

ARTICULO CIENTIFICO

PRUEBAS ESTADÍSTICAS EN NÚMEROS PSEUDOALEATORIOS

Merino Vidal Mateo Alejandro

Universidad Mayor de San Simón

Cochabamba, Bolivia

202301308@est.umss.edu

Abstract

El presente trabajo se realizó mediante el análisis de un conjunto de elementos representativos de la población, denominado muestra, con la finalidad de aplicar pruebas estadísticas que permitan medir la calidad de las secuencias de números pseudoaleatorios, basadas en los criterios de uniformidad e independencia. Cada prueba se fundamenta y explica para posteriormente aplicarse de forma práctica sobre una muestra conformada por valores dentro del rango [0,1], con el propósito de obtener los resultados experimentales y teóricos empleados en la decisión estadística.

Keywords: [uniformidad,independencia,frecuencia,prueba de promedios,chi-cuadrado,kolmogorov-smirnov,series,póker]

1. Introducción

En la actualidad, los sistemas informáticos dependen de valores generados al azar para realizar simulaciones, optimizar procesos y representar situaciones de la vida real. Estos valores permiten explorar distintos escenarios, analizar resultados posibles y comprender mejor el comportamiento de fenómenos complejos, considerando el factor de la aleatoriedad.

El uso de números aleatorios se extiende a múltiples campos como la criptografía, la estadística, la inteligencia artificial y la seguridad informática. Debido a esto, es posible crear llaves de encriptación, realizar muestreos o desarrollar modelos predictivos con el fin de analizar el comportamiento de un sistema. Sin embargo, los números producidos por un algoritmo o método computacional no provienen del azar puro, sino de fórmulas matemáticas que lo imitan, por lo que en realidad no se están generando números completamente aleatorios, por eso se les llama números pseudoaleatorios, ya que la única fuente generadora de números aleatorios puros es la naturaleza.

Dado que estas secuencias son generadas de manera determinista, es necesario comprobar si presentan el comportamiento de una verdadera muestra aleatoria. Para ello se utilizan pruebas estadísticas que analizan el cumplimiento de dos criterios importantes: la distribución equilibrada de los valores dentro del intervalo [0,1] y la ausencia de relación entre ellos.

Es importante verificar estos criterios, ya que de su cumplimiento depende la fiabilidad de los resultados que ofrece el generador. Un conjunto de números que no muestre uniformidad o independencia puede distorsionar los análisis y afectar la validez de cualquier simulación o modelo basado en ellos.

En el presente informe se aplican y analizan cinco pruebas estadísticas diseñadas para este propósito: la prueba de promedios (normal), la de frecuencias (chi-cuadrado), la de Kolmogórov-Smirnov, la de las series y la de póker. Cada una observa un aspecto distinto del comportamiento numérico, aportando evidencias sobre la confiabilidad de los valores generados.

2. Antecedentes

Desde hace siglos, el ser humano ha estado fascinado por los fenómenos impredecibles y por todo aquello que parece escapar a su control. La aleatoriedad ha formado parte de la vida cotidiana desde los juegos de azar hasta la toma de decisiones y la observación de la naturaleza. En sus inicios, los primeros intentos por estudiar y representar el azar se dieron mediante actividades simples como el lanzamiento de monedas, el uso de dados o la extracción de bolas de una urna.

Con el avance de la ciencia, estos experimentos dieron paso al desarrollo de modelos matemáticos más estructurados que buscaban describir y medir la incertidumbre. Sin embargo, a medida que los problemas se volvían más complejos, los métodos manuales resultaban insuficientes. Fue entonces cuando la aparición de las computadoras marcó un gran avance, ya que permitió generar grandes cantidades de datos en poco tiempo, estableciendo el punto de partida para la simulación de procesos aleatorios y de sistemas imposibles de analizar con los métodos tradicionales.

De esta necesidad surgieron los llamados generadores de números pseudoaleatorios, algoritmos capaces de producir secuencias numéricas que imitan el comportamiento del azar. Sin embargo, al estar basados en fórmulas matemáticas, se trata de procesos deterministas: a partir de un valor inicial o semilla, generan resultados que parecen impredecibles. No obstante, a pesar de su apariencia aleatoria, las secuencias pueden repetirse si se utiliza la misma semilla, demostrando su comportamiento predecible.

Con el tiempo, se comprendió que no todos los generadores ofrecían el mismo nivel de calidad. Algunos producían secuencias con períodos cortos o patrones que se repetían, mientras que otros lograban una distribución más homogénea. Para poder determinar si un conjunto de números era realmente “bueno” desde el punto de vista estadístico, se empezaron a desarrollar varias pruebas de aleatoriedad, destinadas a medir propiedades fundamentales como la independencia y la uniformidad.

Estas pruebas buscan responder a una pregunta fundamental: ¿la muestra obtenida proviene de una población verdaderamente aleatoria? Para ello, analizan si los valores están distribuidos de manera equilibrada dentro de su rango y si cada uno de ellos se comporta de forma independiente respecto al resto.

Actualmente, las pruebas de uniformidad e independencia son fundamentales en la evaluación de cualquier generador pseudoaleatorio, ya que constituyen la base para determinar si las secuencias pueden utilizarse de forma confiable en múltiples aspectos, como simulaciones, criptografía, modelado numérico o inteligencia artificial. El cumplimiento de estas propiedades garantiza que los resultados obtenidos no estén sesgados y que representen de manera válida los

fenómenos aleatorios que se desean replicar.

3. Marco Teórico

3.1 Probabilidad y Estadística

Las pruebas estadísticas en números pseudoaleatorios permiten determinar si una secuencia generada por un algoritmo se comporta como una sucesión verdaderamente aleatoria. Aunque estos números no provienen de un fenómeno físico, su análisis busca verificar si presentan las propiedades que se esperan del azar.

Una de las primeras características que se evalúan es la **uniformidad**, la cual hace referencia a la manera en que los números se distribuyen dentro de un intervalo dado. En un generador ideal, todos los valores posibles tienen la misma probabilidad de ocurrir, sin concentrarse en una región específica del rango.

- **Uniformidad:** Propiedad que asegura que los números pseudoaleatorios estén distribuidos de manera homogénea dentro del intervalo definido, generalmente $[0, 1]$. Si el generador es bueno, los números se reparten de forma equitativa.

Otra propiedad fundamental es la **independencia**. Esta garantiza que el valor de un número en la secuencia no influya en los que vienen después. En otras palabras, conocer un valor no permite predecir el siguiente, lo que mantiene el carácter aleatorio del conjunto.

- **Independencia:** Propiedad que indica que los valores generados no presentan relación entre sí. Cada número debe comportarse como un resultado aislado, sin patrones o correlaciones detectables.

Para verificar estas propiedades, las pruebas estadísticas utilizan la **frecuencia** con la que aparecen los números dentro de intervalos determinados. Si la secuencia es realmente uniforme, las frecuencias observadas deberían aproximarse a las esperadas.

- **Frecuencia:** Número de veces que un valor o grupo de valores aparece dentro de la secuencia. Se compara con la frecuencia teórica que se esperaría si los datos fueran perfectamente aleatorios.

En el análisis estadístico también se diferencian los conceptos de **población** y **muestra**, fundamentales para comprender la estimación de parámetros. La población representa el conjunto total de posibles observaciones, mientras que la muestra es una parte reducida de ella, obtenida para realizar inferencias.

- **Población:** Conjunto completo de elementos o resultados posibles en un estudio. En el contexto de números pseudoaleatorios, puede interpretarse como todos los números que un generador podría producir.
- **Muestra:** Subconjunto tomado de la población. En la práctica, es el conjunto finito de números generados que se analiza para inferir las propiedades de la secuencia total.

Dentro de la población existe un parámetro llamado **media poblacional**, que representa el valor esperado o promedio teórico de todos los posibles resultados. En una muestra, el equivalente es la **media muestral**, que estima la media poblacional a partir de los datos observados.

- **Media poblacional (μ):** Valor promedio de toda la población, calculado mediante la integral del producto de cada valor posible por su función de densidad.
- **Media muestral (\bar{x}):** Promedio de los valores observados en la muestra. Se calcula sumando los datos y dividiéndolos entre el tamaño de la muestra: $\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$.

Para medir la dispersión de los datos se utilizan la **desviación estándar** y la **varianza poblacional**. La varianza indica cuánto se alejan los valores respecto a la media, mientras que la desviación estándar es su raíz cuadrada, expresando la dispersión en las mismas unidades de los datos.

- **Varianza poblacional (σ^2):** Medida de la variabilidad total de la población, definida como la esperanza matemática del cuadrado de las desviaciones respecto a la media.
- **Desviación estándar (σ):** Raíz cuadrada de la varianza, que refleja el grado de dispersión promedio de los datos en torno a la media.

A continuación, se presentan las demostraciones matemáticas que sustentan el cálculo de estos parámetros, considerando una distribución uniforme continua en el intervalo $[0, 1]$.

Demostraciones

Media poblacional:

$$\mu = \int_0^1 xf(x) dx = \int_0^1 x(1) dx = \frac{x^2}{2} \Big|_0^1 = \frac{1}{2}$$

Varianza poblacional:

$$\sigma^2 = \int_0^1 (x - \mu)^2 f(x) dx = \int_0^1 (x - \frac{1}{2})^2 (1) dx$$

$$\sigma^2 = \int_0^1 (x^2 - x + \frac{1}{4}) dx = \left(\frac{x^3}{3} - \frac{x^2}{2} + \frac{x}{4} \right) \Big|_0^1 = \frac{1}{3} - \frac{1}{2} + \frac{1}{4} = \frac{1}{12}$$

Por tanto, para una distribución uniforme en el intervalo $[0, 1]$ se cumple que:

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

3.2 Pruebas de Aleatoriedad

3.2.1 Nivel de significancia α .

El símbolo α representa el *nivel de significancia* de una prueba: es el porcentaje de “riesgo” que estamos dispuestos a aceptar de equivocarnos al declarar que hay diferencia cuando en realidad no la hay (falso positivo). En la práctica, elegir $\alpha = 0.05$ significa trabajar con un 5% de tolerancia al error y, por complemento, con un 95% de confianza $(1 - \alpha)$. Valores más pequeños de α implican decisiones más exigentes.

3.2.2 Pruebas de uniformidad

Prueba de los promedios (“Normal”)

Esta prueba permite comprobar si la media de la muestra de números pseudoaleatorios coincide con la media teórica de una distribución uniforme $U(0, 1)$. Si los números son verdaderamente uniformes, el promedio de los valores generados debe ser cercano a 0.5.

Hipótesis:

$$H_0 : \mu = \frac{1}{2} \quad \text{vs} \quad H_1 : \mu \neq \frac{1}{2}$$

Procedimiento:

1. Determinar el tamaño de la muestra n .
2. Calcular la media muestral:

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$$

3. Calcular la desviación estándar teórica. Para una distribución $U(0, 1)$, se tiene $\sigma = \sqrt{1/12}$.
4. Calcular el estadístico de prueba:

$$Z_0 = \frac{(\bar{x} - 0.5)\sqrt{n}}{\sqrt{1/12}}$$

Este valor mide cuántas desviaciones estándar se encuentra la media muestral respecto a la teórica.

5. Determinar el valor crítico $z_{\alpha/2}$ a partir de la tabla de la distribución normal estándar, tal y como se observa en la Tab.1. Posteriormente, comparar.
6. Si $|Z_0| < z_{\alpha/2}$, se acepta la hipótesis nula y se concluye que los datos provienen de una distribución uniforme. En caso contrario, se rechaza H_0 .

Si $|Z_0| < z_{\alpha/2} \Rightarrow$ Uniformidad. Si $|Z_0| \geq z_{\alpha/2} \Rightarrow$ No uniformidad.

Table 1. Valores de la distribución normal estándar $N(0, 1)$

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	0.0000	0.0040	0.0080	0.0120	0.0160	0.0199	0.0239	0.0279	0.0319	0.0359
0.1	0.0398	0.0438	0.0478	0.0517	0.0557	0.0596	0.0636	0.0675	0.0714	0.0753
0.2	0.0793	0.0832	0.0871	0.0910	0.0948	0.0987	0.1026	0.1064	0.1103	0.1141
0.3	0.1179	0.1217	0.1255	0.1293	0.1331	0.1368	0.1406	0.1443	0.1480	0.1517
0.4	0.1554	0.1591	0.1628	0.1664	0.1700	0.1736	0.1772	0.1808	0.1844	0.1879
0.5	0.1915	0.1950	0.1985	0.2019	0.2054	0.2088	0.2123	0.2157	0.2190	0.2224
0.6	0.2257	0.2291	0.2324	0.2357	0.2389	0.2422	0.2454	0.2486	0.2517	0.2549
0.7	0.2580	0.2611	0.2643	0.2673	0.2704	0.2734	0.2764	0.2794	0.2823	0.2852
0.8	0.2881	0.2910	0.2939	0.2967	0.2995	0.3023	0.3051	0.3078	0.3106	0.3133
0.9	0.3159	0.3186	0.3212	0.3238	0.3264	0.3289	0.3315	0.3340	0.3365	0.3389
1.0	0.3413	0.3438	0.3461	0.3485	0.3508	0.3531	0.3554	0.3577	0.3599	0.3621
1.1	0.3643	0.3665	0.3686	0.3708	0.3729	0.3749	0.3770	0.3790	0.3810	0.3830
1.2	0.3849	0.3869	0.3888	0.3907	0.3925	0.3944	0.3962	0.3980	0.3997	0.4015
1.3	0.4032	0.4049	0.4066	0.4082	0.4099	0.4115	0.4131	0.4147	0.4162	0.4177
1.4	0.4192	0.4207	0.4222	0.4236	0.4251	0.4265	0.4279	0.4292	0.4306	0.4319
1.5	0.4332	0.4345	0.4357	0.4370	0.4382	0.4394	0.4406	0.4418	0.4429	0.4441
1.6	0.4452	0.4463	0.4474	0.4484	0.4495	0.4505	0.4515	0.4525	0.4535	0.4545
1.7	0.4554	0.4564	0.4573	0.4582	0.4591	0.4599	0.4608	0.4616	0.4625	0.4633
1.8	0.4641	0.4649	0.4656	0.4664	0.4671	0.4678	0.4686	0.4693	0.4699	0.4706
1.9	0.4713	0.4719	0.4726	0.4732	0.4738	0.4744	0.4750	0.4756	0.4761	0.4767
2.0	0.4772	0.4778	0.4783	0.4788	0.4793	0.4798	0.4803	0.4808	0.4812	0.4817
2.1	0.4821	0.4826	0.4830	0.4834	0.4838	0.4842	0.4846	0.4850	0.4854	0.4857
2.2	0.4861	0.4864	0.4868	0.4871	0.4875	0.4878	0.4881	0.4884	0.4887	0.4890
2.3	0.4893	0.4896	0.4898	0.4901	0.4904	0.4906	0.4909	0.4911	0.4913	0.4916
2.4	0.4918	0.4920	0.4922	0.4925	0.4927	0.4929	0.4931	0.4932	0.4934	0.4936
2.5	0.4938	0.4940	0.4941	0.4943	0.4944	0.4946	0.4947	0.4948	0.4950	0.4951
2.6	0.4952	0.4953	0.4955	0.4956	0.4957	0.4958	0.4959	0.4960	0.4961	0.4962
2.7	0.4963	0.4964	0.4965	0.4966	0.4967	0.4968	0.4969	0.4970	0.4971	0.4972
2.8	0.4973	0.4974	0.4974	0.4975	0.4976	0.4977	0.4977	0.4978	0.4979	0.4979
2.9	0.4980	0.4981	0.4981	0.4982	0.4982	0.4983	0.4984	0.4984	0.4985	0.4985
3.0	0.4987	0.4987	0.4987	0.4988	0.4988	0.4989	0.4989	0.4989	0.4990	0.4990

Prueba de frecuencias (Chi-cuadrado)

Esta prueba evalúa si los números de la muestra están uniformemente distribuidos dentro del intervalo $[0, 1]$, comparando las frecuencias observadas con las esperadas bajo el supuesto de uniformidad.

Hipótesis:

$$H_0 : \text{Los datos provienen de } U(0, 1) \quad H_1 : \text{No provienen de } U(0, 1)$$

Procedimiento:

1. Determinar el tamaño muestral n .
2. Elegir el número de clases o intervalos k y dividir el rango $[0, 1]$ en k subintervalos de igual amplitud $\frac{1}{k}$.
3. Calcular la frecuencia esperada:

$$FE = \frac{n}{k}$$

Esta representa el número de valores que teóricamente deberían caer en cada intervalo si la distribución fuera perfectamente uniforme.

4. Contar la frecuencia observada FO_i de cada intervalo, es decir, el número real de valores que se encuentran en él.
5. Calcular el estadístico de prueba:

$$\chi_0^2 = \sum_{i=1}^k \frac{(FO_i - FE)^2}{FE}$$

6. Obtener el valor crítico $\chi_{\alpha, k-1}^2$ de la tabla de la distribución Chi-cuadrado, tal y como se observa en la Tab.2. Posteriormente, comparar.
7. Si $\chi_0^2 < \chi_{\alpha, k-1}^2$, se acepta H_0 y se concluye que los datos son uniformes; en caso contrario, se rechaza H_0 .

Si $\chi_0^2 < \chi_{\alpha, k-1}^2 \Rightarrow$ Uniformidad. Si $\chi_0^2 \geq \chi_{\alpha, k-1}^2 \Rightarrow$ No uniformidad.

Table 2. Valores de la distribución Chi-cuadrado (χ^2) para distintos niveles de significancia

ν	0.995	0.990	0.975	0.950	0.500	0.250	0.010	0.005
1	0.00	0.00	0.00	0.00	0.45	1.32	6.63	7.88
2	0.01	0.02	0.05	0.10	1.39	2.77	9.21	10.60
3	0.07	0.11	0.22	0.35	2.37	4.11	11.34	12.84
4	0.21	0.30	0.48	0.71	3.36	5.39	13.28	14.86
5	0.41	0.55	0.83	1.15	4.35	6.63	15.09	16.75
6	0.68	0.87	1.24	1.64	5.35	7.84	16.81	18.55
7	0.99	1.24	1.69	2.17	6.35	9.04	18.48	20.28
8	1.34	1.65	2.18	2.73	7.34	10.22	20.09	21.96
9	1.73	2.09	2.70	3.33	8.34	11.39	21.67	23.59
10	2.16	2.56	3.25	3.94	9.34	12.55	23.21	25.19
11	2.60	3.05	3.82	4.57	10.34	13.70	24.72	26.76
12	3.07	3.57	4.40	5.23	11.34	14.85	26.22	28.30
13	3.57	4.11	5.01	5.89	12.34	15.98	27.69	29.82
14	4.07	4.66	5.63	6.57	13.34	17.12	29.14	31.32
15	4.60	5.23	6.27	7.26	14.34	18.25	30.58	32.80
16	5.14	5.81	6.91	7.96	15.34	19.37	32.00	34.27
17	5.70	6.41	7.56	8.67	16.34	20.49	33.41	35.72
18	6.26	7.01	8.23	9.39	17.34	21.60	34.81	37.16
19	6.84	7.63	8.91	10.12	18.34	22.72	36.19	38.58
20	7.43	8.26	9.59	10.85	19.34	23.83	37.57	40.00
25	10.52	11.52	13.12	14.61	24.34	28.66	44.31	46.93
30	13.79	14.95	16.79	18.49	29.34	33.67	50.89	53.67
40	20.71	22.16	24.43	26.51	39.34	43.77	63.69	66.77
50	27.99	29.71	32.36	34.76	49.33	53.68	76.15	79.49
60	35.53	37.48	40.48	43.19	59.33	63.30	88.38	91.95
70	43.28	45.44	48.76	51.74	69.33	72.65	100.42	104.22
80	51.17	53.54	57.15	60.39	79.33	81.84	112.33	116.32
90	59.20	61.75	65.65	69.13	89.33	91.86	124.12	128.30
100	67.33	70.06	74.22	77.93	99.33	101.88	135.81	140.17

Prueba de Kolmogórov–Smírnov

Esta prueba compara la función de distribución empírica de la muestra con la teórica $F(x) = x$ de la distribución uniforme. No requiere agrupar los datos y es adecuada para tamaños muestrales pequeños y medianos.

Hipótesis:

$$H_0 : F(x) = x \text{ para todo } x \in [0, 1] \quad H_1 : F(x) \neq x$$

Procedimiento:

1. Ordenar los datos en forma ascendente $x_{(1)} \leq x_{(2)} \leq \dots \leq x_{(n)}$.
2. Calcular la función empírica acumulada para cada $x_{(i)}$:

$$F_n(x_{(i)}) = \frac{i}{n}$$

3. Calcular el estadístico:

$$D = \max_{1 \leq i \leq n} |F_n(x_{(i)}) - x_{(i)}|$$

4. Obtener el valor crítico $D_\alpha(n)$ según el tamaño muestral n y el nivel de significancia α , de la tabla del test de Kolmogórov–Smírnov, tal y como se observa en la Tab.3. Posteriormente, comparar.
5. Si $D < D_\alpha(n)$, se acepta H_0 ; si $D \geq D_\alpha(n)$, se rechaza H_0 .

Si $D < D_\alpha(n) \Rightarrow$ Uniformidad. Si $D \geq D_\alpha(n) \Rightarrow$ No uniformidad.

Table 3. Valores críticos $D_\alpha(n)$ para la prueba de Kolmogórov–Smírnov

Tamaño de muestra n	$\alpha = 0.10$	$\alpha = 0.05$	$\alpha = 0.01$
1	0.950	0.975	0.995
2	0.776	0.842	0.929
3	0.642	0.708	0.829
4	0.564	0.624	0.734
5	0.510	0.563	0.669
6	0.470	0.521	0.618
7	0.438	0.486	0.577
8	0.411	0.457	0.543
9	0.388	0.432	0.514
10	0.368	0.409	0.486
11	0.352	0.391	0.468
12	0.338	0.375	0.450
13	0.352	0.361	0.433
14	0.314	0.349	0.418
15	0.304	0.338	0.404
16	0.295	0.328	0.392
17	0.286	0.318	0.381
18	0.278	0.309	0.371
19	0.272	0.301	0.363
20	0.264	0.294	0.352
25	0.240	0.264	0.317
30	0.220	0.242	0.290
35	0.210	0.230	0.270
40	0.210	0.230	0.252
50	0.188	0.210	0.226
60	0.172	0.192	0.207
70	0.160	0.182	0.192
80	0.150	0.170	0.180
90	0.141	0.160	0.171
100	0.134	0.150	0.162

Para $n > 100$, $D_{0.10} \approx \frac{1.22}{\sqrt{n}}$, $D_{0.05} \approx \frac{1.36}{\sqrt{n}}$, $D_{0.01} \approx \frac{1.63}{\sqrt{n}}$.

3.2.3 Pruebas de independencia

Prueba de las series

Esta prueba evalúa si los números generados son independientes entre sí, observando las relaciones entre números consecutivos. Si los pares (x_i, x_{i+1}) están distribuidos uniformemente sobre el cuadrado unitario, se considera que los números son independientes.

Procedimiento:

1. Formar los pares consecutivos $(x_1, x_2), (x_2, x_3), \dots, (x_{n-1}, x_n)$.
2. Dividir el cuadrado unitario $[0, 1] \times [0, 1]$ en $k \times k$ celdas de igual tamaño.
3. Contar el número de pares que caen en cada celda (i, j) , obteniendo las frecuencias observadas FO_{ij} .
4. Calcular la frecuencia esperada:

$$FE = \frac{n - 1}{k^2}$$

5. Calcular el estadístico:

$$\chi_0^2 = \frac{k^2}{n - 1} \sum_{i=1}^k \sum_{j=1}^k (FO_{ij} - FE)^2$$

6. Obtener el valor crítico $\chi_{\alpha, k^2 - 1}^2$ de la tabla de la distribución Chi-cuadrado, tal y como se observa en la Tab.2. Posteriormente, comparar.
7. Si $\chi_0^2 < \chi_{\alpha, k^2 - 1}^2$, se concluye independencia; de lo contrario, dependencia.

Si $\chi_0^2 < \chi_{\alpha, k^2 - 1}^2 \Rightarrow$ Independencia. Si $\chi_0^2 \geq \chi_{\alpha, k^2 - 1}^2 \Rightarrow$ No independencia.

Prueba de Póker

Esta prueba analiza los cinco dígitos significativos de cada número pseudoaleatorio, clasificándolos en categorías similares a las jugadas de póker. Si los dígitos son independientes, las frecuencias de las categorías deben coincidir con las probabilidades teóricas.

Categorías y probabilidades teóricas:

Sea n el número de dígitos significativos (por ejemplo, $n = 5$). Las categorías y sus probabilidades se determinan considerando todas las posibles combinaciones de dígitos entre 0 y 9, por tanto el total de casos es 10^n . Las fórmulas generales son:

$$\begin{aligned}
 P(\text{Todos distintos, TD}) &= \frac{10 \times 9 \times 8 \times 7 \times 6}{10^n} \\
 P(1 \text{ Par, 1P}) &= \frac{10 \times 1 \times 9 \times 8 \times 7}{10^n} \binom{5}{2} \\
 P(2 \text{ Pares, 2P}) &= \frac{10 \times 1 \times 9 \times 1 \times 8}{10^n} \binom{5}{2} \binom{3}{2} \\
 P(\text{Trío, T}) &= \frac{10 \times 1 \times 1 \times 9 \times 8}{10^n} \binom{5}{3} \\
 P(\text{Full, F}) &= \frac{10 \times 1 \times 1 \times 9 \times 1}{10^n} \binom{5}{3} \binom{2}{2} \\
 P(\text{Póker, P}) &= \frac{10 \times 1 \times 1 \times 1 \times 9}{10^n} \binom{5}{4} \\
 P(\text{Quintilla, Q}) &= \frac{10 \times 1 \times 1 \times 1 \times 1}{10^n} \binom{5}{5}
 \end{aligned}$$

Nota: Estas expresiones se pueden ajustar a otros valores de n (por ejemplo, 3 o 4 dígitos), modificando la cantidad de factores del numerador. En tales casos, algunas configuraciones (como Full o Quintilla) podrían no formarse.

Procedimiento:

1. Extraer los cinco dígitos significativos de cada número.
2. Clasificar cada número en una de las siete categorías.
3. Contar las frecuencias observadas FO_j .
4. Calcular las frecuencias esperadas:

$$FE_j = n \times p_j$$

donde p_j representa la probabilidad teórica de formar cada tipo de combinación o “mano” (como todos distintos, pareja, trío, etc.), calculada mediante las fórmulas combinatorias correspondientes.

Si alguna $FE_j < 5$, combinar categorías hasta que todas las frecuencias esperadas sean mayores o iguales a 5.

5. Calcular el estadístico:

$$\chi_0^2 = \sum_j \frac{(FO_j - FE_j)^2}{FE_j}$$

6. Obtener el valor crítico $\chi_{\alpha, v}^2$ con $v = m - 1$ grados de libertad de la tabla de la distribución Chi-cuadrado, tal y como se observa en la Tab.2. Posteriormente, comparar.
7. Si $\chi_0^2 < \chi_{\alpha, v}^2$, se acepta independencia; si no, dependencia.

Si $\chi_0^2 < \chi_{\alpha, v}^2 \Rightarrow$ Independencia. Si $\chi_0^2 \geq \chi_{\alpha, v}^2 \Rightarrow$ No independencia.

4. Descripción del Problema

En la generación computacional de números aleatorios, los valores obtenidos no provienen del azar puro, sino de algoritmos deterministas que buscan imitarlo. Estos métodos, denominados generadores pseudoaleatorios, producen secuencias numéricas que parecen aleatorias, pero cuya estructura depende de parámetros iniciales y fórmulas predefinidas.

El principal problema radica en que no todas las secuencias generadas poseen la misma calidad estadística. Algunas pueden presentar acumulaciones en ciertas regiones del intervalo o patrones de repetición entre valores consecutivos, lo que afecta la representatividad de los resultados y distorsiona cualquier análisis o simulación que dependa de dichos números.

Para evaluar la confiabilidad de estas secuencias, es necesario verificar dos propiedades esenciales: la uniformidad y la independencia. La primera garantiza que los valores estén distribuidos de manera homogénea en el rango $[0, 1]$, mientras que la segunda asegura que cada número se genere sin depender de los anteriores.

El análisis de estas propiedades se realiza mediante distintas pruebas estadísticas, como las de promedios, frecuencias, Kolmogórov–Smírnov, series y póker. Cada una detecta desviaciones específicas en el comportamiento de los datos y permite determinar si los números generados pueden considerarse suficientemente aleatorios para su aplicación en estudios y modelos computacionales.

5. Objetivos

- Aplicar distintas pruebas estadísticas a una secuencia de números pseudoaleatorios con el fin de evaluar su grado de uniformidad e independencia.
- Analizar los resultados obtenidos en cada prueba, identificando posibles desviaciones respecto a los valores teóricos esperados según la distribución uniforme.
- Comparar la capacidad de detección de cada prueba (promedios, chi–cuadrado, Kolmogórov–Smírnov, series y póker), determinando qué aspectos del comportamiento pseudoaleatorio evalúan con mayor precisión.
- Presentar los fundamentos teóricos, el procedimiento y la interpretación de los resultados de manera estructurada, garantizando la validez del análisis estadístico aplicado a la muestra.

6. Desarrollo de la Solución

6.1 Metodología e implementación

Para la experimentación se desarrolló un conjunto de funciones en el lenguaje **R**, ejecutadas en el entorno **RStudio** y documentadas en **LATEX**. El propósito fue aplicar y comprobar las principales pruebas estadísticas de uniformidad e independencia sobre una muestra de números pseudoaleatorios.

El código se estructuró de forma modular, donde cada prueba cuenta con su propia función y un bloque principal encargado de la interacción con el usuario. En total, el sistema implementa cinco pruebas: de promedios (normal), de frecuencias (chi–cuadrado), de Kolmogórov–Smírnov, de series y de póker. Todas permiten evaluar diferentes aspectos del comportamiento pseudoaleatorio de una secuencia.

El programa principal (`main`) solicita al usuario el tipo de prueba a ejecutar, los valores de la muestra, el nivel de significancia γ , en los casos necesarios, el número de intervalos k . Con base en esa información, el sistema dirige el flujo hacia la función correspondiente y presenta el resultado con el estadístico calculado, el valor crítico y la decisión final sobre la hipótesis.

Las funciones principales implementadas son las siguientes:

- `pruebaPromedios()`: Calcula el estadístico Z_0 para comprobar si la media muestral coincide con 0.5, valor esperado en una distribución uniforme. Evalúa la hipótesis de uniformidad según el nivel de significancia ingresado.
- `pruebaFrecuencia()`: Divide el intervalo $[0, 1]$ en k subintervalos, cuenta las frecuencias observadas y compara con las esperadas mediante el estadístico χ^2 . Determina si los datos siguen una distribución uniforme.
- `pruebaKolmogorov()`: Ordena la muestra, construye la función empírica acumulada y obtiene el máximo desvío respecto a la función teórica $F(x) = x$. Compara con el valor crítico D_α de la tabla de Kolmogórov-Smírnov.
- `pruebaSeries()`: Forma pares consecutivos (x_i, x_{i+1}) , los ubica dentro de una malla $k \times k$ y aplica el contraste Chi-cuadrado sobre las frecuencias bidimensionales, verificando la independencia entre los valores.
- `pruebaPoker()`: Extrae los cinco dígitos significativos de cada número y los clasifica en categorías combinatorias (todos distintos, pareja, trío, etc.). Calcula las frecuencias esperadas según probabilidades teóricas derivadas de combinatoria y evalúa independencia con χ^2 .

Cada ejecución muestra en consola los valores de los estadísticos calculados y teóricos, junto con la interpretación automática del resultado (*uniformidad* o *independencia*). Además, se agregaron condiciones para fusionar categorías cuando alguna frecuencia esperada resulta menor que 5, asegurando la validez del contraste.

En las siguientes figuras se incluyen capturas del funcionamiento del programa en **RStudio**, mostrando ejemplos de entrada, resultados de salida y validaciones obtenidas para cada una de las pruebas implementadas.

Para los ensayos prácticos se utilizó una muestra de 50 valores pseudoaleatorios comprendidos en el intervalo $[0, 1]$, generados de forma uniforme. Se adoptó un nivel de significancia de $\alpha = 0.05$ (5 %) para todas las pruebas, y cuando la metodología lo requirió (como en las pruebas de frecuencias y series), se fijó un número de intervalos de $k = 3$.

La muestra utilizada fue la siguiente:

```
0.14382, 0.63759, 0.02716, 0.77493, 0.58247, 0.31968, 0.45832, 0.89541, 0.67425, 0.23897,
0.04638, 0.72154, 0.51963, 0.95327, 0.18549, 0.37285, 0.83612, 0.25746, 0.61379, 0.49723,
0.05871, 0.79134, 0.66745, 0.32489, 0.11957, 0.94481, 0.46319, 0.28675, 0.53964, 0.71238,
0.35918, 0.84572, 0.09734, 0.56581, 0.41697, 0.19365, 0.64852, 0.21948, 0.97813, 0.73649,
0.15483, 0.87961, 0.08247, 0.27953, 0.68714, 0.39452, 0.59386, 0.48729, 0.12675, 0.99238
```

En las siguientes figuras se incluyen capturas del funcionamiento del programa en **RStudio**, mostrando ejemplos de entrada, resultados de salida y validaciones obtenidas para cada una de las pruebas implementadas.

Prueba de Promedios (Normal)

```
> main()
Pruebas Estadisticas
1:Prueba de promedios/normal(uniformidad)
2:Prueba de frecuencias/chi-cuadrado(uniformidad)
3:Prueba de las series(independencia)
4:Prueba de Kolmogorov (uniformidad)
5:Prueba de poker (independencia)
Elija la prueba que quiere realizar: 1
Ingresa los elementos de la muestra separados por espacios: 0.14382 0.63759 0.02716 0.77493 0.58247 0.31968 0.45832 0.89541 0.67425 0.2389
7285 0.83612 0.25746 0.61379 0.49723 0.05871 0.79134 0.66745 0.32489 0.11957 0.94481 0.46319 0.28675 0.53964 0.71238 0.35918 0.84572 0.09
97813 0.73649 0.15483 0.87961 0.08247 0.27953 0.68714 0.39452 0.59386 0.48729 0.12675 0.99238
Ingresa el nivel de significancia en porcentaje sobre el 100%: 5
estadisticoZo: 0.2916069
estadisticoTeoricoZo: 1.959964
La muestra proviene de una poblacion uniforme con media igual a 1/2. Por tanto, posee "uniformidad"
```

Figure 1. Ejecución de la prueba de promedios o normal.

Prueba de Frecuencias (Chi-cuadrado)

```
> main()
Pruebas Estadisticas
1:Prueba de promedios/normal(uniformidad)
2:Prueba de frecuencias/chi-cuadrado(uniformidad)
3:Prueba de las series(independencia)
4:Prueba de Kolmogorov (uniformidad)
5:Prueba de poker (independencia)
Elija la prueba que quiere realizar: 2
Ingresa los elementos de la muestra separados por espacios: 0.14382 0.63759 0.02716 0.77493 0.58247 0.31968 0.45832 0.89541 0.67425 0.2389
7285 0.83612 0.25746 0.61379 0.49723 0.05871 0.79134 0.66745 0.32489 0.11957 0.94481 0.46319 0.28675 0.53964 0.71238 0.35918 0.84572 0.09
97813 0.73649 0.15483 0.87961 0.08247 0.27953 0.68714 0.39452 0.59386 0.48729 0.12675 0.99238
Ingresa el nivel de significancia en porcentaje sobre el 100%: 5
Ingresa el numero de intervalos k: 3
chiCalculado: 0.16
chiTeorico: 5.991465
La muestra proviene de una poblacion uniforme. Por tanto, posee "uniformidad"
```

Figure 2. Ejecución de la prueba de frecuencias o chi-cuadrado.

Prueba de Kolmogórov-Smírnov

```
> main()
Pruebas Estadisticas
1:Prueba de promedios/normal(uniformidad)
2:Prueba de frecuencias/chi-cuadrado(uniformidad)
3:Prueba de las series(independencia)
4:Prueba de Kolmogorov (uniformidad)
5:Prueba de poker (independencia)
Elija la prueba que quiere realizar: 4
Ingresa los elementos de la muestra separados por espacios: 0.14382 0.63759 0.02716 0.77493 0.58247 0.31968 0.45832 0.89541 0.67425 0.2389
7285 0.83612 0.25746 0.61379 0.49723 0.05871 0.79134 0.66745 0.32489 0.11957 0.94481 0.46319 0.28675 0.53964 0.71238 0.35918 0.84572 0.09
97813 0.73649 0.15483 0.87961 0.08247 0.27953 0.68714 0.39452 0.59386 0.48729 0.12675 0.99238
Ingresa el nivel de significancia en porcentaje sobre el 100%: 5
dMaximoObservado: 0.06351
dTeorico: 0.188
La muestra proviene de una poblacion uniforme. Por tanto, posee "uniformidad"
```

Figure 3. Ejecución de la prueba de Shitikov (Kolmogórov-Smírnov).

Prueba de las Series

```
> main()
Pruebas Estadísticas
1:Prueba de promedios/normal(uniformidad)
2:Prueba de frecuencias/chi-cuadrado(uniformidad)
3:Prueba de las series(independencia)
4:Prueba de Kolmogorov (uniformidad)
5:Prueba de poker (independencia)
Elija la prueba que quiere realizar: 3
Ingrera los elementos de la muestra separados por espacios: 0.14382 0.63759 0.02716 0.77493 0.58247 0.31968 0.45832 0.89541 0.67425 0.23897
7285 0.83612 0.25746 0.61379 0.49723 0.05871 0.79134 0.66745 0.32489 0.11957 0.94481 0.46319 0.28675 0.53964 0.71238 0.35918 0.84572 0.09734
97813 0.73649 0.15483 0.87961 0.08247 0.27953 0.68714 0.39452 0.59386 0.48729 0.12675 0.99238
Ingrera el nivel de significancia en porcentaje sobre el 100%: 5
Ingrera el numero de intervalos k: 3
factor 0.1836735
par 0.14382 0.63759
par 0.63759 0.02716
par 0.02716 0.77493
par 0.77493 0.58247
par 0.58247 0.31968
par 0.31968 0.45832
par 0.45832 0.89541
par 0.89541 0.67425
par 0.67425 0.23897
par 0.23897 0.04638
par 0.04638 0.72154
par 0.72154 0.51963
par 0.51963 0.95327
par 0.95327 0.18549
par 0.18549 0.37285
par 0.37285 0.83612
par 0.83612 0.25746
par 0.25746 0.61379
par 0.61379 0.49723
par 0.49723 0.05871
par 0.05871 0.79134
par 0.79134 0.66745
par 0.66745 0.32489
par 0.32489 0.11957
par 0.11957 0.94481
par 0.94481 0.46319
```

Figure 4. Ejecución de la prueba de las series.

```
par 0.94481 0.46319
par 0.46319 0.28675
par 0.28675 0.53964
par 0.53964 0.71238
par 0.71238 0.35918
par 0.35918 0.84572
par 0.84572 0.09734
par 0.09734 0.56581
par 0.56581 0.41697
par 0.41697 0.19365
par 0.19365 0.64852
par 0.64852 0.21948
par 0.21948 0.97813
par 0.97813 0.73649
par 0.73649 0.15483
par 0.15483 0.87961
par 0.87961 0.08247
par 0.08247 0.27953
par 0.27953 0.68714
par 0.68714 0.39452
par 0.39452 0.59386
par 0.59386 0.48729
par 0.48729 0.12675
par 0.12675 0.99238
3 7 7 7 4 5 8 5 3
5.444444
chicCalculado: 5.183673
chiTeorico: 15.50731
La muestra proviene de una poblacion independiente. Por tanto, posee "independencia"
```

Figure 5. Ejecución de la prueba de las series.

Prueba de Póker

```
> main()
Pruebas Estadísticas
1:Prueba de promedios/normal(uniformidad)
2:Prueba de frecuencias/chi-cuadrado(uniformidad)
3:Prueba de las series(independencia)
4:Prueba de Kolmogorov (uniformidad)
5:Prueba de poker (independencia)
Elija la prueba que quiere realizar: 5
Ingresá los elementos de la muestra separados por espacios: 0.14382 0.63759 0.02716 0.77493 0.58247 0.31968 0.45832 0.89541 0.67425 0.2389
7285 0.83612 0.25746 0.61379 0.49723 0.05871 0.79134 0.66745 0.32489 0.11957 0.94481 0.46319 0.28675 0.53964 0.71238 0.35918 0.84572 0.09
97813 0.73649 0.15483 0.87961 0.08247 0.27953 0.68714 0.39452 0.59386 0.48729 0.12675 0.99238
Ingresá el nivel de significancia en porcentaje sobre el 100%: 5
44 6 0
15.12 25.2 9.68
chiCalculado: 79.4709
chiTeorico: 5.991465
La muestra proviene de una población no independiente
```

Figure 6. Ejecución de la prueba de póker.

7. Conclusiones

- Al usar varias pruebas (promedios, frecuencias, Kolmogórov–Smírnov, series y póker) se obtuvo una mirada completa de la muestra: unas revisan cómo se reparten los valores y otras si los datos se relacionan entre sí.
- La decisión final no depende solo del número que sale en el cálculo, sino también del tamaño de la muestra, del nivel de significancia elegido y, cuando aplica, del número de intervalos; si estos parámetros se eligen mal, las conclusiones pueden ser engañosas.
- La prueba de frecuencias funciona bien, pero exige que las “cajas” tengan suficientes datos; cuando no es así, combinar categorías evita errores y hace más confiable el resultado.
- La prueba de Kolmogórov–Smírnov es útil para notar si, en general, la muestra se parece a una distribución uniforme sin tener que armar tablas; resulta clara en muestras pequeñas y medianas.
- Las pruebas de independencia (series y póker) mostraron que una secuencia puede verse “pareja” en el reparto de valores y, aun así, tener patrones entre términos o entre dígitos; por eso no basta con revisar solo la uniformidad.
- En conjunto, las pruebas confirman que para confiar en una secuencia pseudoaleatoria hace falta que coincidan los dos aspectos: que los valores se repartan bien (uniformidad) y que no dependan unos de otros (independencia); solo así sirve con confianza en simulaciones y modelos.

8. Bibliografía

References

- [1] García Gómez, José Alfredo; Martínez De La Cruz, Miguel Ángel; Jauregui Wade, Lucila; Valles Rivera, Diana; Sánchez Vasconcelos, Ángel Gabriel. (2024). *Importancia del uso de los generadores de números pseudoaleatorios en los contenidos de la asignatura de*

- simulación*. Innovación y Desarrollo Tecnológico, Vol. 16, Núm. 3. Instituto Tecnológico de Villahermosa.
- [2] ScienceDirect. (s.f.). *Random Number Generation*. Recuperado de: <https://www.sciencedirect.com/topics/computer-science/random-number-generation>
 - [3] Biebighauser, D. (2019). *Testing Random Number Generators*. Informe técnico universitario, Universidad de Minnesota. Recuperado de: <https://www-users.cse.umn.edu/~garrett/students/reu/pRNGs.pdf>
 - [4] Maldonado, M. J.; Álvarez, G.; Ramírez, P. (2024). *Comprehensive Method for Measuring Randomness in Pseudo-Random Number Generators*. Revista Mexicana de Ciencias de la Computación, Vol. 11, Núm. 3, pp. 155–178.
 - [5] Lorek, P.; Kubiak, M.; Gajek, J. (2020). *On testing pseudorandom generators via statistical goodness-of-fit and dependence tests*. Simulation Modelling Practice and Theory, Vol. 106, Art. 102183.