PRUEBA DE ALEATORIEDAD

Prueba de las corridas

La aleatoriedad es un campo de definición que, en matemáticas, se asocia a todo proceso cuyo resultado no es previsible más que en razón de la intervención del azar.
El resultado de todo suceso aleatorio no puede determinarse en ningún caso antes de que este se produzca.
El estudio de los fenómenos aleatorios queda dentro del ámbito de la teoría de la probabilidad y, en un marco más amplio, en el de la estadística.
Las propiedades estadísticas que deben poseer los números pseudoaleatorios generados por los métodos congruencia les tienen que ver con independencia y aleatoriedad estadísticas.

Si un conjunto de datos no pasa la prueba de aleatoriedad, entonces puede ser sustituida por otra serie de dato aleatorizados que pase el test de aleatoriedad

Pruebas de aleatoriedad

Las dos propiedades más importantes que deben satisfacer los números de un conjunto ri, son uniformidad e independencia-aleatoridad. En esta ocasión hablaremos de las pruebas estadísticas que tratan de corroborar si los números en el intervalo (0, 1) son independientes o, en otras palabras, si parecen pseudoaleatorios.

Para probar la aleatoriedad de los números de un conjunto ri, primero es preciso formular las siguientes hipótesis:

Ho: los números del conjunto ri, son aleatorios
H1: los números del conjunto ri, no son aleatorios

Pruebas de Independencia de Corridas Arriba y Abajo

La 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).

Para el presente trabajo haremos un hincapié en la Prueba de Independencia de Corridas Arriba y Abajo de la Media

Prueba de corridas arriba y abajo

El procedimiento de esta prueba consiste en determinar una secuencia de números (S) que sólo contiene unos y ceros, de acuerdo con una comparación entre r_i y r_{i-1}. Después se determina el número de corridas observadas, C_O (una corrida se identifica como la cantidad de unos o ceros consecutivos). Luego se calcula el valor esperado, la varianza del número de corridas y el estadístico Z_0, mediante las ecuaciones:

$$\mu_{C_O} = \frac{2n-1}{3}$$

$$\sigma^2_{C_O} = \frac{16n-29}{90}$$

$$Z_0 = \left| \frac{C_O - \mu_{C_O}}{\sigma_{C_O}} \right|$$

Si el estadístico Z_0 es mayor que el valor crítico de $Z_{\alpha/2}$, se concluye que los números del conjunto r_i no son independientes. De lo contrario no se puede rechazar que el conjunto de r_i sea independiente.

Considere el siguiente conjunto r_i de 21 números:

$$r_i = \{0.89, 0.26, 0.01, 0.98, 0.13, 0.12, 0.69, 0.11, 0.05, 0.65,$$
$$0.21, 0.04, 0.03, 0.11, 0.07, 0.97, 0.27, 0.12, 0.95, 0.02, 0.06\}$$

La secuencia de unos y ceros se construye de esta manera: se coloca un cero si el número r_i es menor que o igual al número r_i anterior; en caso de ser mayor que el número r_i anterior, se pone un uno. Si se considera la secuencia de los 21 números del conjunto r_i que se dio antes, la secuencia de unos y ceros es:

$$S = \{\mathbf{0},\mathbf{0},1,\mathbf{0},\mathbf{0},1,\mathbf{0},\mathbf{0},1,\mathbf{0},\mathbf{0},\mathbf{0},1,0,1,\mathbf{0},\mathbf{0},1,\mathbf{0},1\}$$

Observe que la secuencia S contiene $n - 1$ números, en este caso 20. Esto se debe a que el primer número $r_i = 0.89$ no tiene número anterior con el cual compararlo. Recuerde que una corrida se forma con unos consecutivos o ceros consecutivos. Por ejemplo los primeros dos ceros de la secuencia forman la primer corrida, que se dice que tiene una longitud de dos; el tercer número de la secuencia, uno, forma la segunda corrida con longitud de uno; después siguen dos ceros, los cuales forman la tercera corrida con longitud de dos; después sigue un uno, que forma la cuarta corrida con longitud de uno, etcétera. Mediante el proceso anterior se determina r_i que el número de corridas de la secuencia es $C_o = 14$.

Realizar la prueba de corridas arriba y abajo con un nivel de aceptación de 95% al siguiente conjunto de 40 números r_i:

0.34	0.83	0.96	0.47	0.79	0.99	0.37	0.72	0.06	0.18
0.67	0.62	0.05	0.49	0.59	0.42	0.05	0.02	0.74	0.67
0.46	0.22	0.99	0.78	0.39	0.18	0.75	0.73	0.79	0.29
0.11	0.19	0.58	0.34	0.42	0.37	0.31	0.73	0.74	0.21

Realizaremos la asignación de unos y ceros por renglón (o fila). Por lo tanto, la secuencia S es:

$$S = \{1,1,\mathbf{0},1,1,\mathbf{0},1,\mathbf{0},1,1,\mathbf{0},\mathbf{0},1,1,\mathbf{0},\mathbf{0},\mathbf{0},1,\mathbf{0},\mathbf{0},\mathbf{0},1,\mathbf{0},\mathbf{0},\mathbf{0},1,\mathbf{0},1,\mathbf{0},\mathbf{0},1,1,\mathbf{0},1,\mathbf{0},\mathbf{0},1,1,\mathbf{0}\}$$

obteniéndose un valor de $C_o = 24$ y $\alpha = 5\%$.

A continuación se presentan los cálculos correspondientes al valor esperado y a la varianza del número de corridas:

$$\mu_{C_o} = \frac{2n-1}{3} = \frac{2(40)-1}{3} = 26.333$$

$$\sigma^2_{C_o} = \frac{16n-29}{90} = \frac{16(40)-29}{90} = 6.788$$

$$Z_0 = \left|\frac{C_o - \mu_{C_o}}{\sigma_{C_o}}\right| = \left|\frac{24-26.333}{\sqrt{6.788}}\right| = 0.8954$$

Como el estadístico Z_0 es menor que el valor de tabla de la normal estándar para $Z_{\alpha/2} = Z_{5\%/2} = 1.96$, se concluye que no se puede rechazar que los números del conjunto r_i son independientes. Es decir, de acuerdo con esta prueba, los números son aptos para usarse en simulación.

Prueba de corridas arriba y abajo de la media

El procedimiento de esta prueba consiste en determinar una secuencia de unos y ceros, de acuerdo con una comparación entre los números del conjunto r_i y 0.5. Después se determina el número de corridas observadas, C_O, y los valores de n_0 y n_1. C_O es el número de corridas en la secuencia, el cual está determinado de la misma manera que en la prueba de corridas arriba y abajo; n_0 es igual a la cantidad de ceros en la secuencia, y n_1 es igual a la cantidad de unos en la secuencia, y se cumple que $n_0 + n_1 = n$. (Recuerde que una corrida se identifica como la cantidad de unos o ceros consecutivos). Luego se calcula el valor esperado, la varianza del número de corridas, y el estadístico Z_0 con las siguientes ecuaciones:

$$\mu_{C_O} = \frac{2n_0n_1}{n} + \frac{1}{2}$$

$$\sigma^2_{C_O} = \frac{2n_0n_1(2n_0n_1 - n)}{n^2(n-1)}$$

$$Z_0 = \frac{C_O - \mu_{C_O}}{\sigma_{C_O}}$$

Si el estadístico Z_0 está fuera del intervalo: $-Z_{\frac{\alpha}{2}} \quad Z_O \quad Z_{\frac{\alpha}{2}}$, se concluye que los números del conjunto r_i no son independientes. De lo contrario no se puede rechazar que el conjunto de r_i es independiente.

Considere la siguiente secuencia de 10 números de un conjunto r_i:

$$r_i = \{0.67, 0.62, 0.05, 0.49, 0.59, 0.42, 0.05, 0.02, 0.74, 0.67\}$$

La secuencia de unos y ceros se construye de la siguiente manera: se asigna un uno si el número r_i es mayor que o igual a 0.5. En caso contrario se asignará un cero. Al seguir esta regla, la secuencia de unos y ceros es:

$$S = \{1, 1, \mathbf{0}, \mathbf{0}, 1, \mathbf{0}, \mathbf{0}, \mathbf{0}, 1, 1\}$$

El número de corridas se determina de la misma manera que en la prueba de corridas arriba y abajo. En este caso se tiene que el número de corridas de la secuencia S es $C_O = 5$. Por otra parte, la secuencia tiene 5 ceros y 5 unos, así que $n_0 = 5$ y $n_1 = 5$.

Para saber si el estadístico está fuera del intervalo se emplea la siguiente fórmula:

$$-Z_{\frac{\alpha}{2}} \leq Z_0 \leq Z_{\frac{\alpha}{2}}$$

Para hallar los valores de los intervalos hacemos uso de la siguiente tabla:

$1-\alpha$	90%	92%	94%	95%	96%	97%	98%	99%
α	10%	8%	6%	5%	4%	3%	2%	1%
$Z_{\alpha/2}$	1,645	1,751	1,881	1,960	2,054	2,170	2,326	2,576
Z_{α}	1,282	1,405	1,555	1,645	1,751	1,881	2,054	2,326

Siendo:

$1 - \alpha$ = Nivel de confianza

α = Nivel de significación

Si el estadístico Zo está fuera del intervalo, se concluye que los números del conjunto *ri* no son independientes. De lo contrario no se puede rechazar que el conjunto de *ri* es independiente.

Realizar la prueba de corridas arriba y abajo, con un nivel de aceptación de 95%, al siguiente conjunto de 50 números r_i:

0.809	0.042	0.432	0.538	0.225	0.88	0.688	0.772	0.036	0.854
0.397	0.268	0.821	0.897	0.07	0.721	0.087	0.35	0.779	0.482
0.136	0.855	0.453	0.197	0.444	0.799	0.809	0.691	0.545	0.857
0.692	0.055	0.348	0.373	0.436	0.29	0.015	0.834	0.599	0.724
0.564	0.709	0.946	0.754	0.677	0.128	0.012	0.498	0.6	0.913

Construiremos la secuencia de unos y ceros por renglón quedando de la siguiente manera:

$$S=\{1,\mathbf{0,0},1,\mathbf{0},1,1,1,\mathbf{0},1,\mathbf{0,0},1,1,\mathbf{0},1,\mathbf{0,0},1,\mathbf{0,0},1,\mathbf{0,0,0},\\1,1,1,1,1,1,\mathbf{0,0,0,0,0,0},1,1,1,1,1,1,1,1,\mathbf{0,0,0},1,1\}$$

A partir de la secuencia anterior se determina que hay 21 corridas, 23 ceros y 27 unos. Por lo tanto, $C_O=21$, $n_0=23$ y $n_1=27$. A continuación se presentan los cálculos del valor esperado y de la varianza del número de corridas:

$$\mu_{C_O}=\frac{2n_0n_1}{n}+\frac{1}{2}=\frac{2(23)(27)}{50}+\frac{1}{2}=25.34$$

$$\sigma^2_{C_O}=\frac{2n_0n_1(2n_0n_1-n)}{n^2(n-1)}=\frac{2(23)(27)[2(23)(27)-50]}{(50)^2(50-1)}=12.08542$$

$$Z_0=\frac{C_O-\mu_{C_O}}{\sigma_{C_O}}=\frac{21-25.34}{\sqrt{12.08542}}=-1.2484$$

Como el valor de Z_0 cae dentro del intervalo $-1.96\le Z_0=-1.2484\le 1.96$, se dice que no se puede rechazar que los números del conjunto r_i son independientes con un nivel de confianza de 95 %. De acuerdo con esta prueba, el conjunto de números r_i se puede usar en un estudio de simulación.