Naipes, Algoritmos de Aleatorización y Valores Esperados

Matías Salim Seda Auil
Universidad de Chile

En el siguiente trabajo se analizará un algoritmo de aleatorización a través de un caso específico de estudio: la aleatorización de un mazo de cartas a través de un método particular. Se estudiará el tiempo promedio esperado del algoritmo, su relación con los números armónicos y se complementarán los resultados teóricos con datos obtenidos a través de una simulación computacional que replica el proceso estudiado.

PRESENTACIÓN DEL PROBLEMA: ALEATORIZAR UN MAZO DE CARTAS

Supóngase que tiene un mazo de cartas que se encuentra ordenado. Gustaría tener un procedimiento (algoritmo) para aleatorizar la baraja de cartas, es decir, una forma para *desordenar* la cartas tales que todas se encuentran es posiciones aleatorias (en contraste a las posiciones iniciales que, ya que el mazo está ordenado, son posiciones conocidas).

Existen muchos métodos para aleatorizar una baraja de cartas aunque, en este caso, el interés se centrará en el siguiente procedimiento:

1. Inicialmente, con la baraja ordenada, se toma la carta que se encuentra en la posición más inferior del mazo y se coloca en un lugar aleatorio del mazo.
2. Se repite el procedimiento anterior con la nueva carta (puede ser la misma carta anterior) que se encuentra en la posición más inferior del mazo.
3. El procedimiento se repite hasta que, finalmente, la carta que se hallaba, inicialmente, en la posción más superior del mazo, se encuentra en la posición más inferior. En este punto, se ubica esta última carta en un lugar cualquiera de la baraja y luego el mazo está completamente aleatorizada.

Respecto al algoritmo, la pregunta en particular que se quiere resolver es: ¿En promedio, en cuántos pasos el mazo queda completamente aleatorizado?

Nótese que, fundamental, la duración del algoritmo anterior depende de la carta que, inicialmente, se encontraba en la posición más superior del mazo. Así, preguntarse por la duración del algoritmo (la cantidad de pasos que se necesitan para aleatorizar las cartas en la baraja) es equivalente a preguntarse en cuántos pasos la carta más superior en el mazo llega a la posición más inferior de la baraja, más un último paso extra (aleatorizar la última carta).

RESOLUCIÓN DEL PROBLEMA: CONTANDO LOS PASOS

Supóngase que se tiene una baraja de n cartas. Se define la variable aleatoria X como la cantidad de pasos necesarios para aleatorizar dicha baraja con el algoritmo descrito anteriormente. Se quiere conocer el valor de $E[X]$, es decir, el valor esperado para la cantidad de pasos necesarios para aleatorizar el mazo.

Como se sabe, la cantidad de pasos necesarios para aleatorizar la baraja es equivalente a la cantidad de pasos necesarios para que la carta más superior en el mazo llega a la posición más inferior de la baraja, más un paso extra. Definiendo la variable aleatoria Y como la cantidad de pasos para que la carta más superior en el mazo llega a la posición más inferior de la baraja, se tiene que:

$$E[X] = E[Y] + 1$$

Respecto a Y, la cantidad de pasos necesarios para mover la carta superior a la posición más inferior es equivalente a la cantidad de pasos necesarios para mover la carta superior desde su posición inicial a la posición siguiente, luego de la posición siguiente a la subsiguiente y así hasta la posición más inferior. Así, se define la variable aleatoria Y_i como la cantidad de pasos necesarios para mover la carta superior de la posición i a la posición $i+1$. Por ejemplo, Y_1 es la variable aleatoria que mide la cantidad de pasos necesarios para mover la carta superior de la posición 1 (primera posición de la carta, es decir, la posición más superior en la baraja) a la posición 2 (posición inferior siguiente a la posición más superior en la baraja).

De esa forma, la cantidad de pasos necesarios para mover la carta superior a la posición más inferior de la barajas es equivalente a la cantidad de pasos necesarios para mover la carta superior desde la posición 1 a la posición 2 más la cantidad de pasos necesarios para mover la carta superior desde la posición 2 a la posición 3 y así hasta finalmente hasta la posción n. Luego, se tiene que:

$$Y = \sum_{i=1}^{n-1} Y_i$$

Y, por lo tanto, se cumple que:

$$E[Y] = \sum_{i=1}^{n-1} E[Y_i]$$

Para $E[Y_i]$, por definición, se tiene que:

$$E[Y_i] = \sum_{y \in R_{Y_i}} y \cdot P\{Y_i = y\}$$

En particular, $P\{Y_i = y\}$ es la probabilidad de mover la carta superior de la posición i a la posición $i+1$ en y pasos. Ahora, para mover la carta superior de la posición i a la posición $i+1$ se necesita que, cuando se aleatorizce la carta más inferior de la baraja, ésta quede en alguna posición menor o igual a la posición i. La probabilidad de que la carta más inferior se coloque en alguna posición menor o igual a la posición i es $\frac{i}{n}$. Así, se tiene que $P\{Y_i = y\} = (1 - \frac{i}{n})^{y-1} \cdot \frac{i}{n}$ (dado que se quiere que se realice en y pasos). Además, $1 \leq y < \infty$, dado que la carta más inferior se puede colocar en alguna posición menor o igual a la posición i en 1 paso o más. Así, se tiene que:

$$E[Y_i] = \sum_{y=1}^{\infty} y \cdot (1 - \frac{i}{n})^{y-1} \cdot \frac{i}{n}$$

Desarrollando la expresión:

$$E[Y_i] = \frac{i}{n} \cdot \sum_{y=0}^{\infty} (y+1) \cdot (1 - \frac{i}{n})^{y}$$

$$E[Y_i] = \frac{i}{n} \cdot \left(\sum_{y=0}^{\infty} y \cdot (1 - \frac{i}{n})^{y} + \sum_{y=0}^{\infty} (1 - \frac{i}{n})^{y} \right)$$

Dado que $1 - \frac{i}{n} < 1$, se tiene que:

$$E[Y_i] = \frac{i}{n} \cdot \left(\frac{1 - \frac{i}{n}}{(1 - (1 - \frac{i}{n}))^2} + \frac{1}{1 - (1 - \frac{i}{n})} \right)$$

$$E[Y_i] = \frac{i}{n} \cdot \left(\frac{1 - \frac{i}{n}}{\frac{i}{n}^2} + \frac{1}{\frac{i}{n}} \right)$$

$$E[Y_i] = \frac{i}{n} \cdot \left(\frac{n}{i} \right)^2$$

$$E[Y_i] = \frac{n}{i}$$

Ahora, como $E[Y] = \sum_{i=1}^{n-1} E[Y_i]$, se tiene que:

$$E[Y] = \sum_{i=1}^{n-1} \frac{n}{i}$$

Recordando la definición del número armónico $H_n = \sum_{i=1}^{n} \frac{1}{i}$, se tiene que:

$$E[Y] = n \cdot H_{n-1}$$

Por lo tanto, $E[X]$, es decir, el valor esperado para la cantidad de pasos necesarios para aleatorizar el mazo viene dado por la siguiente expresión:

$$E[X] = n \cdot H_{n-1} + 1$$

ANÁLISIS DE LA SOLUCIÓN: SIMULANDO CON CARTAS

Los resultados teóricos entregan una solución explícita para el valor promedio de pasos en el que se ejecuta el algoritmo. Sin embargo, gustaría conocer más a fondo el comportamiento de la variable aleatoria X, es decir, como se comporta el tiempo de ejecucion del algoritmo de aleatorización. Para realizar lo dicho, se implementó un programa computacional que simuló el método descrito anteriormente, realizando 10000 simulaciones para una baraja de $n = 50$ cartas. Los datos entregados por la simulación fueron visualizados en dos tipos de gráficos distintos que se presentan a continuación.

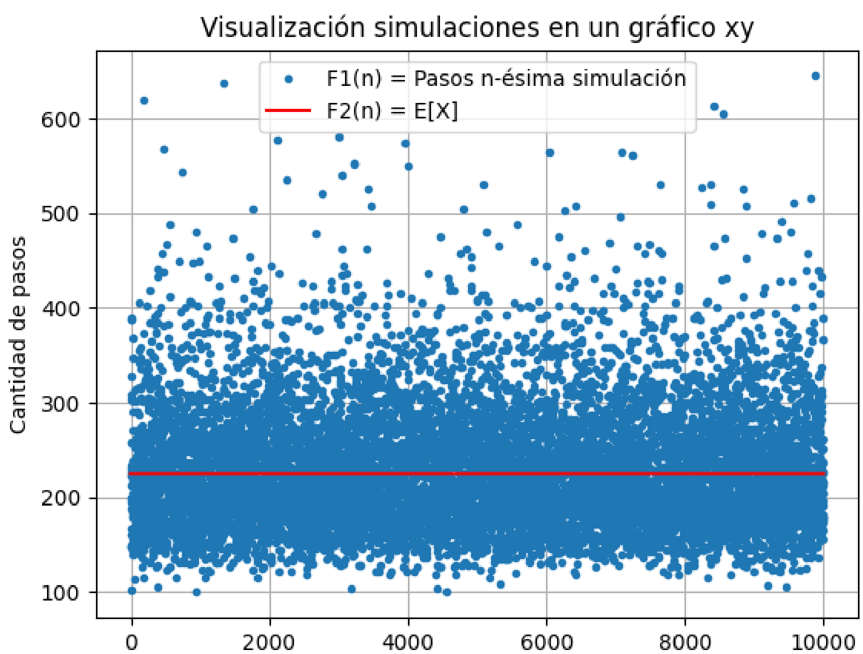

FIG. 1. Primera visualización de los datos de las simulaciones

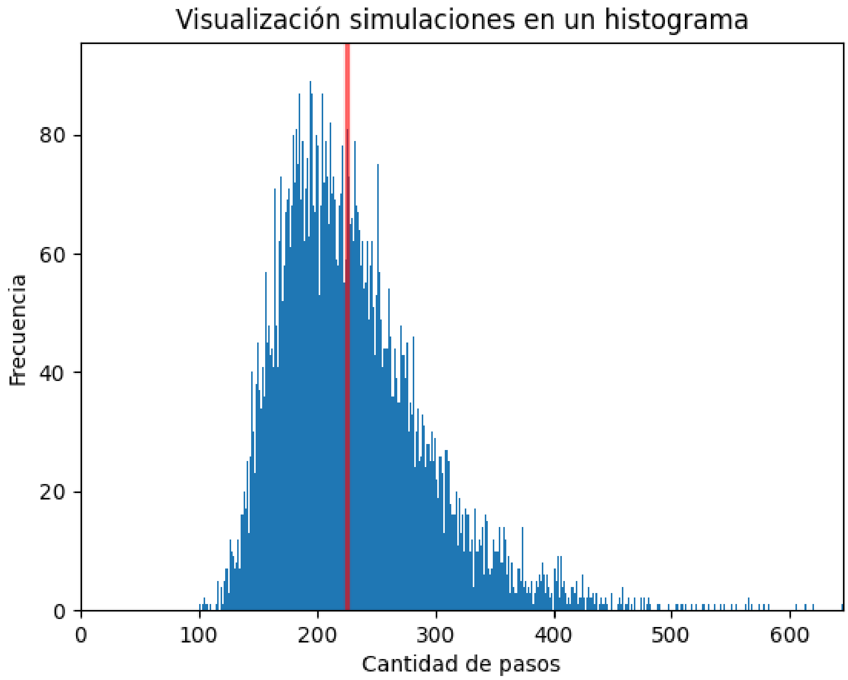

FIG. 2. Segunda visuzalización de los datos de la simulaciones

Lo primero a destacar es la distribución de los datos. A priori, se pensaría que una mayor cantidad de casos se debesiese encontrar en un rango cercano a $E[X]$. La información de los gráficos pareciese confirmar lo dicho, ya que en el rango $[150, 250]$ es donde se observa la mayor cantidad de ocurrencias.

De la misma forma, como X es una variable aleatoria, se esperaría que el número de simulaciones con cantidad de pasos menores a $E[X]$ sea relativamente similar al número de simulaciones con la cantidad de pasos mayores a $E[X]$. Por lo observado los gráficos, a primera vista, pareciese que existe una mayor concentración de datos para las simulaciones con cantidad de pasos menores a $E[X]$.

Respecto a lo anterior, se debe considerar que el rango de simulaciones con cantidad de pasos menores que $E[X]$ es $[50, E[X])$. En cambio, el rango de simulaciones con cantidad de pasos mayores que $E[X]$ es $(E[X], \infty)$. Para el rango de cantidad de pasos menor al valor esperado, se tiene que los datos están más *comprimidos* y, por el otro lado, en el rango de cantidad de pasos menores a el valor esperados, los datos están más *dispersos*. Del mismo modo, en nuestro caso, dado que $n = 50$, se tiene que $E[X] \approx 225$. Como se espera que el promedio de todos los valores sea cercano a $E[X] \approx 225$ y como el rango de datos menores a $E[X]$ es un rango mucho más acotado que el rango para valores mayores a $E[X]$, en general, va existir una mayor cantidad de datos menores a el valor esperado, ya que así se cumple que el promedio de todos los datos es cercano a el valor esperado $E[X]$. En ese sentido, en el mejor caso, el algoritmo puede demorar un tiempo igual a la cantidad de cartas en el mazo. En el peor caso, el algoritmo puede tender a un tiempo de ejecución infinito.

Para visualizar lo anteriormente planteado, con los mismos datos de la simulación anteriror, se graficó un histograma que compara los valores menores a $E[X]$ con los valores mayores a $E[X]$

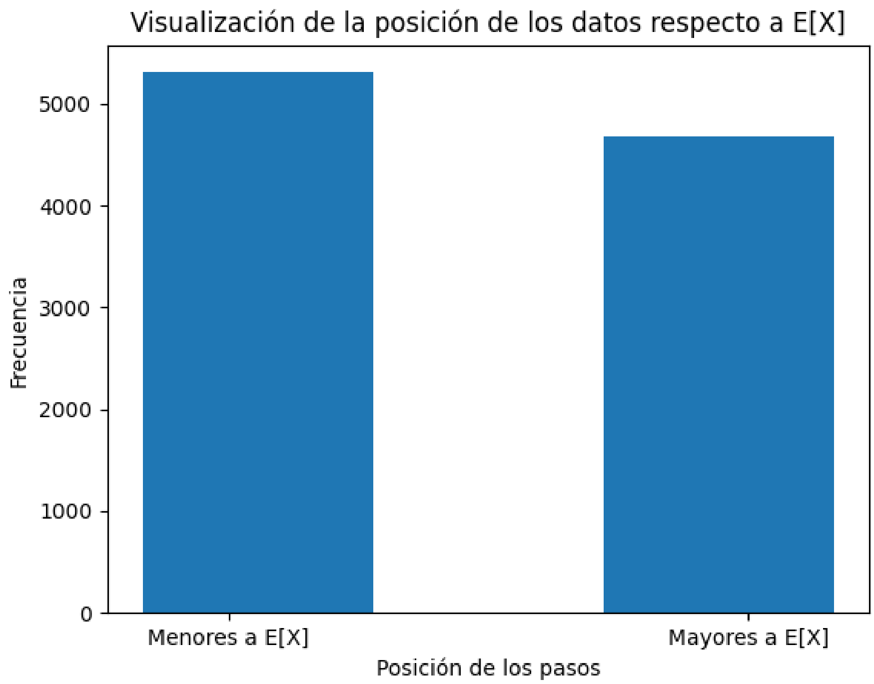

FIG. 3. Comparación posición de los datos

CONCLUSIONES

El estudio del procedimiento de aleatorización de cartas permite desarrollar un análisis más general para un algoritmo de aleatorización que no necesariamente está asociado a la aleatorización de una baraja. De esa forma, el análisis de un caso de estudio permite concluir afirmaciones que abarcan más que solo un ejemplo particular. Del mismo modo, un análisis teórico complementado con datos experimentales permite descubir y proponer proposiciones y aspectos que, en la estudio teórico, no son evidentes a primera vista.