Simulación de una ruleta

Delmonti Agustín
Universidad Tecnológica Nacional Rosario
Zeballos 1341, S2000
agus.delmonti@gmail.com

Abril 2020

ABSTRACT

En este trabajo se realiza la simulación y análisis de resultados de una ruleta europea simple a traves del método de Monte Carlo. El objetivo es sentar las bases y profundizar la comprensión de los conceptos estadísticos utilizados en la simulación de sistemas.

1. Introducción

La ruleta es uno de los juegos de azar mas simples y antiguos. No hay necesidad de simular una ruleta para saber cuales son las probabilidades de que nuestra jugada sea la ganadora o perdedora: estos son datos bien conocidos y fácilmente calculables gracias a los matemáticos y estadísticos que han estudiado el juego desde su origen. Es debido a estas características que la ruleta es el objeto perfecto de estudio para entender las bases de la simulación, pues permite experimentar simulando el sistema y luego comparar los resultados con valores perfectamente conocidos o formular hipótesis corroborables.

El objetivo de este practico es simular una ruleta europea simple y obtener empíricamente valores aproximados a los parámetros de la población. Para esto se crea un modelo de la ruleta que, si bien simplificado, representa de manera acertada su comportamiento.

2. Marco teórico

2.1. Modelización

En la forma mas simple, una ruleta puede ser modelada como una función o caja negra cuya salida es un numero entero aleatorio entre 0 y 36. Cada tirada de la ruleta se obtiene un valor $x \in [0; 36]$ con una probabilidad p_x desconocida asociada al evento de ocurrencia de dicho numero.

2.2. Población

En este trabajo, inicialmente suponemos la distribución de los valores de la ruleta europea como desconocida desconocida. Sin embargo podemos calcular a priori los parámetros teóricos de la población, con el objetivo de que sirvan de guía o comparación a lo largo de los experimentos.

Si todos los eventos son equiprobables, es decir $p_1 = p_2 = ... = p$, entonces la población sigue una *distribución uniforme discreta* $X \sim U(0, 36)$. Si X esta distribuida uniformemente en el conjunto $\{0, 1, 2, ..., N\}$ inmediatamente se desprenden las siguientes ecuaciones:

$$P(X = x) = \frac{1}{N} = \frac{1}{b - a + 1} = \frac{1}{37} \tag{1}$$

$$\mu = \frac{a + b}{2} = 18 \tag{2}$$

$$Var(X) = \frac{N^2 - 1}{12} = \frac{(b - a + 1)^2 - 1}{12} = 114 \tag{3}$$

$$\sigma = \sqrt{114} \tag{4}$$

A través de muestreos sucesivos a la población (tiradas) podemos obtener aproximaciones a los valores teóricos reales con un grado alto de confianza y así poder inferir la distribución de la variable, en lo que se denomina una simulación de Monte Carlo.

2.3. Generación números aleatorios enteros

En la practica, la generación de números es pseudoaleatoria. Los generadores de números pseudoaleatorios generan una secuencia de números, que si bien no son verdaderamente aleatorios, sus propiedades se aproximan suficiente a las propiedades de las secuencias de números aleatorios reales.

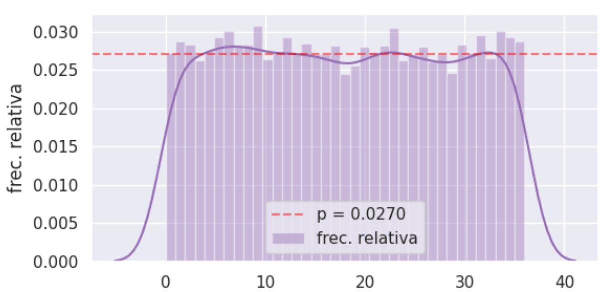
El algoritmo utilizado por los generadores es **determinista**: está completamente determinado por un valor inicial denominado *semilla* o *seed*, la cual si es conocida, se puede utilizar para volver a generar exactamente la misma secuencia de números. En otras palabras, una vez generada la semilla, cada elemento de la sucesión esta predeterminado. Esto por ende garantiza reproducibilidad de los experimentos simulados. [1]

Para mejores resultados y para garantizar la independencia entre los diferentes experimentos, en cada simulación se genera una semilla diferente.

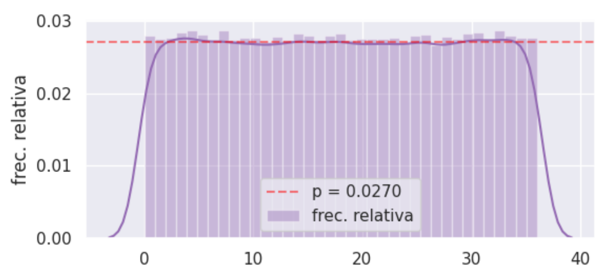
3. Simulación simple

3.1. Distribución

La distribución de frecuencias relativas de muestras obtenida en una corrida de la simulación pueden verse en la figura 1. La distribución presenta una tendencia hacia la uniformidad, cuando el tamaño de la muestra *n* es relativamente grande, lo que coincide con la formulación inicial sobre la población.



(a) muestra n = 10k



(b) muestra n = 100k

Figura 1: distribución de frecuencias de muestras tomada al azar de la simulación.

Como el modelo de la ruleta simplemente es un generador de números pseudoaleatorios, también podemos concluir que el generador toma los números de una distribución uniforme.

3.2. Estructura de la simulación

La simulación de la ruleta se basa en el muestreo sucesivo de los resultados de una tirada. Cada tirada obtenemos un valor $x \in [0, 36]$ generado por el generador de números pseudoaleatorios. Luego se recalculan los estadísticos de la muestra hasta la iteración actual. Este algoritmo se repite hasta llegar al al hiperparametro del modelo N. Así obtendremos una gráfica de la frecuencia relativa de un numero a, la media y la desviación de las sucesivas muestras a medida que aumentan las iteraciones. (Ver algoritmo 1)

Algorithm 1 Ruleta

```
N ← iter
tiradas ← []
frq, mean, std ← []
z ← numero

for k ← 0 to N − 1 do
    a ← random(0, 36)
    append a to tiradas

    z ← count z in tiradas
    append z / k to frq

    append tiradas.mean() to frq
    append tiradas.std() to frq
graficar(frq, mean, std)
```

3.3. Métricas de evaluación

Los resultados de varias simulaciones se pueden observar en la figura 2 (corrida corta), en la figura 3 con (múltiples corridas independientes), y en simulaciones mas largas como la figuras 4a y 4b donde los valores por iteración se promedian entre corridas independientes. La gráfica es errática al principio y tiende a amortiguarse a medida que crece el tamaño n de la muestra, al rededor de los valores teóricos esperados. Esto es lógico, pues mientras mas información tengamos (muestra mas grande) mejores conclusiones se pueden sacar sobre el comportamiento de cierto sistema o modelo (la población) ya que el subconjunto de la muestra representa una parte mas representativa de la población.

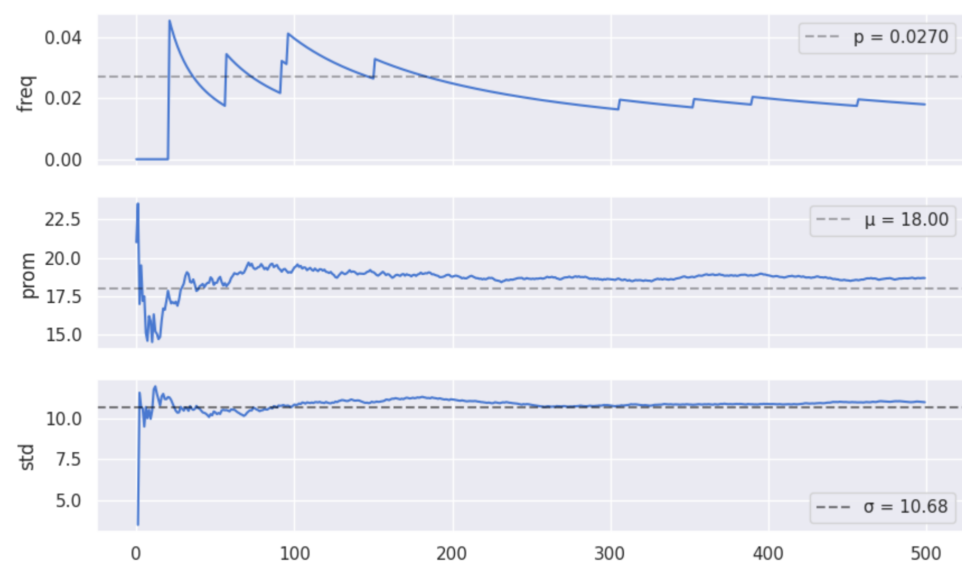


Figura 2: muestra aleatoria (n = 500) de la simulación. (a) Frecuencia relativa x = 7 acumulada. (b) Promedio acumulado. (c) Desviación estándar acumulada.

3.4. Ley de los números grandes

La ley de los números grandes es un teorema que describe el resultado de repetir un experimento un gran numero de veces. Según esta ley, el promedio de los resultados obtenidos de un numero grande de ensayos debe ser cercano al valor esperado y a medida que se conducen mas ensayos tenderá a volverse todavía mas cercano a este valor esperado. [2, p. 181-190][3, Cap 2].

Figura 3: cinco simulaciones independientes (n = 500). (a) Frecuencia relativa acumulada x = 7. (b) Promedio acumulado. (c) Desviación estándar acumulado.

La ley débil de los números grandes establece que un secuencia infinita $X_1, X_2, ...$ de variables aleatorias independientes e idénticamente distribuidas (i.i.d) con un valor esperado $E(X_1) = E(X_2) = ...$, el promedio de la muestra converge en μ

$$\overline{X_n} = \frac{X_1 + X_1 + ... + X_n}{n} \to \mu \qquad \text{cuando } n \to \infty \tag{5}$$

En otras palabras, para cualquier número positivo ϵ se tiene

$$\lim_{n\to\infty} P(|\overline{X}_n - \mu| > \epsilon) = \lim_{n\to\infty} P(|\frac{X_1 + X_1 + ... + X_n}{n} - \mu| > \epsilon) = 0 \tag{6}$$

Interpretando este resultado, la ley débil establece que con una muestra suficientemente grande, habrá una probabilidad muy alta de que el promedio de las observaciones esté cerca del valor esperado μ, con un margen determinado por ϵ. Este fenómeno se puede observar mejor en los resultados de la simulación simple (figura 2) cuando la media se aproxima a μ a medida que crece el tamaño de la muestra.

3.5. Método de Monte Carlo

El método de Monte Carlo es un método de simulación basado en la ley de números grandes para inferir la media de la población. La amplia clase de algoritmos computacionales basados en el método de Monte Carlo que se basan en *muestreo aleatorio repetido* para obtener *resultados numéricos*. [4]

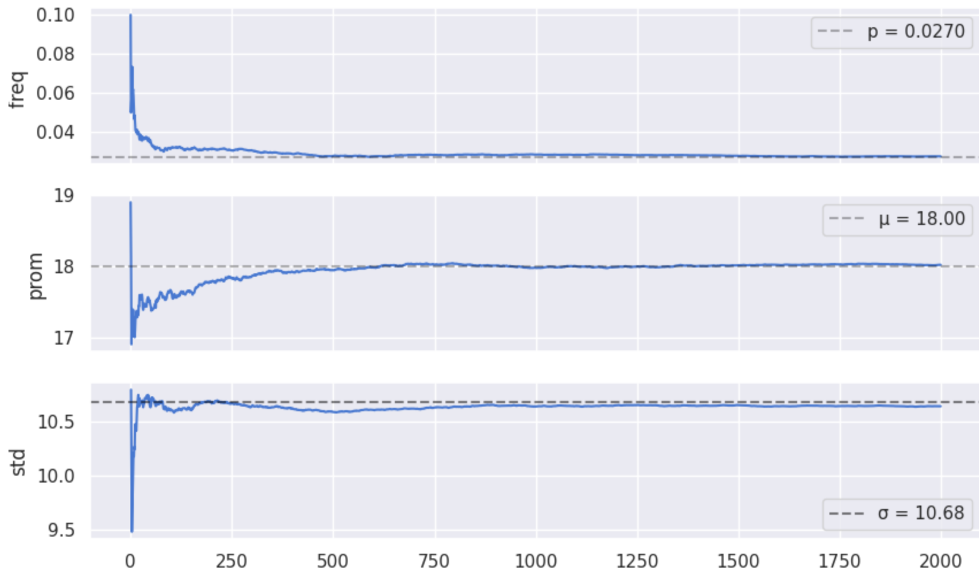
En el caso de la ruleta, el muestreo reiterado se logra con cada tirada independiente de la ruleta, cada una descrita por una variable aleatoria $X_1, X_2, ..., X_n$ proveniente de la misma distribución desconocida a estimar. La independencia y distribución idéntica de las variables permite aplicar la ley débil de los números grandes (ver ecuación 6). Además, la ley explicita que cuanto mayor es el tamaño n de la muestra , mejor tiende a ser la aproximación de la media.

La ley de números grandes es importante porque garantiza resultados estables a largo plazo del promedio de algún evento aleatorio. Por ejemplo, si bien un casino puede perder dinero en un solo giro de la ruleta, sus ganancias tenderán a un porcentaje predecible en un gran número de giros. Cualquier racha ganadora de un jugador eventualmente será vencida por los parámetros reales del juego.[2, p. 181-190]

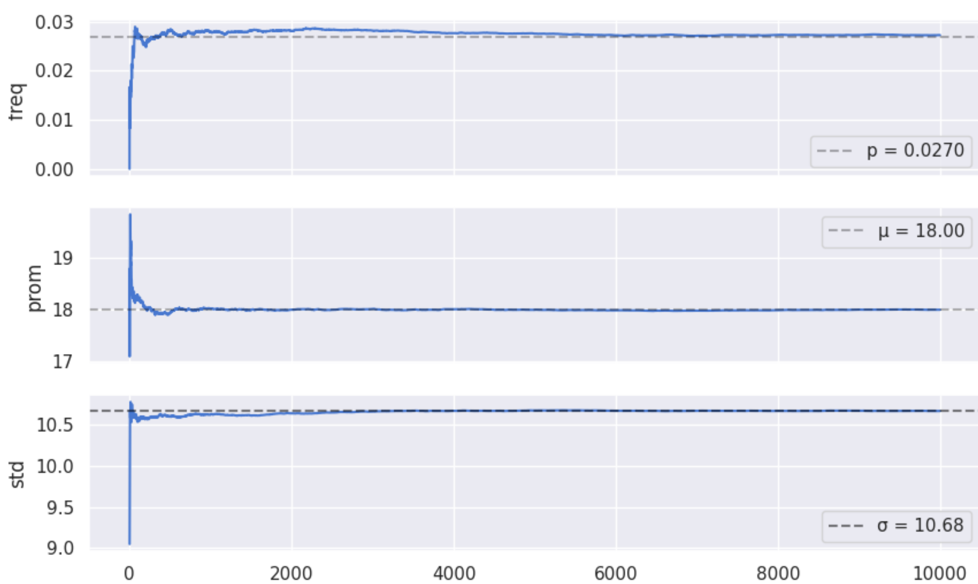
Esto se puede observar claramente en la figura 3, donde 5 simulaciones independientes, sin importar el comportamiento inicial, convergen en los valores esperados a medida que los tamaños de las muestras tienden a valores relativamente grandes.

Aclaración final del método. Si bien sabemos que la simulación converge al valor esperado porque conocemos los parámetros de la población a priori, la ley de los números grandes de todas maneras nos **garantiza** una aproximación

Figura 4: Estadísticos de simulaciones múltiples promediadas por iteración.



(a) Promedio de 30 simulaciones (n = 2k)



(b) Promedio de 30 simulaciones (n = 10k).

al valor real a medida que el tamaño de la muestra crece. Es decir, en cualquier otra simulación que no tengamos los valores esperados como dato, podemos utilizar una o varias simulaciones de Monte Carlo para aproximarnos a los valores reales con un alto grado de confianza.

La razón por la cual este método es importante es principalmente porque, a veces, es difícil o imposible utilizar otros enfoques.

4. Conclusión

A través de simulaciones de Monte Carlo, podemos muestrear sucesivamente un evento aleatorio para obtener aproximaciones de los valores esperados de la población con un grado alto de confianza. Estos métodos resultan útiles cuando los sistemas son complejos en su modelización.

En el infinito, la frecuencia relativa de un elemento de la muestra tiende a la probabilidad real, el promedio muestral $\overline{x}$ tiende a μ y la desviación estándar muestral s tiende a σ.

En el caso de la ruleta, si bien una persona puede tener rachas de victorias, la ley de los números grandes nos asegura que el Casino nunca resultará perdedor. Mientras la persona individual esta interesada en jugar a lo máximo 100 tiradas de ruleta antes de aburrirse y marcharse del lugar, el casino esta interesado en que sucede en 1 millón.

Referencias

[1] Pseudorandom number generator. Pseudorandom number generator — Wikipedia, the free encyclopedia, 2020. [Online; accedido 12 de Abril de 2020].

[2] Michel Dekking. *A modern introduction to probability and statistics: understanding why and how*. Springer, 2010.

[3] Sheldon M. Ross. *Simulation*. Academic, 5th edition, 2012.

[4] Monte Carlo Method. Monte carlo method — Wikipedia, the free encyclopedia, 2020. [Online; accedido 9 de Abril de 2020].