La simulación: un recurso didáctico para facilitar el aprendizaje de la Probabilidad y la Estadística.

Autores: Raúl D. Katz, Mabel A. Medina.
Departamento de Matemática. Escuela de Formación Básica.

Resumen

El presente trabajo hace referencia a una actividad realizada con estudiantes que cursaban la asignatura Probabilidad y Estadística en la especialidad Civil. La misma abarcó la resolución de problemas utilizando simulaciones asistidas por el uso de un software, mediante la aplicación del método de Montecarlo.

La simulación no constituye solamente un medio propicio para obtener respuestas aproximadas en una variedad de problemas. También es un recurso didáctico para facilitar la comprensión de algunas propiedades que resultan complejas por el nivel de abstracción que requieren.

La actividad consta de seis problemas, tres de los cuales están guiados y los restantes quedan propuestos para el alumno. La simulación se realiza con el software libre Scilab.
Este tipo de actividad rompe con la rutina de la clase tradicional e introduce experiencias de aprendizaje activo que contribuyen a elevar la calidad del proceso de enseñanza-aprendizaje en la asignatura Probabilidad y Estadística. Además se observa un alto grado de motivación en los estudiantes cuando obtienen un resultado o respuesta a un problema a través de un proceso de ensayo, o cuando mediante las simulaciones establecen un puente con propiedades que involucran conceptos abstractos.

Cabe aclarar que este recurso didáctico debe apuntar a planteamientos metodológicos apropiados: no se trata de cambiar prácticas comprobatorias por simulaciones comprobatorias. Se deben usar como un recurso más en propuestas multimedia. Para ello es necesaria una meticulosa planificación de las propuestas.

Introducción

El presente trabajo hace referencia a una actividad realizada con estudiantes que cursaban la asignatura Probabilidad y Estadística en la especialidad Civil. La misma abarcó la resolución de problemas utilizando simulaciones asistidas por el uso de un software, mediante la aplicación del método de Montecarlo.

El interés de la simulación en la enseñanza de la Probabilidad y la Estadística es destacada entre otros, en los trabajos de Biehler (1991) y Estepa (1994). Como describe este último autor existen experiencias (o fenómenos) que, por sus características espacio - temporales o de otra índole son difíciles de observar. En estos casos el recurso de la simulación permite condensar el experimento en el tiempo y en el espacio y operar con el experimento simulado para obtener resultados válidos para el experimento original. Asimismo el uso de la simulación pone al alcance de los alumno un nuevo medio para la exploración de conceptos y propiedades, que de otro modo serían mucho más abstractos, contribuyendo de este modo a atenuar el problema de la falta de experiencia aleatoria y mejorar la intuición probabilística que, en general, no se desarrolla espontáneamente (Fischbein, 1975).

Objetivos de la propuesta

En ocasiones la dificultad para resolver un problema se asocia con alguna complejidad matemática que lo rodea. Sin embargo, la comprensión del problema y el recurso matemático disponible para emprender su solución son dos dimensiones que no tienen porqué coincidir. Cuando éste es el caso, la simulación puede ser un medio propicio para obtener respuestas aproximadas.

Atendiendo al continuo avance tecnológico se considera indicado mostrarles a los estudiantes ejemplos sencillos que ilustren la aplicabilidad de esta técnica y ofrecer así oportunidades de prácticas innovadoras que apoyen su formación para la futura actividad laboral.

Por tal razón y en función de las consideraciones realizadas anteriormente se formularon los siguientes objetivos para las actividades propuestas:

a) mostrar una variedad de problemas abordables por simulación.
b) utilizar los recursos de la simulación para facilitar la comprensión de algunas propiedades que resultan complejas por el nivel de abstracción que requieren.

Se parte del supuesto de que a través de estas actividades se favorece el proceso de aprendizaje y el desarrollo de habilidades cognitivas.

Software libre utilizado: Scilab

El Software Libre Scilab fue desarrollado en el INRIA (Institut Nacional de Recherche en Informatique et Automatique, Francia) con la colaboración de la escuela de ingenieros ENPC (Ecole Nationale de Ponts et Chaussées) y es actualmente usado en universidades y liceos franceses y de otros países. El sitio oficial es www.scilab.org.

Sus principales características son:

- Software numérico de cálculo científico.
- Interactivo.
- Programable.
- De libre uso, con la condición de siempre hacer referencia a sus autores.
- Disponible para diferentes plataformas: Windows, Linux, Sun, Alpha, etc.
- De fácil accesibilidad e instalación en versión ejecutable o fuente.

Los problemas propuestos

Problema 1

Un simple cálculo de integración nos permite determinar que el área encerrada por la gráfica de la función $y = x^2$, para x variando en $[0,1]$, es igual a $\frac{1}{3}$.
Estime el valor del área, utilizando los recursos de la simulación.

Problema 2

a) Genere por simulación 100000 valores de una variable aleatoria X uniformemente distribuida en el intervalo [0,1] y represente la distribución de los mismos a través de histogramas, considerando diferentes cantidades de intervalos de clase.
b) Divida los 100000 valores en 1000 muestras de tamaño 100 y calcule la media aritmética y variancia en cada muestra de tamaño 100. Construya luego un histograma que muestre la distribución de las medias aritméticas y variancias muestrales.
c) Repita el procedimiento realizado en b) pero considerando ahora 100 muestras de tamaño 1000.
d) Realice un breve informe sobre sus observaciones, en relación a los histogramas construidos en b) y c).

Problema 3

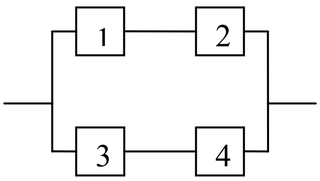
a) Genere por simulación 10000 valores de una variable aleatoria X uniformemente distribuida en el [0,1] y otros 10000 valores de una variable aleatoria Y uniformemente distribuida en el [0,1]. Considere luego Z = X+Y y construya un histograma con los 10000 valores de Z.
b) Reflexione sobre la distribución de los valores de Z y explique porqué son más frecuentes los valores en la proximidad del uno, que los valores cercanos a cero y dos.
c) Calcule la media aritmética y variancia de los 10000 valores correspondientes a X, Y y Z.
d) Considere ahora W = 2X y construya un histograma con los 10000 valores de W. ¿Cuáles son sus observaciones sobre la distribución de dichos valores?
e) Calcule la media aritmética y variancia de los 10000 valores correspondientes a W.
f) Establezca diferencias entre las distribuciones y los valores característicos calculados para los 10000 valores de Z y W.

Problema 4

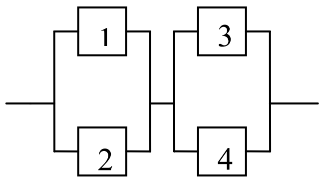
Una persona sale de su casa a las 8:00 hs. para abordar un colectivo que sale de la terminal a las 8:30 hs.
El tiempo que la persona demora en conseguir un taxi que lo traslade a la terminal es una variable aleatoria X con distribución exponencial y media 5 minutos.
El tiempo que demora el taxi en trasladarlo es una variable aleatoria Y con distribución uniforme en el intervalo [15, 20].
Estime, utilizando los recursos de la simulación, la probabilidad de que la persona no alcance a tomar el colectivo.

Problema 5

Considere las siguientes dos configuraciones de un sistema.



Configuración 1



Configuración 2

Las componentes de los sistemas provienen de un proceso del cual se conoce que la distribución de los tiempos hasta la falla (duración de las componentes) sigue una ley normal con media 500 horas y desviación estándar igual a 25 horas.
Para cada configuración utilice los recursos de la simulación para:

a) estudiar la distribución de los tiempos hasta la falla (del sistema).
b) estimar la media, la mediana y la variancia de los tiempos hasta la falla.
c) estimar la probabilidad de que el sistema funcione al cabo de las 500 horas.
d) estimar la probabilidad de que en la configuración 1 el subsistema formado por la rama superior falle antes que el subsistema formado por la rama inferior.
e) estimar la probabilidad de que en la configuración 2 el subsistema de la izquierda falle antes que el subsistema de la derecha.
f) Repita los gráficos y cálculos, considerando otras distribuciones para los tiempos hasta la falla de las componentes.

Problema 6

El número de camiones que pasan por un puesto de peaje en cierto período de una hora sigue una ley de Poisson con media igual a10.
El número de colectivos y autos que pasan por el mismo puesto, en el mismo período, también sigue una ley de Poisson con media igual a 5 y 50 respectivamente.
Cada camión paga un peaje de $10, mientras que los colectivos y autos pagan un peaje de $8 y $5 respectivamente.

a) Calcule la media y la variancia de la recaudación en dicho período (suponga que la cantidad de vehículos de cada tipo que pasan por dicho peaje son variables aleatorias independientes).
b) Estime dicha media y variancia a partir de los datos de una muestra simulada de tamaño 100.
c) A partir de la muestra simulada estime la probabilidad de que la recaudación en dicho período supere los $400.
d) Repita las estimaciones realizadas en b) y c) para otros tamaños de muestra.

Consideraciones sobre los problemas propuestas

En el problema 1 los estudiantes generan por simulación pares ordenados (x_i,y_i) de números reales, donde, tanto los x_i como los y_i son respectivamente valores de variables aleatorias (independientes) X e Y, que se distribuyen uniformemente en el intervalo [0,1]. De este modo el vector (X,Y) se distribuye uniformemente en el cuadrado $[0,1]\times[0,1]$. La probabilidad de que un par ordenado (x_i,y_i) pertenezca a la región $R=\{(x,y):\ 0 \le x \le 1,\ y \le x^2\}$ es igual al área de dicha región dividido por el área del cuadrado. Siendo el área del cuadrado igual a uno, dicha probabilidad es igual al área de la región. En consecuencia estimar el valor del área equivale a estimar el valor de una probabilidad. Por tal razón, los estudiantes calculan la proporción o frecuencia relativa de pares (x_i,y_i) que tienen $y_i \leq x_i^2$; es decir la cantidad de pares ordenados que se encuentran en la región delimitada por la gráfica de la función sobre el intervalo [0,1], dividido por el total de pares ordenados generados.

La ley de los grandes números les brinda los medios para determinar la cantidad de observaciones a realizar, para que la frecuencia relativa de un suceso difiera de la probabilidad del suceso en menos de un valor fijado, con una probabilidad también fijada. Es decir, por aplicación de la ley de los grandes números determinan la cantidad mínima de pares ordenados a simular, para estimar el área de la región con un error menor a un valor fijado y con una confiabilidad también fijada.
Por ejemplo, si se fija el error de estimación en menos de 0.01 y la confiabilidad de la misma en por lo menos 0.95, se deben simular 44445 o más pares ordenados.
Este valor resulta de aplicar: $P[|f_A - p| < \varepsilon] \geq 1-\alpha$ cuando $n \geq \frac{p(1-p)}{\varepsilon^2\alpha}$ [1], donde p es el valor conocido de la probabilidad e igual a $\frac{1}{3}$, ε=0.01 y α=0.05.

El resultado aplicado deriva de la aplicación de la desigualdad de Tchebyshev, por cuanto se desconoce la distribución de la variable aleatoria f_A (frecuencia relativa de ocurrencia del suceso A). Ésta es la razón por la cual se obtiene un valor de n tan grande.
Si se considera que f_A tiende a distribuirse normalmente para valores de n lo suficientemente grande (Teorema del límite central) se puede obtener un valor más pequeño de n planteando $z\sqrt{\frac{p(1-p)}{n}} < \varepsilon$. Reemplazando z por 1.96 (valor que se obtiene de la tabla de la distribución normal cuando se trabaja con una confiabilidad igual a0.95), p por $\frac{1}{3}$ y ε por 0.01 se obtiene que n tiene que ser mayor a 8536.

Los estudiantes estiman el área sobre la base de 8537 pares ordenados simulados y comparar el valor de la estimación con el verdadero valor (igual a $\frac{1}{3}$).
Luego se les propone:

a) Construir un histograma con los valores obtenidos de las repetidas estimaciones del área.
b) Repetir el procedimiento hasta obtener una estimación del área con un error que supere el valor 0.01 y registrar las veces que repitió el procedimiento.
c) Comparar y comentar los valores que obtienen con los resultados de sus compañeros.
d) Repetir la actividad para estimar el área encerrada por la función $f(x) = e^{x^2}$ para $x\,\varepsilon[0,1]$.

[1] La demostración puede consultarse en: Paul L. Meyer (1973). Probabilidad y Aplicaciones Estadísticas. Fondo Educativo Interamericano, S.A. México

En el contexto de la situación planteada la simulación se torna particularmente útil para el aprendizaje de la estimación de un parámetro poblacional. Los estudiantes verifican “empíricamente” que la estimación de un parámetro en la población (área encerrada por la curva en el ejemplo) varía de una muestra a otra y que por lo tanto el estimador es una variable aleatoria con una distribución que se ajusta a un patrón. La construcción de un histograma con los valores de las repetidas estimaciones les permite visualizar la tendencia de ese patrón.

De este modo la actividad que se realiza se constituye en un apropiado complemento de una enseñanza tradicional donde la distribución del estadístico se obtiene a partir de deducciones algebraicas que demandan un elevado nivel de abstracción.

El objetivo de la propuesta 2 pretende ahondar en la distribución de las medias muestrales de una distribución uniforme, observando no solamente la incidencia del tamaño de la muestra en la precisión de la estimación de un parámetro, sino también captar a partir de la construcción de un histograma, basado en una cantidad considerable de medias muestrales, la distribución teórica de la media muestral.

En la propuesta 3 se busca provocar a partir del recurso de la simulación una mayor reflexión sobre el significado de la suma de dos variables aleatorias y establecer la diferencia entre variables aleatorias iguales y variables aleatorias idénticamente distribuidas, conceptos que suelen ser confundidos por la escasa atención que se le presta al lenguaje.

La propuesta 4 requiere encontrar la distribución de la suma de dos variables aleatorias cuando se conoce la distribución de cada una de ellas. Esto demanda plantear una integral de convolución. La evaluación de dichas integrales suele ser complicada y debe hacerse con mucho cuidado. Si bien el recurso analítico se encuentra disponible, las limitaciones del tiempo no permiten profundizar en estas cuestiones teóricas que por otra parte demandan la realización de cálculos complejos. En estos casos el recurso de la simulación proporciona un medio simple y rápido para obtener respuestas aproximadas a los problemas planteados. Una situación similar se presenta en las propuestas 5 y 6.

Resumiendo, en las actividades de simulación propuestas en 1, 2 y 3 se proporciona un medio alternativo para que los estudiantes “sometan a prueba” propiedades tratadas en el salón de clases. Algunas de ellas, por la complejidad matemática de sus pruebas, sólo se enuncian y se aplican luego en la resolución de problemas.
En cambio las propuestas 4, 5 y 6 utilizan la simulación para el tratamiento y resolución de problemas cuyo abordaje no resulta posible cuando se carecen de ciertos fundamentos teóricos, que se pueden alcanzar cuando se disponen de otros tiempos para la enseñanza y el aprendizaje.

Nuestras reflexiones

Consideramos que propuestas de este tipo rompen con la rutina de la clase tradicional e introducen experiencias de aprendizaje activo que contribuyen a elevar la calidad del proceso de enseñanza-aprendizaje en la asignatura Probabilidad y Estadística.

Cabe señalar el alto grado de motivación que muestran los estudiantes cuando obtienen un resultado o respuesta a un problema a través de un proceso de ensayo, o cuando mediante las simulaciones establecen un puente con propiedades que involucran conceptos abstractos.

Las simulaciones deben apuntar a planteamientos metodológicos apropiados: no se trata de cambiar prácticas comprobatorias por simulaciones comprobatorias. Se deben usar como un recurso más en propuestas multimedia. Para ello, es necesaria una meticulosa planificación de las propuestas.

Bibliografía

- Estepa, A. (1994). *Concepciones iniciales sobre la asociación estadística y su evolución como consecuencia de una enseñanza basada en el uso de ordenadores.* Tesis Doctoral. Universidad de Granada.
- Fischbein, E. (1975). *The intuitive sources of probability thinking in children.* Dordrecht: D. Reidel.
- Navidi W. (2006). *Estadística para ingenieros y científicos*. McGrawHill. Mexico 2006.
- Sóbol I. M. (1976). *Método de Montecarlo*. Mir. Moscú.