Simulación en R aplicada a la Ley de los Grandes Números y Teorema Central del Límite

Realizar el ejercicio en R en grupos de 2 estudiantes (obligatoriamente) del mismo turno. Utilizar semillas para poder replicar los resultados. Se editan las respuestas a modo de informe y se entrega un archivo en formato pdf. No hay que entregar el script. La entrega es obligatoria y no tendrá una calificación cuantitativa pero sin embargoe, deberán rehacerlo en caso de errores según el criterio de los docentes. La fecha de entrega es el lunes 6 de julio a las 22hs. La entrega es individual subiendo el archivo al moodle. Recordar colocar en el pdf los nombres y números de libreta de los estudiantes que trabajaron juntos.

Estudiante 1 Apellido y nombre: N$^{\text{O}}$ de libreta:
Estudiante 2 Apellido y nombre: N$^{\text{O}}$ de libreta:

1. En este ejercicio estudiaremos la distribución del promedio de variables independientes e idénticamente distribuidas. A través de los histogramas correspondientes analizaremos el comportamiento de estas distribuciones a medida que promediamos un número creciente de variables aleatorias. Es decir, trataremos de validar empíricamente los resultados de la Ley de los Grandes Números y el Teorema Central del Límite.

   a) Comencemos por tomar un primer conjunto de datos de variables aleatorias $X_1, \ldots, X_{1000}$ independientes con distribución $U(0,1)$. Le pedimos a R que nos genere una muestra de ellas y luego hacemos un histograma. ¿A qué densidad se parece el histograma obtenido?

   b) Considerar dos variables aleatorias X_1 y X_2 independientes con distribución $U(0,1)$ y el promedio de ambas, es decir,

   $$\overline{X} = \frac{X_1 + X_2}{2}\,.$$

   Generando una muestra de dos variables aleatorias con distribución $U(0,1)$ computar la variable promedio. Replicar 1000 veces y a partir de los valores replicados realizar un histograma. ¿Qué características tiene este histograma?

   c) Aumentemos a cinco las variables promediadas. Considerar ahora 5 variables aleatorias uniformes independientes, es decir $X_1, X_2, \ldots, X_5$ i.i.d. con $X_i \sim U(0,1)$ y definir

   $$\overline{X} = \frac{1}{5}\sum_{i=1}^{5} X_i\,.$$

   Generando muestras de cinco variables aleatorias con distribución $U(0,1)$ computar la variable promedio. Repetir 1000 veces y realizar un histograma para los valores obtenidos. Comparar con el histograma anterior. ¿Qué se observa?

d) Aumentemos aún más la cantidad de variables promediadas. Generando muestras de 30 variables aleatorias con distribución $U(0,1)$ repetir el ítem anterior. ¿Qué se observa?

e) Ídem anterior generando muestras de 500 variables aleatorias. ¿Qué pasa si se aumenta el tamaño de la muestra? Observar que para poder comparar los histogramas de los distintos conjuntos de datos será necesario tenerlos dibujados en la misma escala tanto para el eje horizontal como para el vertical. Por eso, en general es más cómodo hacer boxplots para comparar distintos conjuntos de datos.

f) Finalmente hacerlo también para 1200, y hacer un boxplot de los 6 conjuntos de datos en el mismo gráfico. En este gráfico se verá que a medida que aumenta el n los valores de los promedios tienden a concentrarse, ¿alrededor de qué valor? Calcule media y varianza muestral para cada conjunto de datos. ¿Puede dar los valores teóricos a los que deberían parecerse? Realice un qqplot para cada uno de los 6 conjuntos de datos. ¿Son esperables los resultados?

g) El teorema central del límite nos dice que cuando hacemos la siguiente transformación con los promedios,

$$\frac{\overline{X}_n - E\left(X_1\right)}{\sqrt{\frac{Var(X_1)}{n}}},$$

la distribución de estas variables aleatorias se aproxima a la de la normal estándar, cuando n es suficientemente grande. Para comprobarlo empíricamente, hagamos esta transformación en los 6 conjuntos de datos y luego comparemos los datos transformados mediante histogramas y boxplots.

h) Repetir los ítems anteriores generando ahora variables aleatorias con distribución de Cauchy $\mathcal{C}(0,1)$. Comparar los resultados obtenidos. La densidad de una Cauchy es

$$f_X\left(x\right) = \frac{1}{\pi\left(1+x^2\right)},$$

que es una densidad simétrica alrededor del cero, con colas que acumulan más probabilidad que la normal estándar, y que no tiene esperanza ni varianza finitas.