Tarea 2

Conceptos y aplicaciones de convergencia de variables aleatorias
Fecha de entrega: viernes 17 de noviembre del 2023

Instrucciones: Responde las siguientes preguntas y justifica con los procedimientos correspondientes.

Importante: La tarea deberá cumplir con los siguientes requisitos para tener derecho a ser calificado: 1) nombres de los integrantes del equipo (máximo 3 alumnos) en la parte superior derecha de la primera hoja sin ocupar mucho espacio, 2) tener excelente presentación, 3) sin desperdiciar espacios, 4) gráficas bien diseñadas, cuidando escalas y espacios, 5) no incluyas código ni partes de bases de datos y 6) cuando corresponda justificando con prosa de forma concisa, bien pensada y escrita sus respuestas a cada pregunta.

1. Explica con tus propias palabras lo que entiendes por muestra aleatoria (m.a.).

2. Considera m.a. $X_1, X_2, \ldots, X_n$ de una v.a. $X \sim \text{Poisson}(\lambda)$.

   2.1. Formula la Ley Débil de los Grandes Números (LDGN) y el Teorema Central de Límite (TCL) para este caso.

   2.2. Para entender de mejor manera el punto 2.1. vamos a realizar una simulación. Genera $m = 5,000$ muestras aleatorias $X_1, X_2, \ldots, X_n$ y muestra la convergencia de Z_n a una $N(0,1)$. Asume $\lambda = 1/2$ y utiliza $n = 5, 10, 100$ y $1,000$. Grafica los histogramas correspondientes y ajusta en cada caso la distribución de la normal estándar. Describe lo que observas.

   2.3. Ahora grafica los histogramas para $\bar{X}_n$ con $n = 5, 10, 100$ y $1,000$ utilizando las $m = 5,000$ iteraciones. Describe lo que observas.

   2.4. Repite las simulaciones del punto 2.3., pero ahora generando m.a. de una v.a. $X \sim$ Cauchy. Describe lo que observas e intenta explicar lo que sucede.

3. Considera m.a. $X_1, X_2, \ldots, X_n$ de una v.a. $X \sim \text{Bernoulli}(p)$.

   3.1. Formula la Ley Débil de los Grandes Números (LDGN) y el Teorema Central de Límite (TCL) para este caso.

   3.2. Para entender de mejor manera el punto 3.1. vamos a realizar una simulación. Genera $m = 5,000$ muestras aleatorias $X_1, X_2, \ldots, X_n$ y muestra la convergencia de Z_n a una $N(0,1)$. Asume $p = 1/3$ y utiliza $n = 5, 10, 20$ y 40. Grafica los histogramas correspondientes y ajusta en cada caso la distribución de la normal estándar. Describe lo que observas.

   3.3. Repite el ejercicio asumiendo $p = 0.01$ y $p = 0.99$. ¿Puedes encontrar alguna n "suficientemente grande" para que la aproximación vía el TCL funcione adecuadamente en estos casos?

   3.4. ¿Qué puedes concluir?

4. Vamos a aplicar el TCL a los datos del Censo de Población y Vivienda 2021, `http://sigma.iimas.unam.mx/carloserwin/`. Primero, construye una base de datos del Censo a nivel municipio, en México hay 2,469 municipios. Haz el histograma de la población de 15 años y más

analfabeta por municipio, esta será la variable que analizarás (puedes analizar la variable que quieras). Selecciona $m = 5,000$ muestras aleatorias simples con reemplazo (MASCR) $X_1, X_2, \ldots, X_n$, considera los tamaños de muestra $n = 5, 10, 20$ y 50 y genera los histogramas en cada caso. También incluye la distribución normal que mejor ajusta cada histograma. Describe lo que observas. ¿Hay alguna diferencia si en lugar de seleccionar MASSCR seleccionas muestras aleatorias simples sin reemplazo (MASSR)? La función `sample` de R te será útil, consulta la opción `replace = TRUE` o `FALSE`.