Técnicas de Muestreo I

Patricia Isabel Romero Mares

Departamento de Probabilidad y Estadística
IIMAS UNAM

agosto 2023

Números aleatorios

¿Qué son los números aleatorios?

Es una secuencia de números que cumplen con que:

- los valores tienen una distribución conocida sobre un intervalo definido.
- es imposible predecir valores futuros basados en los valores pasados o presentes.

¿Para qué se usan en Estadística?

- Selección de muestras aleatorias.
- Asignación de tratamientos a las unidades experimentales en experimentos.
- Estudios de simulación de procesos estocásticos, cálculo de integrales complejas, simulación de procesos físicos y biológicos, entre otros.
- Métodos de remuestreo (bootstrap, jackknife).

¿Cómo se genera una secuencia de números aleatorios?

- Antes del desarrollo de métodos confiables para la generación de números aleatorios, se construyeron tablas de éstos generadas por un procedimiento físico.
- Las tablas más usadas fueron las ”tablas RAND”(RAND Corporation, 1955) que contienen un millón de dígitos aleatorios generados por un proceso físico.
- Estos dígitos aleatorios han sido sujetos a muchas pruebas estadísticas sin evidencia de que no provienen de una distribución uniforme discreta sobre los puntos 0,...,9.

Ejemplo de una página de las tablas RAND

400

TABLE OF RANDOM DIGITS

19950	92630	78240	19267	95457	53497	23894	37708	79862	76471	66418
19951	79445	78735	71549	44843	26104	67318	00701	34986	66751	99723
19952	59654	71966	27386	50004	05358	94031	29281	18544	52429	06080
19953	31524	49587	76612	39789	13537	48086	59483	60680	84675	53014
19954	06348	76938	90379	51392	55887	71015	09209	79157	24440	30244
19955	28703	51709	94456	48396	73780	06436	86641	69239	57662	80181
19956	68108	89266	94730	95761	75023	48464	65544	96583	18911	16391
19957	99938	90704	93621	66330	33393	95261	95349	51769	91616	33238
19958	91543	73196	34449	63513	83834	99411	58826	40456	69268	48562
19959	42103	02781	73920	56297	72678	12249	25270	36678	21313	75767
19960	17138	27584	25296	28387	51350	61664	37893	05363	44143	42677
19961	28297	14280	54524	21618	95320	38174	60579	08089	94999	78460
19962	09331	56712	51333	06289	75345	08811	82711	57392	25252	30333
19963	31295	04204	93712	51287	05754	79396	87399	51773	33075	97061
19964	36146	15560	27592	42089	99281	59640	15221	96079	09961	05371
19965	29553	18432	13630	05529	02791	81017	49027	79031	50912	09399
19966	23501	22642	63081	08191	89420	67800	55137	54707	32945	64522
19967	57888	85846	67967	07835	11314	01545	48535	17142	08552	67457
19968	55336	71264	88472	04334	63919	36394	11196	92470	70543	29776
19969	10087	10072	55980	64688	68239	20461	89381	93809	00796	95945
19970	34101	81277	66090	88872	37818	72142	67140	50785	21380	16703
19971	53362	44940	60430	22834	14130	96593	23298	56203	92671	15925
19972	82975	66158	84731	19436	55790	69229	28661	13675	99318	76873
19973	54827	84673	22898	08094	14326	87038	42892	21127	30712	48489
19974	25464	59098	27436	89421	80754	89924	19097	67737	80368	08795
19975	67609	60214	41475	84950	40133	02546	09570	45682	50165	15609
19976	44921	70924	61295	51137	47596	86735	35561	76649	18217	63446
19977	33170	30972	98130	95828	49786	13301	36081	80761	33985	68621
19978	84687	85445	06208	17654	51333	02878	35010	67578	61574	20749
19979	71886	56450	36567	09395	96951	35507	17555	35212	69106	01679
19980	00475	02224	74722	14721	40215	21351	08596	45625	83981	63748
19981	25993	38881	68361	59560	41274	69742	40703	37993	03435	18873
19982	92882	53178	99195	93803	56985	53089	15305	50522	55900	43026
19983	25138	26810	07093	15677	60688	04410	24505	37890	67186	62829
19984	84631	71882	12991	83028	82484	90339	91950	74579	03539	90122
19985	34003	92326	12793	61453	48121	74271	28363	66561	75220	35908
19986	53775	45749	05734	86169	42762	70175	97310	73894	88606	19994
19987	59316	97885	72807	54966	60859	11932	35265	71601	55577	67715
19988	20479	66557	50705	26999	09854	52591	14063	30214	19890	19292
19989	86180	84931	25455	26044	02227	52015	21820	50599	51671	65411
19990	21451	68001	72710	40261	61281	13172	63819	48970	51732	54113
19991	98062	68375	80089	24135	72355	95428	11808	29740	81644	86610
19992	01788	64429	14430	94575	75153	94576	61393	96192	03227	32258
19993	62465	04841	43272	68702	01274	05437	22953	18946	99053	41690
19994	94324	31089	84159	92933	99989	89500	91586	02802	69471	68274
19995	05797	43984	21575	09908	70221	19791	51578	36432	33494	79888
19996	10395	14289	52185	09721	25789	38562	54794	04897	59012	89251
19997	35177	56986	25549	59730	64718	52630	31100	62384	49483	11409
19998	25633	89619	75882	98256	02126	72099	57183	55887	09320	73463
19999	16464	48280	94254	45777	45150	68865	11382	11782	22695	41988

¿Cómo se genera una secuencia de números aleatorios?

- Se han generado secuencias de números aleatorios a través de procesos físicos en la computadora, por ejemplo, utilizando cierta medición en los drives de los discos o en los semiconductores.
- Existe una serie de procedimientos determinísticos que generan números llamados **pseudoaleatorios**, los cuales se **comportan** como una muestra aleatoria con una distribución conocida.


DILBERT By Scott Adams
TOUR OF ACCOUNTING
OVER HERE WE HAVE OUR RANDOM NUMBER GENERATOR.
www.dilbert.com
scottadams@aol.com
NINE NINE NINE NINE NINE NINE
© 2001 United Feature Syndicate, Inc.
ARE YOU SURE THAT'S RANDOM?
THAT'S THE PROBLEM WITH RAN-DOMNESS: YOU CAN NEVER BE SURE.

Características de los números pseudoaleatorios

- Los procedimientos se especializan en generar números pseudoaleatorios con distribución uniforme entre 0 y 1 $U(0,1)$ y estos se transforman a cualquier distribución de probabilidad.
- Existen los generadores lineales congruenciales, de la forma:

  $$x_i = (ax_{i-1} + c) \bmod m \quad 0 \leq x_i < m$$

  Si $c = 0$ se llama generador congruencial multiplicativo
  Si $c \neq 0$ se llama generador congruencial mixto
- Se calcula $u_i = x_i/m;\ \ u_i \in (0,1)$

Características de los números pseudoaleatorios

- La **semilla** de este generador es el valor inicial x_0.
- Si se empieza con la misma semilla, la secuencia de números aleatorios generada es la misma (reproducibilidad).
- m es el periodo del generador, es decir, es la longitud del ciclo de la secuencia, a partir del número $m+1$ se repite la secuencia.

Procedimientos para generar números aleatorios en R

- Existen muchos procedimientos para generar números aleatorios.
- R tiene 8 diferentes.
- Los números aleatorios generados por estos procedimientos pasan una serie de pruebas estadísticas para medir su bondad.

Procedimientos para generar números aleatorios en R

- Wichmann-Hill
- Marsaglia-Multicarry
- Super-Duper
- **Mersenne-Twister** Es el que usa R si no se especifica
- Knuth-TAOCP-2002
- Knuth-TAOCP
- L'Ecuyer-CMRG
- ‘User-supplied”

El generador de Wichmann-Hill como ejemplo

$$
\begin{aligned}
x_i &= 171 x_{i-1} \operatorname{mod} 30269 \\
y_i &= 172 y_{i-1} \operatorname{mod} 30307 \\
z_i &= 170 z_{i-1} \operatorname{mod} 30323
\end{aligned}
$$

$$
u_i = \left(\frac{x_i}{30269} + \frac{y_i}{30307} + \frac{z_i}{30323} \right) \operatorname{mod} 1
$$

- Tres semillas x_0, y_0, z_0
- El periodo es del orden de 10^{12}

Algunas distribuciones que se pueden generar con R

- uniforme - *runif*
- normal - *rnorm*
- Ji-cuadrada - *rchisq*
- t - *rt*
- F - *rf*
- binomial - *rbinom*
- Poisson - *rpois*
- geométrica - *rgeom*

Ejemplo de simulación, cálculo de π

Supongamos que hay un cuadrado con lados igual a uno, por lo tanto, el área del cuadrado es uno.

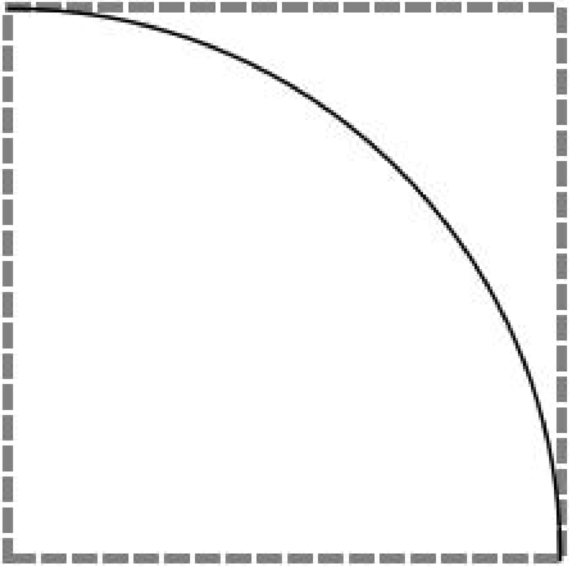
Inscribamos un cuarto de círculo en el cuadrado, centrado en el vértice inferior izquierdo (véase figura).

Este cuarto de círculo es parte de un círculo con radio uno.

El círculo completo tiene área igual a π, por lo que el cuarto de círculo tiene área $\pi/4$.

Vamos a estimar esta área.

Ejemplo de simulación, cálculo de π



Ejemplo de simulación, cálculo de π

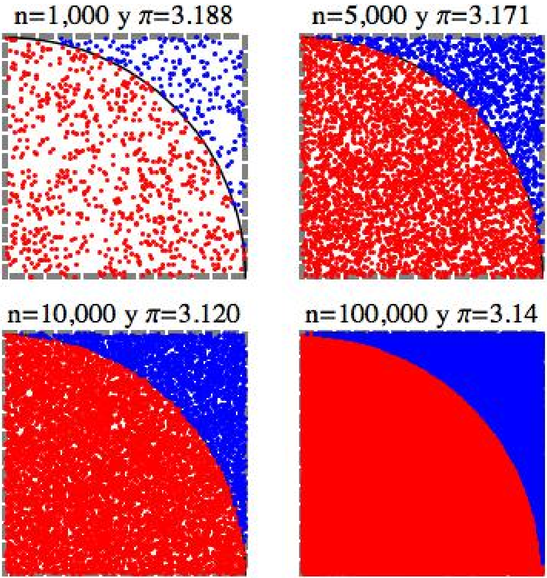
Generaremos al azar puntos en el cuadrado, es decir, generaremos puntos de la forma (x, y) donde x y y se distribuyen uniformemente en el $(0, 1)$.

La proporción de puntos que se localicen dentro del círculo nos dará una estimación de $\pi/4$, así que habrá que multiplicar esta proporción por cuatro para estimar π.

En la siguiente gráfica se muestran cuatro simulaciones con diferente número de puntos generados y el valor aproximado de π.

Se observa que a medida que generamos más puntos mejora la estimación.

Ejemplo de simulación, cálculo de π



Programa de R

```
N <- 100000
puntos <- rep(0,N)
x <- runif(N)
y <- runif(N)
puntos <- (x^2 + y^2) <= 1
pi <- sum(puntos)/N*4
```

Bibliografía

- Easterday K. y Smith T. (1991). *A Monte Carlo Application to Approximate Pi*. The Mathematics Teacher, Vol. 84, No. 5
- Gentle, J.E. (2003). *Random Number Generation and Monte Carlo Methods*. 2nd. ed. Springer
- Romero, P. y Rueda, R. (2013). *Ars Conjectandi y el Método Montecarlo*. Revista Digital Universitaria, Vol. 14, No. 11
- The RAND Corporation (1955). *A Million Random Digits with 100,000 Normal Deviates*. The Free Press, New York