SIMULACIÓN EN EXCEL DE VARIABLES ALEATORIAS DISCRETAS

Giovanni Sanabria Brenes
gsanabriab@yahoo.com
Instituto Tecnológico de Costa Rica – Universidad de Costa Rica, Costa Rica

Tema: V.5 - TIC y Matemática.
Modalidad: Taller
Nivel educativo: Terciario - Universitario
Palabras clave: Variable aleatoria discreta, probabilidad, simulación, Ley de los grandes números

Resumen

Según la ley de los Grandes Números, dada una variable aleatoria X, si se obtiene una cantidad suficientemente grande de valores de X, el diagrama de bastones de las frecuencias relativas se aproxima al histograma de la variable X. Las probabilidades frecuenciales utilizadas en el diagrama definen la distribución de probabilidad empírica u observada de X y la función fX es llamada la distribución teórica de X. El presente taller plantea actividades que, a través de la simulación de algunos experimentos aleatorios utilizando Excel, evidencien: la relación entre las distribuciones empírica y teórica, y la estimación de distribuciones de variables aleatorias.

1. Introducción

Dado un experimento y un evento determinado A, la probabilidad de A es la medida de posibilidad de que ocurra A. Para que la probabilidad sea útil debe existir una correspondencia entre la probabilidad y la realidad, es decir, si el experimento se repite varias veces, la frecuencia relativa observada (probabilidad frecuencial) con que ocurre un evento debe ser cercana a la medida de la posibilidad de que ocurra. Donde la probabilidad frecuencial es $\dfrac{\#\text{ de experimentos donde el evento ocurre}}{\#\text{ de experimentos realizados}}$

La Ley de los Grandes Números establece que: Dado un experimento, sea A un evento, si el experimento se repite un número suficientemente grande de veces, entonces la probabilidad frecuencial de A será muy cercana al valor real de la probabilidad.

Así, si se generan una cantidad suficientemente grande de valores de una variable aleatoria X, la frecuencia relativa de los valores observados de X que se encuentran en un intervalo I se aproxima a la probabilidad teórica de P(X∈I).

¿Cómo generar valores aleatorios de X? Actualmente, el avance que han tenido los ordenadores han permitido el desarrollo de varios software que permite la implementación de simulaciones de forma rápida de una cantidad considera de valores de una variables aleatorias.

El presente taller plantea actividades que, a través de la simulación de algunos experimentos aleatorios utilizando Excel, evidencien: la relación entre las distribuciones empírica y teórica, y la estimación de distribuciones de variables aleatorias. Este propuesta, basada en la simulación en Excel, se complementa con los trabajos Sanabria & Núñez (2010) y Sanabria & Núñez (2011).

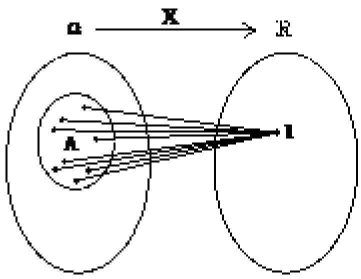
2. Algunos conceptos importantes

Seguidamente se presentan los fundamentos teóricos de variables aleatorias discretas, indicados en Sanabria (2012), Devore (1998) y Walpole et al (1999).

Definición. Variable aleatoria discreta (v.a.d.) Se dice que X es una variable aleatoria discreta si X es variable aleatoria cuyo rango R_X es finito o infinito numerable.

Recuerde que dada una experiencia aleatoria, el espacio muestral Ω es conjunto de posibles resultados, un evento es un subconjunto de Ω, y la probabilidad P asigna un valor entre 0 y 1 a cada evento, este valor mide la posibilidad de que suceda el evento.

Dada la variable aleatoria $X:\Omega\rightarrow\mathbb{R}$, el objetivo es calcular probabilidades para los valores de la variable, específicamente sobre el rango de la variable. Por ejemplo, si X es el número de escudos obtenidos al lanzar una moneda cinco veces, puede ser de interés calcular la probabilidad de que X=1 o de que X=5. Para ello se utiliza el cálculo de probabilidades desarrollado en Ω identificando, por ejemplo, X=1 con un evento $A\subseteq\Omega$:



El evento A es el conjunto de preimágenes que son enviadas a 1, es como observar el valor X=1 desde Ω, donde se tiene la teoría de probabilidad desarrollada. Así, es intuitivo definir la probabilidad de que X=1 por P(X=1)=P(A). Seguidamente se establece la notación que identifica conjuntos de valores en el rango de una v.a.d. X con eventos en Ω.

Notación. Dada una v.a.d. $X:\Omega\to\mathbb{R}$, sea $m\in R_X$. Se tomará en cuenta la siguiente notación de eventos:

1. X=m es el evento $\{e\in\Omega \mid X(e)=m\}$
2. X<m es el evento $\{e\in\Omega \mid X(e)<m\}$
3. X≤m es el evento $\{e\in\Omega \mid X(e)\leq m\}=(X<m)\cup(X=m)$

Similarmente se definen los eventos X>k y X≤k.

Definición. Sea X una v.a.d. con espacio muestral Ω, sea P una probabilidad sobre Ω. La función f_X: $R_X \to\mathbb{R}$ definida por $f_X(m) = P(X=m)$ es llamada función de distribución de probabilidad para X. La función f_X también se denomina ley de probabilidad o función de densidad.

Note que si m≠n, entonces los eventos X=m y X=n son eventos disjuntos, pues como X es una función, no existe una preimagen que se envíe al valor m y también al valor n. Por lo tanto, $P((X=m)\cup(X=n))=P(X=m)+P(X=n)$

Esto facilita mucho el cálculo de probabilidades. Por ejemplo, si X es una v.a.d. con rango $R_X=\{1,2,3,4,5\}$, entonces $P(X\leq 3) = P((X=1)\cup(X=2)\cup(X=3))$
$= P(X=1)+P(X=2)+P(X=3) = f_X(1)+ f_X(2)+ f_X(3)$.

Ejercicio. Se lanza un dado tres veces. Sea X el número de lanzamientos en los que se obtuvo un resultado múltiplo de tres.

1. Determine el rango de X.
2. Halle la función de distribución de probabilidad de X.
3. Realice el histograma de la v.a.d. X.
4. Determine la probabilidad de obtener menos de dos múltiplos de tres.

Ejercicio. (¿Qué sucede si el rango no es finito?) Sea X una v.a.d. con función de probabilidad f_X definida por $f_X(x)=1/2^x$, con x=1,2,3,...

1.Verifique que f_X tiene las condiciones necesarias de una función de probabilidad.

2.Determine P(π<X<8).

3.Halle la probabilidad de que el valor de X sea mayor a 3.

Los ejercicios anteriores evidencian la utilidad que tiene la función de distribución para calcular probabilidades. Sin embargo, en ciertas ocasiones el uso directo de f_X para calcular probabilidades puede ser insuficiente, por ejemplo, si $R_X = \mathbb{N}$ e interesa calcular la probabilidad de que $11 \leq X \leq 100$. Para solucionar dicho problema se utiliza la función de distribución acumulada.

Definición. (Función de distribución acumulada de probabilidad de una v.a.d.) Sea X una v.a.d. y f_X una ley de probabilidad para X, se dice que $F_X:\mathbb{R}\rightarrow\mathbb{R}$ es una función de distribución acumulada de probabilidad, o de masa, para X si

$$F_X(m) = \sum_{\substack{k\in R_X \\ k\leq m}} f_X(k) = \sum_{\substack{k\in R_X \\ k\leq m}} P(X=k) = P(X \leq m)$$

Ejercicio. Sea X el número de lanzamientos en los que se obtuvo par al lanzar un dado tres veces.

1. Determine la función de distribución acumulada para X y grafìquela.
2. Calcule $P(X \geq 2)$.

Teorema. (Características de la función de distribución acumulada) Sea X una v.a.d y F_X la función de distribución acumulada de probabilidad de X. Se cumple que

1. $F_X(x) \geq 0$, para todo $x \in \mathbb{R}$
2. F_X es creciente

3.$P(X>m)=1- F_X(m)$ para $m\in\mathbb{R}$.

4.$P(a<X\leq b)= F_X(b)- F_X(a)$, para $a,b\in\mathbb{R}$ con $a<b$.

5. $F_X(x_n)= F_X(x_n)- F_X(x_{n-1})$, para todo $x_n \in R_X$ que no sea el mínimo de R_X.

Definición. (Esperanza) La esperanza, media o valor esperado de una v.a.d. X es el promedio ponderado de los valores del rango de X según las probabilidades de que X tome cada uno de estos valores. Así, si X es una v.a.d. con función de probabilidad f_X tal que $\sum_{k\in R_X} |k| f_X(k)$ converge, se define la esperanza de X por $E(X) = \mu_X = \sum_{k\in R_X} |k| f_X(k)$.

3. Simulación para variables aleatorias discretas

Según la ley de los Grandes Números, dada una variable aleatoria X, si se obtiene una cantidad suficientemente grande de valores de X, el diagrama de bastones de las frecuencias relativas se aproxima al histograma de la variable X. Las probabilidades frecuenciales utilizadas en el diagrama definen la distribución de probabilidad empírica u observada de X y la función f_X es llamada la distribución teórica de X. Así, cuantas más veces se repita la experiencia aleatoria, la distribución empírica estará más cerca de la distribución teórica.

Ejercicio. Se lanza una moneda tres veces. Considere la v.a.d. X definida como el número de lanzamientos en los que se obtuvo escudo.

1.Determine la función de distribución de probabilidad de X.

2.Utilice simulación en Excel para obtener 200 valores aleatorios de X. Halle una función de probabilidad empírica de X a partir de estos valores.

3.Realice el diagrama de bastones de las frecuencias relativas observadas y el histograma de X.

El ejercicio anterior evidencia la utilidad que tiene la simulación para hallar aproximadamente la función de distribución f_X de una v.a.d. X, especialmente si esta es difícil de encontrar.

Otra utilidad que tiene la simulación es utilizar la función de distribución de una v.a.d. X para determinar la probabilidad de un evento asociado al valor de X. Específicamente, el método de Montecarlo consiste en asignar valores a las variables aleatorias según su distribución acumulada y el valor obtenido en un número aleatorio entre 0 y 1 (random de una computadora).

Ejercicio. En un torneo entre cuatro equipos del grupo A se tienen las siguientes puntuaciones

Equipo	Puntos
San José	24
Alajuela	22
Cartago	21

Heredia	19

De este grupo solo clasifican dos equipos a la siguiente fase y falta un partido para cada equipo: San José vs Cartago, Alajuela vs Heredia.

Si un equipo gana, obtiene tres puntos; si empata, un punto, y se pierde, cero puntos. Considere las variables

X : de puntos que obtiene San José ante Cartago

Y : de puntos que obtiene Alajuela ante Heredia

Mediante un modelo se platearon las siguientes distribuciones de probabilidad para dichas variables

$$f_X(m) = \begin{cases} 0.25 & si \quad m=0 \\ 0.3 & si \quad m=1, \\ 0.45 & si \quad m=3 \end{cases} \qquad f_Y(m) = \begin{cases} 0.35 & si \quad m=0 \\ 0.3 & si \quad m=1 \\ 0.35 & si \quad m=3 \end{cases}$$

Determine, utilizando el método de Montecarlo, la probabilidad de que clasifique Alajuela.

Ejercicio. El Ministerio de Salud está preocupado por la cantidad de niños con obesidad en el país. Un estudio realizado en familias con dos niños determinó que la probabilidad de que ninguno de ellos tenga obesidad es del 25%; de que solo un niño padezca esta enfermedad es del 30%, y de que ambos sean obesos es del 45%. Sea X el número de niños obesos en una familia con dos niños.

1.Determine la esperanza de X.

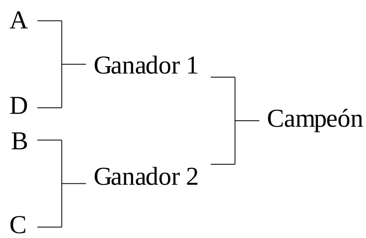
2.Interprete el valor de la esperanza.

3.Determine empíricamente una aproximación a la esperanza de X utilizando simulación en Excel.

Ejercicio. Cuatro equipos clasificaron a la etapa eliminatoria con los puntajes:

Equipo	Puntos
A	37
B	33
C	32
D	28

Estos equipos se enfrentarán de la siguiente forma para definir el campeón del torneo:



Como es una fase eliminatoria, en cada partido se gana o se pierde.

a. Considere la v.a.d. X que toma el valor 0 si A gana y el valor 1 si A pierde. Similarmente, considere las variables Y y Z asociadas a los otros encuentros. Realice un modelo que le permita determinar las funciones de distribución de probabilidad de las variables X,Y,Z utilizando como base los puntajes con que los equipos clasifican.

b. Utilice el método de Montecarlo para determinar aproximadamente la probabilidad de que cada equipo sea campeón.

Ejercicio. Sea X una v.a.d. cuya función de distribución es

$$f_X(m) = \begin{cases} 0.1 & si \quad m = 0 \\ 0.3 & si \quad m = 1 \\ 0.6 & si \quad m = 2 \end{cases}$$

Determine empíricamente una aproximación de la esperanza de X y la esperanza de Y=12X-4 utilizando simulación en Excel.

El juego M Tres consiste en lanzar varias veces un dado hasta obtener un múltiplo de tres. Sea X el número de lanzamientos realizados antes de obtenerlo.

1.Determine f_X, F_X y E(X).

2. Determine empíricamente una aproximación a la esperanza de X utilizando simulación en Excel.

4. Conclusión

A través de las simulaciones hechas en Excel y bajo ciertas consideraciones teóricas, se logra generar una cantidad considerable de valores de una variable aleatoria discreta X que permiten estimar probabilidades sobre X y el valor esperado X.

La utilización de este trabajo en los cursos de probabilidad permitirá introducir a los estudiantes en la probabilidad numérica como una opción para el cálculo de probabilidad, además de la probabilidad teórica o algebraica.

Además, el uso de la simulación permite mantener ligado el concepto de probabilidad con su utilidad práctica gracias a la Ley de los Grandes Números.

Referencias bibliográficas

Devore, J. (1998). *Probabilidad y Estádistica para Ingeniería y Ciencias.* México: International Thomson Editores, 4a ed.

Sanabria, G (2012). *Comprendiendo las Probabilidades.* Costa Rica: Editorial Tecnológica de Costa Rica.

Sanabria, G. & Núñez, F. (2010). *Una propuesta para introducir el estudio de las probabilidades: Probabilidad Frecuencial*. En Facultad de Ciencias Naturales, Universida Estatal a Distancia. Memorias III Encuentro de Enseñanza de la Matemática UNED, realizado en el INBio Parque, Heredia, Costa Rica, 3 y 4 de setiembre 2010. InBio Parque, Heredia, Costa Rica.

Sanabria, G. & Núñez, F. (2011). *Introducción a la probabilidad utilizando la simulación en Excel*. Memorias del 1er Encuentro Internacional de Enseñanza de la Probabilidad y la Estadística (EIEPE), del 12 al 15 de julio de 2011. México: Facultad de Ciencias Físico Matemáticas de la Benemérita Universidad Autónoma de Puebla.

Walpole, R, Myers, R, Myers, S. (1999). *Probabilidad y estadística para ingenieros.* USA: Prentice-Hall Hispanoamericana. S.A, Sexta Ed.