* dígitos distintos de *k*, antes de que vuelva a parecer *k*, se dice que existe un hueco de tamaño *x*. Por ejemplo:

4, 8, 9, 7, 9, 8, 3, 3, 3, 9, 9, 0, 6, 3, 0, 3, 3, 4, 3, 5, 5, 8, 2, 9, 5, 5, 2, 5, 1, 5, 4, 8, 7, 9, 0, 6, 4, 8, 9,
2, 3, 9, 6, 0, 1, 5, 6, 8, 7, 7, 0, 9, 9, 7, 6, 3, 6, 3, 3, 5, 2, 7, 4, 0, 3, 1, 1, 4, 4, 2, 3, 4, 0, 4, 6, 0, 2, 7,
8, 5, 6, 8, 4, 0, 8, 8, 5, 0, 6, 5, 2, 7, 6, 6, 3, 9, 4, 6, 9, 1, 8, 9, 4, 5, 0, 2, 0, 4, 8, 1, 4, 5, 0, 2, 8.

Se puede tomar cualquier números aleatorio; en este caso se toma el número cero, el cual aparece 13 veces y por ende habrá 12 huecos. El primero de longitud 2, el segundo de 19, el tercero de 8, etc. Otro ejemplo tomamos el número cuatro, el cual aparece 15 veces y tendrá 14 huecos. El primero de longitud 16, el segundo de 12, el tercero de 5, etc. Para fines de esta prueba, nos interesa la frecuencia con la que se presentan los diversos huecos.

Para una secuencia dada de dígitos, anotamos el número de veces que aparecen los huecos de longitudes 0, 1, 2,.... . Podemos aplicar este procedimiento a un dígito simple entre 0 y 1. Después de tomar nota de la frecuencia con que aparece cada hueco, comparemos la frecuencia acumulativa relativa (S_x) observada con la frecuencia acumulativa teórica. Suponiendo que los dígitos están ordenados aleatoriamente, la distribución de frecuencias acumulativas relativas está dada por:

$$S(x) = (m / T) \qquad \text{donde: } m \text{ es frecuencia del hueco}$$

$$T \text{ Total de huecos}$$

Y la distribución de frecuencias acumulativas teóricas está dada por:

- La probabilidad de un hueco de una cierta longitud puede ser determinada por una prueba Bernoulli.

$$P(hueco\ de\ n) = P(x \neq 3)P(x \neq 3)....P(x \neq 3)P(x = 3)$$

Si únicamente consideramos dígitos del 0 al 9, entonces;

$$P(hueco\ de\ x) = (0.9)^x 0.1 \quad para \quad x = 0,1,2... \qquad (13)$$

Teóricamente la distribución de frecuencia para dígitos ordenados aleatoriamente esta dada por;

$$P(hueco \leq x) = F(x) = 0.1\sum_{n=0}^{x}(0.9)^n = 1 - 0.9^{x+1} \qquad (14)$$

Ejemplo

Basándonos en la frecuencia con que se producen los huecos, determínese si se puede suponer que los dígitos están ordenados aleatoriamente. Sea el nivel de significancia de $\alpha = 0.05$.

2, 9, 3, 1 , 6, 3, 0, 4, 6, 3, 2, 8, 7, 0, 8, 1, 3, 1, 8, 3, 6, 0, 7, 9, 6, 1, 3, 4, 8, 6, 3, 4, 9, 1, 4, 2, 8, 1, 0, 5, 5, 9, 2, 3, 1, 4, 0, 5, 8, 8, 9, 8, 3, 9, 9, 3, 3, 5, 9, 1, 1, 5, 3, 6, 8, 4, 7, 7, 9, 6, 0, 4, 0, 6, 0, 5, 7, 3, 1, 5, 9, 5, 4, 0, 1, 4, 6, 0, 0, 5, 4, 6, 2, 4, 8, 4, 2, 0, 5, 4, 4, 1, 0, 2, 0, 5, 4, 1, 3, 7, 5, 3, 3, 1, 6, 7, 1, 0, 2, 9, 6, 7, 0, 1, 7.

El número de huecos registrados será la cantidad de números analizados menos el número de números aleatorios generados (en este caso son 10, puesto que cada dígito se debe presentar, por lo menos, una última vez).

Total de huecos (T) = N – 10 donde N es el tamaño de la muestra

(T) = 125 – 10 = 115.

Después se verifica cual fue la mayor longitud del hueco, y dependiendo de ésta usted elegirá cuantos intervalos requiere. Por ejemplo: si tiene una longitud de hueco igual a 49 y desea 10 intervalos entonces el primer intervalo será de 0 – 4, el segundo de 5 – 9, el tercero de 10 – 14, etc. Si quisiera solo 5 intervalos entonces quedará el primero de 0 – 9, el segundo de 10 – 19, el tercero de 20 – 29, el cuarto de 30 – 39 y el quinto de 40 – 49. Para el ejemplo se tiene que la mayor longitud de hueco es de 50 y se dividió en 17 intervalos.

Enseguida se analizan cada uno de los números aleatorios generados para determinar su longitud de hueco y obtener la frecuencia en los intervalos generados. Por ejemplo: si tomamos el número aleatorio siete (7) su primera longitud de hueco es de 9; y caerá en el intervalo 9 – 11, entonces ese intervalo tendrá su primera frecuencia. Si el mismo número aleatorio u otro número cayeran en ese mismo intervalo entonces se sumaria la segunda frecuencia para este intervalo; y así sucesivamente para todos los intervalos. La suma de las frecuencias de todos los intervalos (en este ejemplo son 17) es igual a el total de huecos (T = 115) .

Pasos a seguir en la prueba.

Paso 1.

Especifique la fdp para la distribución de frecuencia teórica dada por la ecuación (14) basado en el ancho del intervalo de clase seleccionado.

Paso 2.

Arregle los huecos observados en una distribución acumulada con esas mismas clases.

Paso 3.

Encuentre *D*, La máxima desviación entre *F(x)* y $S_n(x)$ como en la ecuación

$$D = \max \left| F(x) - S_n(x) \right|$$

Paso 4.

Determine el valor crítico D_α , de la tabla de Kolmogorov–Smirnov para el valor específico de α y el tamaño de muestra *N*.

Paso 5.

Si el valor calculado de *D* es mayor que el valor tabulado de D_α la hipótesis nula de independencia es rechazada.

El valor exacto de α puede ser encontrado usando la metodología descrita por Conmover [1980].

Resumimos la prueba en la tabla siguiente:

Longitud de hueco	Ocurrencias	Frecuencia Acumulada	Frecuencia	F_x (x) | Diferencia |
0 - 2	27	0.234	0.271	0.027
3 - 5	30	0.495	0.469	0.026
6 - 8	23	0.695	0.613	0.082 *
9 - 11	11	0.792	0.718	0.074
12 - 14	8	0.86	0.794	0.066
15 - 17	3	0.888	0.85	0.038
18 - 20	2	0.905	0.891	0.014
21 - 23	3	0.931	0.92	0.011
24 - 26	1	0.94	0.942	0.002
27 - 29	1	0.948	0.958	0.01
30 - 32	2	0.965	0.969	0.004
33 - 35	1	0.974	0.978	0.004
36 - 38	1	0.983	0.984	0.001
39 - 41	0	0.983	0.988	0.005
42 - 44	1	0.99	0.991	0.001
45 - 47	0	0.99	0.994	0.004
48 - 50	1	1	0.995	0.005
	Total	115		

Para determinar la frecuencia acumulativa relativa se basa en la fórmula:

$S(x) = (m/T)$

1er. Intervalo $S(0-2) = (27/115) = 0.234.$

2do. Intervalo $S(3-5) = (57/115) = 0.495.$

3er. Intervalo $S(6-8) = (80/115) = 0.695.$

4to. Intervalo $S(9-11) = (91/115) = 0.792.$

y así sucesivamente hasta acabar con los intervalos.

Para determinar la frecuencia acumulativa relativa se basa en la fórmula:

$F_x(X) = 1-(0.9)^{x+1}$

1er. Intervalo $F_x(0-2) = 1-(0.9)^{2+1} = 0.271.$

2do. Intervalo $F_x(3-5) = 1-(0.9)^{5+1} = 0.469.$

3er. Intervalo $F_x(6-8) = 1-(0.9)^{8+1} = 0.613.$

4to. Intervalo $F_x(9-11) = 1-(0.9)^{11+1} = 0.718.$

Posteriormente se obtiene la diferencia máxima absoluta entre las dos frecuencias acumulativas $D^* = 0.082$. Esta diferencia se compara con la diferencia de confiabilidad. La diferencia de confiabilidad esta dada por la siguiente fórmula:

$$D_{Nivel\ de\ confiabilidad} = \frac{valor\ en\ la\ tabla\ Kolmogorov - Smirnov}{\sqrt{T}} \qquad (15)$$

Donde el nivel de confiabilidad es igual a 1 – nivel de significancia (1-.95)=0.05. Valor de la tabla con $\alpha_{\mathbf{0.05}}$ y N>35 **(**tamaño muestral 125) = 1.36 **(apéndice de tablas)**

$$D_{0.95} = \frac{1.36}{\sqrt{115}} = 0.127$$

Puesto que D^* (0.082) < $D_{0.95}$ (0.127); rechazamos la hipótesis de que los dígitos están ordenados aleatoriamente.

III.4.2 La prueba Chi-Cuadrada (χ^2)

Esta prueba puede ser realizada de dos maneras: considerando a números pseudoaleatorios generados como dígitos o como números reales.

III.4.2.1 Números pseudoaleatorios considerados como Dígitos

La prueba consiste en contar el número de dígitos que aparece entre ocurrencias sucesivas de un mismo dígito. Por ejemplo, 58425 ilustra un hueco de tamaño 3 entre los dos cincos, y así para cada uno de los números pseudoaleatorios, encontrando desde 0,1,2,3... hasta n tamaños de huecos.

La probabilidad de cada una de los tamaños de hueco(i = 0,1,2,3..) se obtiene con la siguiente expresión:

$$P_i = 0.1(0.9)^i \quad \text{para i =0,1,2,3.....}$$

sin embargo, como teóricamente el valor del tamaño del hueco puede ser infinito, es conveniente agrupar las probabilidades para valores de **i** mayores o iguales para un valor determinado de **n**. Tal sumatoria se obtiene de acuerdo a la siguiente expresión:

$$P_i \geq \sum_{m=0}^{\infty} 0.1(0.9)^{m+n} = (0.9)^n \qquad (16)$$

La frecuencia observada (FO) es el número de ocurrencias de los diferentes tamaños de huecos en la serie números pseudoaleatorios.

Para obtener la frecuencia esperada usamos la siguiente expresión:

$$(\sum \mathrm{fo_i})(0.1)(0.9)^i = (\sum \mathrm{fo_i})\mathrm{P_i} \qquad (17)$$

Una vez calculados fo y fe, calculamos el estadístico X^2_C, donde;

$$X^2c = \sum_{i=0}^{n} \frac{(fo_i - fe_i)^2}{fe_i} \quad (18)$$

El cual se compara con $X^2_{\alpha,\ n}$. Si $X^2_C < X^2_{\alpha,\ n,}$ entonces los números pseudoaleatorios pasan la prueba de la distancia. Es importante señalar que el valor seleccionado de n, debe ser tal que la suma de las frecuencias esperadas de todos los tamaños de huecos agrupados, sea mayor que 5.

Ejemplo:

Se tiene una serie de números pseudoaleatorios :

1 2 3 5 7 3 9 1 6 8

5 2 4 9 5 3 1 6 7 4

9 9 1 5 7 3 9 4 1 2

9 6 6 7 0 3 4 2 8 9

0 2 6 0 8 9 1 7 4 2

para n=3

i	P_i	f_o	f_e	$\frac{(fo_i - fe_i)^2}{fe_i}$
0	0.1	2	3.6	
1	0.1(0.9)	0	3.24	
2	$0.1(0.9)^2$	2	2.916	
≥3	$(0.9)^3$	32	26.244	

Cuando f_e <5 se agrupa la fila en sus datos con la fila inmediata superior, lo cual sucede con las filas 1, 2 y 3, por lo que se agrupan las filas 1,2 y 3.

i	P_i	f_o	f_e	$\frac{(fo_i - fe_i)^2}{fe_i}$
0	0.271	4	9.756	3.396
≥1	0.729	32	26.244	1.262
Total	1.0	36	36	4.658

Si $X^2c < X^2_{0.05,3,}$ entonces los números generados pasan la prueba de distancia. Como $X^2c=4.658 > X^2_{0.05,1}$

Los números no pasan la prueba de distancia.

Nota: $X^2_{0.05,3,}$ valor obtenido de la tabla de Ji con valor de significancia de 0.05 y n = 3 **(apéndice de tablas)**

III.4.2.2 Números pseudoaleatorios considerados como números reales.

Si los números pseudoaleatorios generados son considerados como reales, entonces, para realizar esta prueba es necesario seleccionar un intervalo (α ,β) el cual debe estar contenido en el intervalo de (0,1), es decir 0 ≤ α ≤ β ≤ 1. Enseguida para cada número pseudoaleatorio generado se pregunta si es o no elemento del intervalo (α,β). Si U_j (número uniforme generado) es elemento de (α ,β), U_{j+1} hasta U_{j+1} no son elementos de dicho intervalo y $U_{j+i+1,}$ vuelve a ser elemento del intervalo (α,β), entonces se tiene un hueco de tamaño i.

Ejemplo:
Sean α = .3 y β = .5 téngase los Números Pseudoaleatorios generados
0.32415, 0.22257, 0.19147, 0.755103, 0.49383

Entonces los números que caen el intervalo de .3 y .5 son 0.32415 y 0.49383 por lo tanto se tiene un hueco de 3.

La distribución de probabilidad del tamaño del hueco es:

$$P_i = θ(1- θ)^2, \quad \text{para } i= 0, 1, 2...$$

Donde θ = β -α representa la probabilidad de caer en el intervalo (α , β)
Se necesita agrupar las probabilidades para valores de i ≥ n. Y la formula para lograrlo es:

$$P_{i \geq n} = \sum(1- θ))^{m+n} = (1- θ)^n$$

Con estas formulas se obtienen las frecuencias esperadas presentadas en la tabla siguiente.

i	***Pi***	***Foi***	***Fei***	***X^2c***
0	θ	Fo_0	$\sum fo_0$ θ.	$(fo_i-fe_i)^2 / fe_i$
1	θ.(1- θ)	Fo_1	$\sum fo_1$ θ (1- θ)	$(fo_1-fe_1)^2 / fe_1$
2	θ.$(1- θ)^2$	Fo_2	$\sum fo_2$ θ $(1- θ)^2$	$(fo_2-fe_2)^2 / fe_2$
....	....	....	....	....
....	θ.$(1- θ)^i$	....	$\sum fo_i$ θ $(1- θ)^i$	....

i .... ≥n	.... $\theta_{.}(1-\theta)^n$	fo_i .... fo_n	.... $\sum fo_n\, \underset{\sim}{\theta}\, (1-\theta)^n$	$(fo_i-fe_i)^2 / fe_i$ .... $(fo_i-fe_i)^2 / fe_i$
TOTAL	1.0	. $\sum fo_i$	$\sum fo_i$	$\sum_{i=0}^{n} \frac{(fo_i - fe_i)^2}{fe_i}$

Utilizando la ecuación de:

$$X^2 c = \sum \frac{(fo_i - fe_i)^2}{fe_i}$$

Comparamos que el resultado obtenido con $X^2_{a,n}$. Se toma la decisión de aceptar o rechazar la prueba de distancia. Es muy importante señalar que los valores de α y β no tienen ninguna influencia en la bondad de la prueba y es necesario señalar que el valor de n debe ser seleccionado, de tal manera que la suma de las frecuencia esperada de todos los tamaños de huecos agrupados sea mayor que 5.

Ejemplo:
Tomamos los números presentados en la siguiente **tabla 2**

0.78961	0.05230	0.10699	0.55877	0.14151
0.76086	0.12079	0.27738	0.65726	0.79269
0.80548	0.82654	0.29453	0.20852	0.42989
0.58518	0.98611	0.34488	0.34358	0.11537
0.89898	0.57880	0.67621	0.05010	0.00121
0.28269	0.73059	0.70119	0.18284	0.49962
0.38618	0.76910	0.68334	0.55170	0.10850
0.79982	0.45679	0.21631	0.87616	0.55743
0.58962	0.33216	0.03185	0.61168	0.09264
0.69623	0.17028	0.05475	0.91512	0.76262
0.29931	0.30861	0.83358	0.51781	0.03272
0.57410	0.26593	0.85903	0.43308	0.35286
0.24000	0.65559	0.38507	0.90829	0.94187
0.93655	0.88809	0.81772	0.36982	0.19904
0.54325	0.62400	0.09133	0.41678	0.33954

0.58244	0.85853	0.88752	0.33729	0.15506
0.23949	0.53559	0.33381	0.49383	0.75103
0.19962	0.65002	0.74579	0.79113	0.63453
0.19147	0.40644	0.08128	0.73435	0.22724
0.22287	0.07281	0.64183	0.44267	0.72102

con un valor de n=3, los valores de α= .3 y β= .7 entonces las frecuencias observadas y esperadas para los diferentes tamaños de hueco serian como aparecen en la siguiente tabla.

i	P_i	Foi	Fe_i	X^2c
0	.400	19	18.8	0.00212
1	.240	10	11.28	0.14524
2	.144	11	7.68	2.6462
≥3	.216	7	10.152	0.97863
TOTAL	1.0	47	47	3.77221

Si se especifica el valor arbitrario de α= 0.05 entonces $X^2_{0.05,3}$ = 7.81 ;este valor sacado de la tabla de valores de X^2 .
Como X^2c=3.77221 es menor que $X^2_{0.05,3}$= 7.81**(apéndice de tablas)**, entonces los números pseudoaleatorios presentados en la pasan la prueba de distancia.

III.5 Prueba de Póquer

La prueba POKER se utiliza para analizar la frecuencia con la que se repiten los dígitos en números aleatorios individuales. Para determinar si los números aleatorios generados cumplen con las propiedades especificadas (uniformidad e independencia) se tendrán las hipótesis siguientes :

H_0 si $X^2_{\text{confiabilidad}} > \Sigma (O_i - E_i)^2 / E_i$; se aprueba que los dígitos están ordenados al azar.
H_1 si $X^2_{\text{confiabilidad}} < \Sigma (O_i - E_i)^2 / E_i$; se rechaza que los dígitos están ordenados al azar.

Se utiliza para analizar la frecuencia con la que se repiten los dígitos en números aleatorios individuales. Por ejemplo, si nos ocupamos de números aleatorios de cinco dígitos, nos interesara la frecuencia con que ocurre lo que sigue en los números individuales:

1.- Los cinco son diferentes.
2.- Hay exactamente un par.
3.- Dos pares diferentes.
4.- Tres dígitos iguales.
5.- Tres dígitos iguales y un par.
6.- Cuatro dígitos iguales.
7.- Cinco dígitos iguales.

Por supuesto, el número de esas combinaciones que se pueden dar depende del número de dígitos que constituyen cada uno de los números aleatorios.

Para aplicar la prueba del póquer:

a) Escogemos primeramente un nivel de significancia, α, y enumeramos el grado de repetición de los dígitos.
b) A continuación, calculamos la probabilidad de aparición de cada una de esas combinaciones.
c) Luego, se examina la frecuencia con que se presenta cada combinación en la secuencia de números estudiados.
d) Posteriormente, se puede comparar la frecuencia observada con que aparece cada combinación con la frecuencia esperada, mediante la prueba de la ji cuadrada. Para comprobar que los datos pertenecen a una distribución Uniforme, se debe de cumplir la condición de que $X^2_{Calculada} < x^2_{\alpha/1,g.l.}$. Donde $x^2_{\alpha/2,g.l}$ se obtiene de la tabla de la distribución Ji cuadrada, con un nivel de significancia α y y los grados de libertad g.l. = No. de parámetros de la distribución de probabilidad a probar menos l.(en nuestro caso estamos probando la uniformidad y la distribución uniforme no tiene parámetros)

Como ejemplo, supóngase que tenemos que aplicar la prueba de póquer a *N* números aleatorios de cinco dígitos. Calcularemos la probabilidad de aparición de cada una de esas combinaciones, bajo la suposición de que los dígitos se presentan de una manera completamente aleatoria.

Todos diferentes $$\frac{10x9x8x7x6}{10^5} = 0.3024$$

Un par $$\frac{10x9x8x7x1}{10^5}\binom{5}{2} = 0.5040$$

Dos pares $$\frac{10x9x8x1x1}{10^5}\binom{5}{2}\binom{3}{2} = 0.1080$$

Tercia $$\frac{10x9x1x1x1}{10^5}\binom{5}{3} = 0.072$$

Full $$\frac{10x9x1x1x1}{10^5}\binom{5}{2}\binom{2}{2} == .0090$$

Póker $$\frac{10x9x1x1x1}{10^5}\binom{5}{4} = 0.0045$$

Quintilla $$\frac{10x1x1x1x1}{10^5}\binom{5}{5} = 0.0001$$

Formulas que ya están establecidas estadísticamente:

Prob(5 dígitos diferentes) = Prob(2° digito ≠ 1°) por Prob(3er ≠ 1° y el 2°) por Prob(4° ≠ de lo 3 primeros) por Prob(5°≠ de los 4 primeros)

Las probabilidades para cada una de las manos de póquer se muestran a continuación:

Prob(exactamente un par) = (0.9)(0.8)(0.1) =0.5040

Prob(dos pares) = $(0.1)^2(0.9)(0.8)$ =0.1080

Prob(tres dígitos iguales) = $(0.9)(0.8)(0.1)^2$ =0.0720

Prob(tres dígitos iguales mas un par) =$(0.9)(0.1)^3$ =0.0090

Prob(cuatro dígitos iguales) = $(0.9)(0.1)^3$ =0.0045

Pprob(cinco dígitos iguales) =$(0.1)^4$ =0.0001

Para obtener el número de veces que se puede esperar cada una de esas combinaciones, se multiplica cada probabilidad por *N.* Por supuesto, el numero de esas combinaciones que se pueden producir depende del numero de dígitos que constituyen cada uno de los números aleatorios.

Ejemplo:

Tenemos que aplicar la prueba del póquer a n números aleatorios de cinco dígitos. Las combinaciones posibles que indican el grado de repetición de los dígitos en un numero aleatorio dado se dieron antes. Calcularemos la probabilidad de aparición de cada una de esas combinaciones, bajo la suposición de que los dígitos se presentan de una manera completamente aleatoria.

Números aleatorios:

.85881	.99700	.75289	.82813	.02818	.36065	.45649	.06451	.07582	.73994
.52480	.03333	.50410	.76568	.11767	.37587	.55763	.33089	.53339	.41700
.24577	.74797	.92023	.93143	.05520	.94996	.35838	.85376	.41727	.08969

Si analizamos el primer dígito 0.85881 contiene una tercia de 8´s , el segundo dígito contiene dos pares uno de 7´s y uno de 9´s, y así sucesivamente se analizan todos los números aleatorios y se cuantifican las diferentes opciones en el juego de póquer agrupándolas para obtener la frecuencia esperada f_e de cada uno de ellos.

Para obtener el número de veces que se puede esperar cada una de esas combinaciones, se multiplica cada probabilidad por n.

Resultados del análisis del póquer:

Como la

Tipo de combinación	Frecuencia Observada f_o	Frecuencia Esperada f_e	$\frac{(f_e - f_o)^2}{f_e}$
Todos Diferentes	5	9.072	1.8277319
Un Par	16	15.12	0.0512169
Dos pares	4	3.24	0.1782716
Tercia	4	2.16	1.5674074
Full	0	0.27	1
Poker	1	0.135	5.542407
Quintilla	0	0.003	1
	∑=30	$X^2_{Calculada} = \sum \frac{(f_e - f_o)^2}{f_e} =$	

frecuencia esperada es menor de 5, se deben agrupar las filas con las inmediatas superiores hasta que la suma se al menos 5. Así;

Como

Tipo de combinación	Frecuencia Observada f_o	Frecuencia Esperada f_e	$\frac{(f_e - f_o)^2}{f_e}$
Todos Diferentes	5	9.072	1.8277319
Un Par	16	15.12	0.0512169
Dos pares	9	5.808	0.1754280
	∑=30	$X^2_{Calculada} = \sum \frac{(f_e - f_o)^2}{f_e} =$	3.6332297

α=0.05 y numero de intervalos es igual a 3, la $X^2_{Tabla} = X^2_{0.05,2} = 5.99$ **(apéndice de tablas)**, y entonces como 3.63 < 5.99 se acepta la hipótesis de que los números están ordenados al azar.

Combinación	frecuencia observada	frecuencia esperada	$\frac{(fo_i - fe_i)^2}{fe_i}$
I	fo	fe	
Todos	460	504	3.841
1 par	480	432	5.333
2 pares	16	27	4.481
3 dígitos iguales	44	37	0.2432
	1000	100	22.8982

Los resultados que se obtuvieron en la tabla fue de la siguiente forma:

4 dígitos distintos = .504 * 100 = 504

1 dígito par = .432 * 100 = 432

2 dígitos pares = .027 * 100 = 27

3 dígitos iguales = .037 * 100 = 37

Por lo tanto para sacar el resultado de la ultima columna se hace mediante la formula que se encuentra en la misma posición de la columna.

Puesto que x^2 0.95(4) = 9.488 <22.8982, que por lo tanto no podemos rechazar la aseveración de que los dígitos al interior de los números aleatorios están ordenados al azar. Por lo tanto si se aprueba la hipótesis.

Nota: este ejemplo fue creado con números aleatorios para cinco dígitos lo cual se puede realizar con números de cuatro dígitos, cinco dígitos, seis dígitos, siete dígitos , etc. Lo cual las formulas van a variar dependiendo de su tamaño de la longitud.

III.6 Prueba de Corridas

Una prueba de Corridas es un método que nos ayuda a evaluar el carácter de aleatoriedad de una secuencia de números estadísticamente independientes y números uniformemente distribuidos. Es decir dado una serie de números determinar si son o no aleatorios.

Existen dos versiones de la prueba de corridas:

- **Prueba de corridas arriba y abajo** (ascendente y descendente).
- **Prueba de corridas arriba y abajo de la media** (promedio).

III.6.1Prueba de corridas Arriba y Abajo para números estadísticamente independientes

Si tenemos una secuencia de números de tal manera que a cada uno de los números siga otro mayor la secuencia dada será ascendente (arriba).

Si cada número va seguido por otro menor, la secuencia será descendente (abajo).

Pasos para evaluar una prueba de corridas:

1. Primeramente le asignaremos un signo a cada número de la secuencia ya sea + ó - , eso dependerá de los siguiente.
2. Si a un número le sigue otro mayor, se le asigna +. Esto es si $\mathbf{X_i < X_i +1}$ el signo asignado será (+). Siendo $\mathbf{X_i}$ un número de la muestra o secuencia de números.
3. Si el número siguiente es menor, se le da un signo -. Esto es si $\mathbf{X_i > X_i +1}$ el signo asignado será (-).
4. Se continuará con la comparación de los números y la asignación de su signo correspondiente hasta N-1. Es decir hasta el penúltimo numero de la secuencia, ya que al último número le sigue un evento nulo(no es posible compararlo con otro número).

Para comprender mejor el método ejemplificaremos con la siguiente secuencia de números:

59,12,19,05,59,58,83,18,36,00,61,47,24,41,42,98,23,67,84,43,29,71,88,74,60,10,46,23,15,11,78,31,11,91,99,57,28,18,32,21,12,95,38,76,07,96,33,63,10,05

De acuerdo al método (prueba de corridas arriba y abajo) se evaluará 59<12, como no lo es se le asignará un signo **-**. Seguiremos comparando 12<19, ya que si lo es se le asigna un signo **+**.

Se continúa con la evaluación quedando de la siguiente manera:

- + - + - + - + - + - - + + + - + + - - + + - - - + - - - + - - + + - - - + - - + - + - + - + - -

Una vez encontrado los signos de cada número de la secuencia dada se procede a calcular el total de corridas que resulta de la suma de suma de corrida ascendente con la descendente.

Una corrida se define como una sucesión de eventos similares, precedidos y seguidos por un evento diferente.

En este ejemplo tendíamos un total de corridas de 33.

Sea ***a*** = 33 el número total de corridas en una secuencia. La media $\boldsymbol{\mu_a}$ y la varianza

$\boldsymbol{\sigma_a}$ de **a** están dadas por:

$$\mu_a = \frac{2N-1}{3} = \frac{2(50)-1}{3} = 33 \qquad \sigma_a{}^2 = \frac{16N-29}{90} = \frac{16(50)-29 = 8.57}{90}$$

Para N>20, es posible aproximarse razonablemente a la distribución de **a** mediante una distribución normal con la media y la varianza que se dan en las anteriores ecuaciones.

Por lo común, esa aproximación sería apropiada para comprobar la aleatoriedad de los números generados por un generador de números aleatorios, puesto que se pueden producir varios centenares de números antes de aplicar una prueba.

Podemos rechazar una hipótesis de que una secuencia de números es aleatoria, porque hay un número excesivo o demasiado bajo de corridas. Por ende se requiere una prueba de colas para determinar si se ha presentado alguno de esos extremos. Como estadística de la prueba utilizaremos:

$$Z = \frac{a1 - \mu_a}{\sigma_a}$$

H_0 : Hipótesis Nula Criterio de Aceptación $|Z| \leq Z_{1-\alpha/2}$. La secuencia de números es independiente y por lo tanto la secuencia es aleatoria

H_1 : Hipótesis Alternativa Criterio de rechazo $|Z| > Z_{1-\alpha/2}$ .La secuencia de números No es Independiente y por lo tanto la secuencia No es aleatoria.

Sustituyendo la media **μ_a** y la varianza **σ_a** , tenemos que:

$$Z = \frac{a - [(2N-1)/3]}{\sqrt{(16N-29)/90}} \qquad Z = \frac{a - \mu_a}{\sqrt{\frac{16N-29}{90}}} = \frac{33-33}{\sqrt{8.57}} = 0.00$$

Si se define el nivel de significancia por medio de $\propto = 0.05$, entonces, $Z_{1-\alpha/2}$ será igual a 1- 0.05/2 = 0.975, buscando este valor en las tablas de Z encontramos que tiene un valor de 1.96 entonces, si el valor absoluto de Z calculada es mayor o igual a la Z de las tablas se rechazará la hipótesis de la independencia de los números (propiedad de los números pseudoaleatorios). Esto es:

Z calculada =0.00 < $Z_{0.975}$ =1.96

Estaremos rechazando la hipótesis de que los números dados no son estadísticamente independientes. Debido a la falsedad de la comparación llegamos a la aceptación de la hipótesis alternativa.

Nota.

- Una secuencia de números puede ser no aleatoria si se tienen demasiadas o muy pocas corridas.

- Si tenemos una secuencia de N números, el número máximo de corridas posibles es N-1. El número mínimo posible es siempre uno.

III.6.2 Prueba de corridas arriba y abajo de la media para números uniformemente distribuidos

El anterior método no es completamente adecuado para evaluar la aleatoriedad de una secuencia de números veamos porque:

Se tiene la siguiente secuencia de 50

59,12,19,05,59,58,83,18,36,00,61,47,24,41,42,98,23,67,84,43,29,71,88,74,60,10,46,23,15,11,78,31,11,91,99,57,28,18,32,21,12,95,38,76,07,96,33,63,10,05

Si tuviéramos que aplicar el anterior método, obtendríamos la siguiente secuencia de signos más y menos:

- + - + - + - + - + - - + + + - + + - - + + - - - + - - - + - - + + - - - + - - + - + - + - + - -

Esta secuencia es idéntica a la que aparece en el ejemplo pasado. Por lo tanto, el análisis precedente sugerirá que esos números son verdaderamente aleatorios. Sin embargo esa aseveración es claramente discutible, puesto que los primeros 25 números caen por encima de la media ($\mu = 49.5$), mientras que los 25 restantes caen bajo la media. El carácter no aleatorio de esta secuencia se sugiere por el hecho de que tenemos una corrida de números por encima de la media, seguida por una corrida por debajo de la media. Es por ello que necesitamos de otro método que nos lleva a la verdadera respuesta. Utilizando entonces el método llamado ***prueba de corridas por arriba y abajo de la media.*** El cual consiste en lo siguiente:

- Denotaremos con un signo - a aquel número que se encuentre por debajo de la media.
- Denotaremos con un signo + a aquel número que se encuentre por arriba de la media.

Ejemplo:

En la secuencia de números

9, 4, 5,6,1,0,6,6,4,9,2,8,4,0,3,7,5,5,5,7,1,8,9,1,0

cuya media es de 4.6, tenemos una asignación de signos más y menos como sigue:

+ - + + - - + + - + - + - - - + + + + + - + + - -

En este caso tenemos una corrida de 1 sobre la media, 1 bajo la media, dos sobre la media, etc. Ahora bien tenemos 7 corridas por encima de la media y 7 por debajo de ella.

Sean **n_1** y **n_2** el número de observaciones individuales por encima y por debajo de la media, respectivamente y sea **b** como el número total de corridas. Note que el número máximo de corridas es $N=n_1+n_2$, y el número mínimo de corridas es 1. Dados **n_1** y **n_2**, la media de **b** - con una corrección de continuidad sugerida por Swed and Eisenhart [1943]- y la varianza de **b** para una

$$\mu_b = \frac{2n_1n_2}{n_1+n_2}+1 \qquad (20)$$

secuencia verdaderamente Independiente están dadas por:

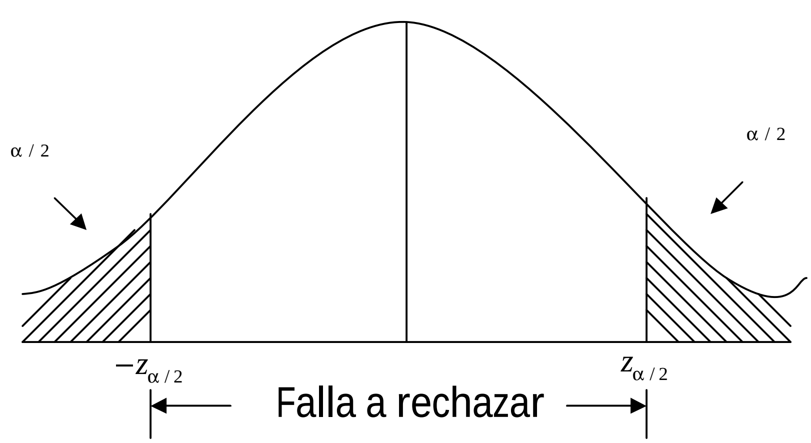
$$\sigma^2_b = \frac{2n_1n_2(2n_1n_2-N)}{N^2(N-1)} \qquad (21)$$

$$Z = \frac{b-\mu_b}{\sigma_b} \qquad (22)$$

para **n_1=**0 **n_2** mayor que 20, **b** se aproxima a una distribución normal y, para este caso, la estadística de la prueba puede ser realizada restando a la media el número de corridas y dividiéndola por la desviación estándar, o;

Puesto que nos interesa la presencia de un número demasiado grande o demasiado pequeño de corridas, conviene nuevamente aplicar una prueba de dos colas. Si se especifica el nivel de significancia por medio de ∝, rechazaremos la hipótesis de aleatoriedad, sí;

$$|Z| \geq Z_1 - \frac{\alpha}{2}$$



Ejemplo:

Determine si la secuencia siguiente de 40 números es tal que la hipótesis de independencia pueda ser rechazada donde α = 0.05.

41, 68, 89, 94, 74, 91, 55, 62, 36, 27, 19, 72, 75, 9, 54, 2, 1, 36, 16, 28, 18, 1, 95, 69, 18, 47, 23, 32, 82, 53, 31, 42, 73, 4, 83, 45, 13, 57, 63, 29

La secuencia de corridas arriba y debajo de la media es la siguiente;

- + + + + + + + - - - + + - + - - - - - - - + + - - - - + + - - + - + - - + + -

Existen 17 corridas en la secuencia, con N=40 y b=17, n1=18 y n2=22

Se determinan μ_b y σ_b usando las ecuaciones (20) y (21) respectivamente

$$\mu_b = \frac{2(18)(22)}{40} + 1 = 20.8 \qquad \sigma_b = \frac{2(18)(22)[2(18)(22) - 40]}{(40)^2(40-1)} = 9.54$$

Debido a que n2 es mayor que 20, la distribución Normal es aceptable, resultando en Z un valor de;

$$Z = \frac{17 - 20.8}{\sqrt{9.54}} = -1.23$$

Ya que $Z_{0.025}$ = 1.96, la hipótesis de independencia no puede ser rechazada sobre la base de esta prueba. ($Z_{Calculada}$= -1.23 < $Z_{0.025}$ = 1.96).

Existe otra prueba estadística para determinar si los números dados se encuentran uniformemente distribuidos (propiedad de los números pseudoaleatorios). Este estadístico es el conocido con el nombre de ji cuadrada (X^2), utilizado ya, para las anteriores pruebas (pruebas de frecuencia, distancia, huecos, etc.). Para ello ya no necesitaremos el número de corrida sino, la longitud de corrida. Por ejemplo supóngase que dentro de una secuencia de 1000 números, se observaran 250 corridas por encima de la media y 250 por debajo de la media. Además, supóngase que cada corrida sea de longitud dos. Para esta situación, el número esperado de corridas por encima y por debajo de la media es 501 (utilizando la formula de la media vista anteriormente). Así pues, al observar 500 corridas en la secuencia, nos veríamos obligados a aceptar la hipótesis de la aleatoriedad. Sin embargo, podríamos esperar encontrar corridas de longitud diferente de dos, dentro de esa serie larga de números. Sea $\mathbf{R_i}$ el número de corridas de longitud **i** en una secuencia de **N** números. Para el valor esperado de $\mathbf{R_i}$ tenemos:

$$E(R_i) = 2N\frac{(n_1)}{N}\frac{(n_2)^2}{N} \qquad (23)$$

$$E(R_i) = \frac{2}{(i+3)!}\left[N(i^2+3i+1)-(i^3+3i^2-i-4)\right], i \le N-1 \qquad (24)$$

para corridas por encima y por debajo de la media y una gran **N**.

Utilizando la prueba $\boldsymbol{\chi^2}$ cuadrada mencionada anteriormente podemos comparar el número observado de corridas de una longitud dada con el número esperado. Es decir si $\mathbf{O_i}$, es el número

$$E(R_i) = \frac{2}{N!}, \qquad i = N-1 \qquad (25)$$

observado de corridas de longitud **i**, tenemos como estadística de prueba:

$$X_2 = \sum_{i=1}^{L} \frac{\left[O_i - E(R_i)\right]^2}{E(R_i)} \qquad (26)$$

En donde L =N para corridas por encima y por debajo de la media y L = N -1 para corridas ascendentes y descendentes.

Ejemplo:

Dada la secuencia de números que sigue, ¿Se puede rechazar la hipótesis de que los números son aleatorios, sobre la base de la distribución de las longitudes de corridas por encima y por debajo de la media? Sea $\alpha = 0.05$.

0.44, 0.04, 0.12, 0.22, 0.64, 0.55, 0.43, 0.92, 0.65, 0.24, 0.69, 0.86, 0.48, 0.78, 0.47, 0.20, 0.80, 0.04, 0.67, 0.28, 0.17, 0.99, 0.02, 0.55, 0.59, 0.66, 0.01, 0.29, 0.47, 0.06, 0.31, 0.72, 0.17, 0.48, 0.74, 0.05, 0.92, 0.15, 0.80, 0.22, 0.86, 0.96, 0.35, 0.29, 0.36, 0.32, 0.51, 0.74, 0.33, 0.78, 0.99, 0.77, 0.57, 0.35, 0.81, 0.53, 0.78, 0.61, 0.52, 0.95, 0.26, 0.21, 0.99, 0.01, 0.95, 0.30, 0.88, 0.80, 0.37, 0.00, 0.01, 0.28, 0.21, 0.34, 0.86, 0.67, 0.67, 0.32, 0.78, 0.76

Para esta secuencia de números tenemos:

- - - - + + - + + - + + - + - - + - + - - + - + + + - - - - - + - - + - + - + - + + - + + - + + + + - + + + + + + - -
+ - + - + + - - - - - - + + + - + +

Puesto que estamos interesados en la longitud de los corridas por encima y por debajo de la media, utilizando la ecuación de el valor esperado ($E(R_i)$), para determinar el numero esperado de corridas de longitud *i*.

Por ejemplo $E(R_i)$ está dada por:

$$E(R_i) = 2(80)\frac{(40)}{(80)}\frac{(40)^2}{(80)} = 20$$

En donde **n1** = 40, **n2** =40. Los cálculos restantes se resumen en la tabla siguiente:

Longitud de corrida	**Corridas Observadas (O_i)**	**Corridas esperadas $E(R_i)$**	$\frac{[Oi - E(Ri)]^2}{E(Ri)}$
1	23	20	0.45
2	11	10	0.10
≥3	6	10	0.4
Total	40	40	0.95

El valor $\mathbf{X^2}$ experimental está dado por:

= 0.95 y $\mathbf{X^2_{0.95}}$ **(3)** = 7.81 **(apéndice de tablas)**

Puesto que $X^2 < X^2_{0.95}$ (3), no podemos rechazar la hipótesis de que los números dados no son aleatorios, sobre la base de esta prueba.

Definiciones

Corrida.- Es la sucesión de eventos similares, precedidos y seguidos por un evento diferente.

Longitud de Corrida.- Es el número de eventos que ocurren en la corrida. Es decir la longitud de la corrida se determina con el número de signos iguales que contienen la secuencia de números.

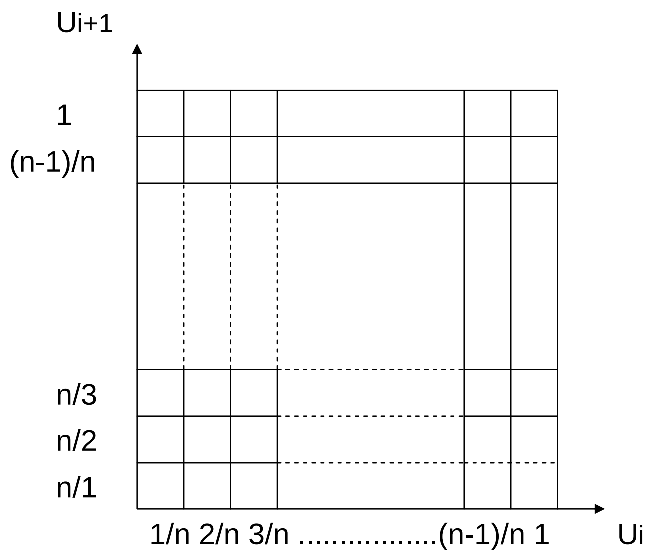
III.7 Prueba de Series

La prueba de series se utiliza para comprobar el grado de aleatoriedad entre números sucesivos.

Usualmente esta prueba consiste en formar parejas de números, las cuales son consideradas como coordenadas en un cuadro unitario dividido en n^2 celdas. El valor de n se toma a criterio de cada uno, de acuerdo al tamaño de la muestra de N números. La prueba consiste en generar N números pseudoaleatorios de los cuales se forman parejas aleatorias entre U_i y U_{i+1}, es decir, si se generan 10 números, entonces las parejas aleatorias que se pueden formar serian:

(U_1,U_2), (U_2,U_3), (U_3,U_4), (U_4,U_5), (U_5,U_6), (U_6,U_7), (U_7,U_8), (U_8,U_9), (U_9,U_{10})

Enseguida, se determina la celda a que pertenece cada pareja ordenada, con lo cual se determina la frecuencia observada de cada celda.



Para determinar la frecuencia observada por ejemplo de la pareja de números (0.17028, 0.57410), (0.57410, 0.18284), (0.18284, 0.14151), (0.14151, 0.65559), (0.65559,0.88809)

para n=2 se obtendría la siguiente tabla de 2x2 y se le asignarían los intervalos para posteriormente determinar donde cae cada pareja de números. Ejemplo:

1	2	1
.5	1	1
	.5	1

La frecuencia esperada de cada celda se obtiene al dividir el total de parejas coordenadas (n-1) por el total de celdas (n^2).

$$(n-1)/n^2$$

Finalmente, conocida la frecuencia observada y esperada de cada celda, se obtiene el estadístico Xo^2 como sigue:

$$X_0 = \frac{n^2}{N-1}\sum_{i=1}^{n}\sum_{j=1}^{n}(fo_{ij} - \frac{N-1}{n^2})^2 \qquad (27)$$

Si $\mathbf{X_o^2 < X^2_{calculada},\ n^2_{-1}}$, entonces no se puede rechazar la hipótesis de que los números provienen de una distribución uniforme.

Por ejemplo, si se aplica esta prueba a los números pseudoaleatorios dado a continuación, y utilizamos un valor de n=5, determinar las frecuencias observadas de los números de la **tabla 2** dada anteriormente en el punto **3.2.2**.

Primero se agrupan todos los números

(0.78961,0.76086), (0.76086,0.80548), (0.80548,0.58518), (0.58518,0.89898), ... hasta N-1

Posteriormente se asigna en la tabla las frecuencias observadas

1	3	3	3	4	2
.8	4	5	5	6	2
.6	3	3	2	6	6
.4	5	5	4	4	4
.2	6	6	6	2	2
	.2	.4	.6	.8	1

Después calcular **Xo²**

$$X_0{}^2 = \frac{25}{99}\left[5(2-3.99)^2 + 5(3-3.96)^2 + 5(4-3.96)^2 + 6(5-3.96)^2 + 4(6-3.96)^2\right] \qquad X_0{}^2 = 11.86$$

y por ultimo comparar que **Xo^2 < $X^2_{calculada}$, n^2_{-1}** para determinar que los números pseudoaleatorios presentados pasan la prueba de uniformidad. α=0.05

Buscando en la tabla de **$X^2_{calculada,\ 0.05,24}$** (**apéndice de tablas),** obtenemos el valor de 36.4 y comparamos

11.86 < 34.4

Como **Xo^2 < $X^2_{calculada,\ 0.05,24.}$** Concluimos entonces los números pseudoaleatorios presentados pasan la prueba de uniformidad.

Apéndice de Tablas

Tabla de Kolmogorov – Smirnov

Tamaño muestral (N)	nivel de significancia (α) 0.20	0.15	0.10	0.05	0.01	Tamaño muestral (N)	nivel de significancia (α) 0.20	0.15	0.10	0.05	0.01
1	0.900	0.925	0.950	0.975	0.995	14	0.274	0.292	0.314	0.349	0.418
2	0.684	0.726	0.776	0.842	0.929	15	0.266	0.283	0.304	0.338	0.404
3	0.565	0.597	0.642	0.708	0.828						
4	0.494	0.525	0.564	0.624	0.733	16	0.258	0.274	0.295	0.328	0.392
5	0.446	0.474	0.510	0.556	0.669	17	0.250	0.266	0.286	0.318	0.381
						18	0.244	0.259	0.278	0.309	0.371
6	0.410	0.436	0.470	0.521	0.617	19	0.237	0.252	0.272	0.301	0.363
7	0.381	0.405	0.438	0.486	0.577	20	0.231	0.246	0.265	0.294	0.356
8	0.358	0.381	0.411	0.457	0.543						
9	0.339	0.360	0.388	0.432	0.514	25	0.21	0.22	0.24	0.27	0.32
10	0.322	0.342	0.368	0.410	0.490	30	0.19	0.20	0.22	0.24	0.29
						mas de					
11	0.307	0.326	0.352	0.391	0.468	35	1.07	1.14	1.22	1.36	1.63
12	0.295	0.313	0.338	0.375	0.450						
13	0.284	0.302	0.325	0.361	0.433						

Swed and Eisenhart [1943].

Tabla Chi – Cuadrada (χ2)

Tabla de probabilidades χ2 Chi - Cuadrada

Grados de	**Nivel de Singnificancia α**									
Libertad	**0.995**	**0.99**	**0.975**	**0.95**	**0.9**	**0.1**	**0.05**	**0.025**	**0.01**	**0.005**
1	---	---	0.001	0.004	0.016	2.706	3.841	5.024	6.635	7.879
2	0.01	0.02	0.051	0.103	0.211	4.605	5.991	7.378	9.21	10.597
3	0.072	0.115	0.216	0.352	0.584	6.251	7.815	9.348	11.345	12.838
4	0.207	0.297	0.484	0.711	1.064	7.779	9.488	11.143	13.277	14.86
5	0.412	0.554	0.831	1.145	1.61	9.236	11.07	12.833	15.086	16.75
6	0.676	0.872	1.237	1.635	2.204	10.645	12.59	14.449	16.812	18.548
7	0.989	1.239	1.69	2.167	2.833	12.017	14.07	16.013	18.475	20.278
8	1.344	1.646	2.18	2.733	3.49	13.362	15.51	17.535	20.09	21.955
9	1.735	2.088	2.7	3.325	4.168	14.684	16.92	19.023	21.666	23.589
10	2.156	2.558	3.247	3.94	4.865	15.987	18.31	20.483	23.209	25.188
11	2.603	3.053	3.816	4.575	5.578	17.275	19.68	21.92	24.725	26.757
12	3.074	3.571	4.404	5.226	6.304	18.549	21.03	23.337	26.217	28.3
13	3.565	4.107	5.009	5.892	7.042	19.812	22.36	24.736	27.688	29.819
14	4.075	4.66	5.629	6.571	7.79	21.064	23.69	26.119	29.141	31.319
15	4.601	5.229	6.262	7.261	8.547	22.307	25	27.488	30.578	32.801
16	5.142	5.812	6.908	7.962	9.312	23.542	26.3	28.845	32	34.267
17	5.697	6.408	7.564	8.672	10.085	24.769	27.59	30.191	33.409	35.718
18	6.265	7.015	8.231	9.39	10.865	25.989	28.87	31.526	34.805	37.156
19	6.844	7.633	8.907	10.117	11.651	27.204	30.14	32.852	36.191	38.582
20	7.434	8.26	9.591	10.851	12.443	28.412	31.41	34.17	37.566	39.997
21	8.034	8.897	10.283	11.591	13.24	29.615	32.67	35.479	38.932	41.401
22	8.643	9.542	10.982	12.338	14.041	30.813	33.92	36.781	40.289	42.796
23	9.26	10.196	11.689	13.091	14.848	32.007	35.17	38.076	41.638	44.181
24	9.886	10.856	12.401	13.848	15.659	33.196	36.42	39.364	42.98	45.559
25	10.52	11.524	13.12	14.611	16.473	34.382	37.65	40.646	44.314	46.928

26	11.16	12.198	13.844	15.379	17.292	35.563	38.89	41.923	45.642	48.29
27	11.81	12.879	14.573	16.151	18.114	36.741	40.11	43.195	46.963	49.645
28	12.46	13.565	15.308	16.928	18.939	37.916	41.34	44.461	48.278	50.993
29	13.12	14.256	16.047	17.708	19.768	39.087	42.56	45.722	49.588	52.336
30	13.79	14.953	16.791	18.493	20.599	40.256	43.77	46.979	50.892	53.672
40	20.71	22.164	24.433	26.509	29.051	51.805	55.76	59.342	63.691	66.766
50	27.99	29.707	32.357	34.764	37.689	63.167	67.51	71.42	76.154	79.49
60	35.53	37.485	40.482	43.188	46.459	74.397	79.08	83.298	88.379	91.952
70	43.28	45.442	48.758	51.739	55.329	85.527	90.53	95.023	100.43	104.22
80	51.17	53.54	57.153	60.391	64.278	96.578	101.9	106.63	112.33	116.32
90	59.2	61.754	65.647	69.126	73.291	107.57	113.1	118.14	124.12	128.3
100	67.33	70.065	74.222	77.929	82.358	118.5	124.3	129.56	135.81	140.17

Referencias Bibliográficas

Azarang, M. R. y García Dunna, E., (1996), *Simulación y Análisis de Modelos Estocásticos* McGrawHill/Interamericana de México, S.A. de C.V., México.

Banks, J. y Carson, J.S., (1984), *Discrete event system simulation*, Prentice-Hall, Englewood Cliffs, N.J..

Bratley, P., Fox, B.L., Schrage, L.E. (1983) A Guide to Simulation. Springer Verlag

Concebís B., *Discrete Systems Simulation*, Mc. Graw Hill

Coss Bu Raúl, (2002), *Simulación Un enfoque práctico* , Limusa

Hillier, F.S. y Lieberman, G.J., (2003), *Introducción a la Investigación de Operaciones*, 5ª. Edición, , McGrawHill/Interamericana de México, S.A. de C.V., México.

Naylor, Balintfy y Burdick, *Técnicas de Simulación de computadoras*, Limusa

Ross, S., (1997), *Simulation*, 2a Edición, Academic Press, USA

Shdmit y Taylor*, Análisis y Simulación de Sistemas Industriales*, Trillas

Taha, H.A., (1991), *Investigación de Operaciones*, 2ª Edición, Alfaomega S.A., México.

Winston, *Investigación de Operaciones*, Gpo. Editorial Iberoamérica

http://random.mat.sbg.ac.at/tests/#Empirical

http://www.barringer1.com/simul.htm

http://www.barringer1.com/MC.htm