Pruebas estadísticas para los números pseudoaleatorios

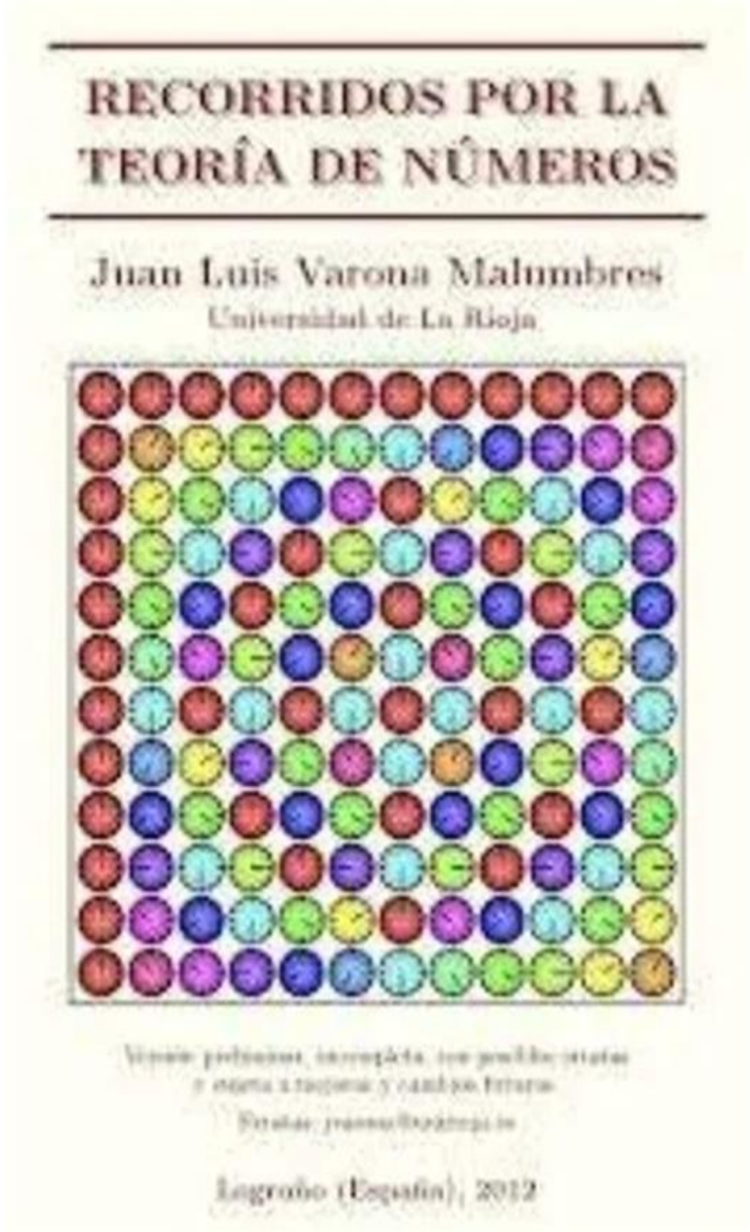
En la sección 2.2 se presentaron diversos algoritmos para construir un conjunto r_i pero ése es sólo el primer paso, ya que el conjunto resultante debe ser sometido a una serie de pruebas para validar si los números que lo integran son aptos para usarse en un estudio de simulación.

A continuación se analizarán las pruebas estadísticas básicas que se emplean generalmente para determinar si un conjunto de números pseudoaleatorios entre cero y uno cumplen con las propiedades básicas de independencia y uniformidad. El objetivo, en otras palabras, es validar que el conjunto r_i realmente está conformado por números aleatorios. Es importante mencionar que las pruebas que se discutirán no son únicas; si desea conocer otras, consulte Banks, Carson, Nelson y Nicol.[1]

CONDICIONES

DEBEN SER:

1. Uniformemente distribuidos
2. Estadísticamente independientes
3. Reproducibles
4. Sin repetición dentro de una longitud determinada de la sucesión
5. Generación a grandes velocidades
6. Requerir el mínimo de capacidad de almacenamiento Instituto



Prueba de Hipótesis para la media

Definición y aspectos generales

- Es un procedimiento basado en la evidencia muestral y en la teoría de probabilidad que se emplea para determinar si la hipótesis es un enunciado racional y no debe rechazarse o si es irracional y debe ser rechazado.
- En un trabajo de investigación se plantean dos hipótesis mutuamente excluyentes; la hipótesis nula y la hipótesis alternativa o de investigación.
- El análisis estadístico de los datos servirá para determinar si se puede aceptar o no la hipótesis nula (H_o).
- Cuando se rechaza H_o significa que el factor estudiado a influido significativamente en los resultados y se acepta la hipótesis alternativa (H_1)

Definición y aspectos generales

- Es importante tener presente que la hipótesis de investigación debe coincidir con la H_1.
- Plantear una hipótesis de investigación que coincida con H_o supondría una aplicación incorrecta del razonamiento estadístico.
- La hipótesis es el elemento que condiciona el diseño de investigación y responde provisionalmente al problema, verdadero motor de la investigación.
- El propósito de la prueba de hipótesis es determinar si el valor supuesto (hipotético) debe aceptarse como verosímil en base a evidencia muestral.

Pasos de la prueba de hipótesis

PASO 1

Planteamiento de la hipótesis nula y alternativa

3 situaciones

1 $\begin{cases} H_o: & \bar{X} = \mu \\ H_1: & \bar{X} \neq \mu \end{cases}$

2 $\begin{cases} H_o: & \bar{X} \leq \mu \\ H_1: & \bar{X} > \mu \end{cases}$

3 $\begin{cases} H_o: & \bar{X} \geq \mu \\ H_1: & \bar{X} < \mu \end{cases}$

PASO 2

Elegir el nivel de significancia (α):

Se define así a la "máxima cantidad de error que estamos dispuestos a aceptar para dar como válida la hipótesis del investigador".

En los sistemas biológicos siempre trabajaremos con un $\alpha = 0.05$ o en su forma 5%.

PASO 2 Elegir el nivel de significancia ($\alpha = 0.05$ o en su forma 5%.

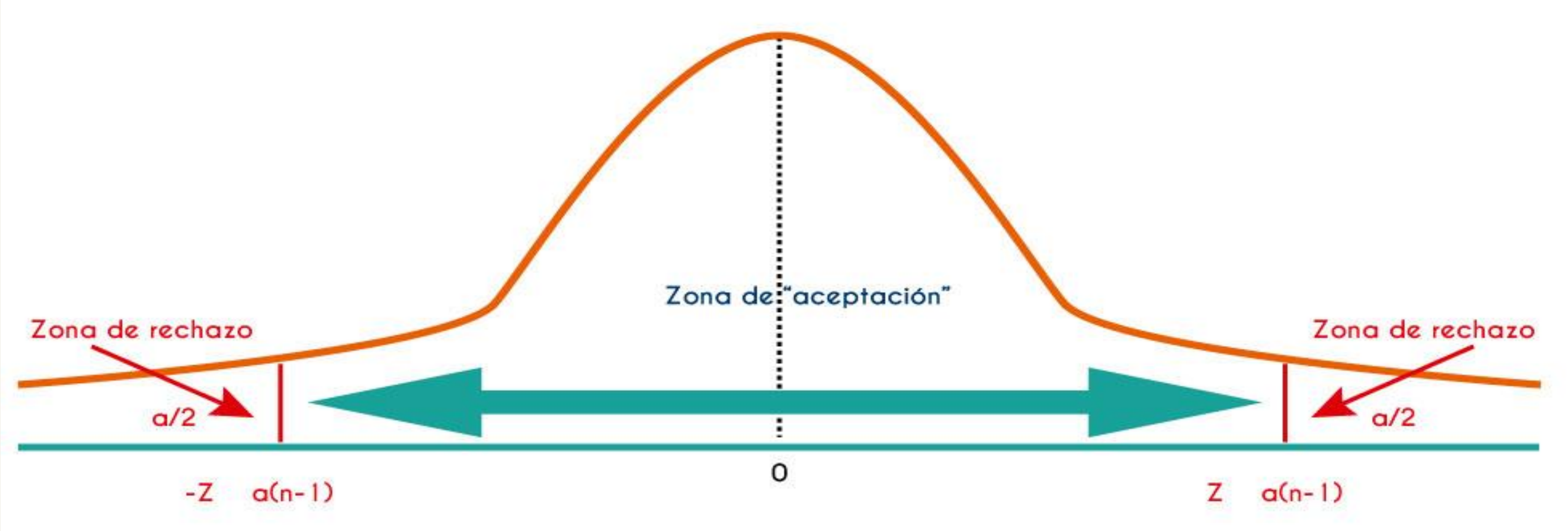
PASO 3 Determinación de la zona de aceptación y rechazo de la hipótesis nula (H_o)

Tomar una decisión estadística.

Siempre que se hace un estudio es necesario establecer con cuánta certeza llegamos a los resultados, esto es, el grado de error que nos permitimos.

En estadística esto se conoce como nivel de significancia o error tipo I e implica la seguridad para generalizar sin equivocarse. Se simboliza con la letra griega α y generalmente se trabaja con un error de 0.05, lo que quiere decir que el investigador tiene el 95% de certeza en los resultados obtenidos y sólo un 5% en contra. Este es el nivel de significancia o de error más usado, aunque hay otros niveles, pues si el investigador es muy estricto, puede trabajar con un error o nivel de significancia de 0.001 o incluso 0.0001.

Gráficamente el nivel de significancia o error bajo la curva normal es la siguiente:



Recuerda que la hipótesis que se somete a prueba es la hipótesis nula (Ho), por lo que las colas señalan en color rojo la zona de rechazo para la hipótesis nula (Ho).
Si establecemos el nivel de significancia de a=0.05 (5%), entonces a/2= 0.025m como se observa en la gráfica.
Para saber dónde está el límite de la zona de rechazo para nuestra muestra, elegimos un nivel de significancia de 5%.

$$1 - \alpha = \text{Nivel de confianza}$$
$$\alpha = \text{Nivel de significación}$$

A continuación te presentamos los valores de Z a diferentes niveles de confianza.

Nivel de confianza	Error	Valor de Z en tablas
90%	10%	1.645
91%	9%	1.663
92%	8%	1.681
93%	7%	1.699
94%	6%	1.71
95%	5%	1.96
96%	4%	2.06
97%	3%	2.08
98%	2%	2.101
99%	1%	2.575

Propiedades de los números pseudoaleatorios entre 0 y 1

En la sección anterior hablamos de cómo generar números aleatorios usando diferentes métodos. Sin embargo, ¿de qué manera se puede garantizar que tales números son realmente aleatorios entre 0 y 1? ¿Cuáles son las características que los identifican?, ¿cuáles son sus parámetros? La respuesta es muy importante, dado que los números aleatorios serán utilizados en la simulación para generar los valores de cualquier variable aleatoria. En gran medida, conocer las propiedades que deben tener estos números aleatorios garantiza una buena simulación, por ello, se enumeran a continuación.

Media de los aleatorios entre 0 y 1. En vista de que estos números deben tener la misma probabilidad de presentarse, es preciso que su comportamiento muestre una distribución de probabilidad uniforme continua, con límite inferior cero y límite superior uno. La función de densidad de una distribución uniforme es la siguiente:

$$f(x) = \frac{1}{b-a} \qquad a \leq x \leq b\text{; en este caso, } a = 0 \text{ y } b = 1$$

Para obtener la media de la distribución multiplicamos la función de densidad por x, y la integramos en todo el rango de la misma distribución de la siguiente manera:

$$E(x) = \int_a^b f(x)dx = \int_a^b \frac{x}{b-a}\,dx = \frac{x^2}{2(b-a)}\Big|_a^b$$

Sustituyendo los valores de $a = 0$ y $b = 1$.

$$E(x) = \frac{1}{2}$$

Por lo tanto, el valor esperado (es decir, la media de los números aleatorios entre 0 y 1) es $\mu = 0.5$.

Prueba de medias

Una de las propiedades que deben cumplir los números del conjunto r_i, es que el valor esperado sea igual a 0.5. La prueba que busca determinar lo anterior es la llamada *prueba de medias*, en la cual se plantean las siguientes hipótesis:

$$H_0: \mu_{r_i} = 0.5$$

$$H_1: \mu_{r_i} \neq 0.5$$

La prueba de medias consiste en determinar el promedio de los n números que contiene el conjunto r_i, mediante la ecuación siguiente:

$$\bar{r} = \frac{1}{n}\sum_{i=1}^{n} r_i$$

Después se calculan los límites de aceptación inferior y superior con las ecuaciones siguientes:

$$LI_{\bar{r}} = \frac{1}{2} - z_{\alpha/2}\left(\frac{1}{\sqrt{12n}}\right)$$

$$LS_{\bar{r}} = \frac{1}{2} + z_{\alpha/2}\left(\frac{1}{\sqrt{12n}}\right)$$

Si el valor de $\bar{r}$ se encuentra entre los límites de aceptación, concluimos que no se puede rechazar que el conjunto r_i tiene un valor esperado de 0.5 con un nivel de aceptación de $1 - \alpha$. En caso contrario se rechaza que el conjunto r_i tiene un valor esperado de 0.5.

Para el cálculo de los límites de aceptación se utiliza el estadístico $z_{\alpha/2}$, el cual se determina por medio de la tabla de la distribución normal estándar (ver apéndice).

Considere los 40 números del conjunto r_i que se presenta a continuación, y determine si tienen un valor esperado de ½ con un nivel de aceptación de 95 por ciento.

0.0449	0.1733	0.5746	0.049	0.8406	0.8349	0.92	0.2564
0.6015	0.6694	0.3972	0.7025	0.1055	0.1247	0.1977	0.0125
0.63	0.2531	0.8297	0.6483	0.6972	0.9582	0.9085	0.8524
0.5514	0.0316	0.3587	0.7041	0.5915	0.2523	0.2545	0.3044
0.0207	0.1067	0.3857	0.1746	0.3362	0.1589	0.3727	0.4145

El conjunto r_i contiene 40 números, por lo tanto, $n = 40$. Un nivel de aceptación de 95% implica que $\alpha = 5\%$. Enseguida procedemos a calcular el promedio de los números y los límites de aceptación:

$$\bar{r} = \frac{1}{n}\sum_{i=1}^{n} r_i = \frac{1}{40}\sum_{i=1}^{40} r_i$$

$$\bar{r} = \frac{1}{40}[0.04487 + 0.17328 + 0.57458 + 0.04901 + \ldots + 0.33616 + 0.15885 + 0.37266 + 0.41453$$

$$\bar{r} = 0.43250$$

$$LI_{\bar{r}} = \frac{1}{2} - z_{\alpha/2}\left(\frac{1}{\sqrt{12n}}\right) = \frac{1}{2} - z_{0.05/2}\left(\frac{1}{\sqrt{12(40)}}\right)$$

$$LI_{\bar{r}} = \frac{1}{2} - (1.96)\left(\frac{1}{\sqrt{12(40)}}\right) = 0.410538649$$

$$LS_{\bar{r}} = \frac{1}{2} + z_{\alpha/2}\left(\frac{1}{\sqrt{12n}}\right) = \frac{1}{2} + z_{0.05/2}\left(\frac{1}{\sqrt{12(40)}}\right)$$

$$LS_{\bar{r}} = \frac{1}{2} + (1.96)\left(\frac{1}{\sqrt{12(40)}}\right) = 0.589461351$$

Como el valor del promedio: $\bar{r} = 0.43250$ se encuentra entre los límites de aceptación, se concluye que no se puede rechazar que el conjunto de 40 números r_i tiene un valor esperado de 0.5, con un nivel de aceptación de 95 por ciento.

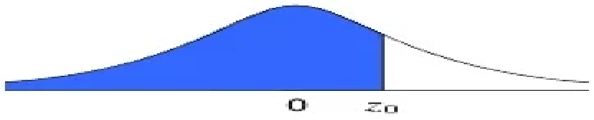
Tabla de la distribución normal N(0,1) para probabilidad acumulada inferior

μ = Media

σ = Desviación típica

$$P(z \le z_0) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z_0} e^{\frac{-z^2}{2}} dz$$

Tipificación: $z_0 = \dfrac{x-\mu}{\sigma}$



z_0	0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09	z_0
0,0	0,5000	0,5040	0,5080	0,5120	0,5160	0,5199	0,5239	0,5279	0,5319	0,5359	0,0
0,1	0,5398	0,5438	0,5478	0,5517	0,5557	0,5596	0,5636	0,5675	0,5714	0,5753	0,1
0,2	0,5793	0,5832	0,5871	0,5910	0,5948	0,5987	0,6026	0,6064	0,6103	0,6141	0,2
0,3	0,6179	0,6217	0,6255	0,6293	0,6331	0,6368	0,6406	0,6443	0,6480	0,6517	0,3
0,4	0,6554	0,6591	0,6628	0,6664	0,6700	0,6736	0,6772	0,6808	0,6844	0,6879	0,4
0,5	0,6915	0,6950	0,6985	0,7019	0,7054	0,7088	0,7123	0,7157	0,7190	0,7224	0,5
0,6	0,7257	0,7291	0,7324	0,7357	0,7389	0,7422	0,7454	0,7486	0,7517	0,7549	0,6
0,7	0,7580	0,7611	0,7642	0,7673	0,7704	0,7734	0,7764	0,7794	0,7823	0,7852	0,7
0,8	0,7881	0,7910	0,7939	0,7967	0,7995	0,8023	0,8051	0,8078	0,8106	0,8133	0,8
0,9	0,8159	0,8186	0,8212	0,8238	0,8264	0,8289	0,8315	0,8340	0,8365	0,8389	0,9
1,0	0,8413	0,8438	0,8461	0,8485	0,8508	0,8531	0,8554	0,8577	0,8599	0,8621	1,0
1,1	0,8643	0,8665	0,8686	0,8708	0,8729	0,8749	0,8770	0,8790	0,8810	0,8830	1,1
1,2	0,8849	0,8869	0,8888	0,8907	0,8925	0,8944	0,8962	0,8980	0,8997	0,9015	1,2
1,3	0,9032	0,9049	0,9066	0,9082	0,9099	0,9115	0,9131	0,9147	0,9162	0,9177	1,3
1,4	0,9192	0,9207	0,9222	0,9236	0,9251	0,9265	0,9279	0,9292	0,9306	0,9319	1,4
1,5	0,9332	0,9345	0,9357	0,9370	0,9382	0,9394	0,9406	0,9418	0,9429	0,9441	1,5
1,6	0,9452	0,9463	0,9474	0,9484	0,9495	0,9505	0,9515	0,9525	0,9535	0,9545	1,6
1,7	0,9554	0,9564	0,9573	0,9582	0,9591	0,9599	0,9608	0,9616	0,9625	0,9633	1,7
1,8	0,9641	0,9649	0,9656	0,9664	0,9671	0,9678	0,9686	0,9693	0,9699	0,9706	1,8
1,9	0,9713	0,9719	0,9726	0,9732	0,9738	0,9744	0,9750	0,9756	0,9761	0,9767	1,9
2,0	0,9772	0,9778	0,9783	0,9788	0,9793	0,9798	0,9803	0,9808	0,9812	0,9817	2,0
2,1	0,9821	0,9826	0,9830	0,9834	0,9838	0,9842	0,9846	0,9850	0,9854	0,9857	2,1
2,2	0,9861	0,9864	0,9868	0,9871	0,9875	0,9878	0,9881	0,9884	0,9887	0,9890	2,2
2,3	0,9893	0,9896	0,9898	0,9901	0,9904	0,9906	0,9909	0,9911	0,9913	0,9916	2,3
2,4	0,9918	0,9920	0,9922	0,9925	0,9927	0,9929	0,9931	0,9932	0,9934	0,9936	2,4
2,5	0,9938	0,9940	0,9941	0,9943	0,9945	0,9946	0,9948	0,9949	0,9951	0,9952	2,5
2,6	0,9953	0,9955	0,9956	0,9957	0,9959	0,9960	0,9961	0,9962	0,9963	0,9964	2,6
2,7	0,9965	0,9966	0,9967	0,9968	0,9969	0,9970	0,9971	0,9972	0,9973	0,9974	2,7
2,8	0,9974	0,9975	0,9976	0,9977	0,9977	0,9978	0,9979	0,9979	0,9980	0,9981	2,8
2,9	0,9981	0,9982	0,9982	0,9983	0,9984	0,9984	0,9985	0,9985	0,9986	0,9986	2,9
3,0	0,99865	0,99869	0,99874	0,99878	0,99882	0,99886	0,99889	0,99893	0,99896	0,99900	3,0
3,1	0,99903	0,99906	0,99910	0,99913	0,99916	0,99918	0,99921	0,99924	0,99926	0,99929	3,1
3,2	0,99931	0,99934	0,99936	0,99938	0,99940	0,99942	0,99944	0,99946	0,99948	0,99950	3,2
3,3	0,99952	0,99953	0,99955	0,99957	0,99958	0,99960	0,99961	0,99962	0,99964	0,99965	3,3
3,4	0,99966	0,99968	0,99969	0,99970	0,99971	0,99972	0,99973	0,99974	0,99975	0,99976	3,4
3,5	0,99977	0,99978	0,99978	0,99979	0,99980	0,99981	0,99981	0,99982	0,99983	0,99983	3,5
3,6	0,99984	0,99985	0,99985	0,99986	0,99986	0,99987	0,99987	0,99988	0,99988	0,99989	3,6
3,7	0,99989	0,99990	0,99990	0,99990	0,99991	0,99991	0,99992	0,99992	0,99992	0,99992	3,7
3,8	0,99993	0,99993	0,99993	0,99994	0,99994	0,99994	0,99994	0,99995	0,99995	0,99995	3,8
3,9	0,99995	0,99995	0,99996	0,99996	0,99996	0,99996	0,99996	0,99996	0,99997	0,99997	3,9

$1-\alpha$	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

www.vaxasoftware.com/indexes.html