Probabilidad y Estadística

Trabajo Práctico

Ing. Gabriel Pena

Importante: el trabajo puede resolverse utilizando cualquier plataforma o lenguaje de programación que el estudiante considere adecuado. Desde la cátedra recomendamos Python, R, MATLAB/Octave y Wolfram Mathematica. No obstante, y a menos que se especifique lo contrario, el trabajo debe completarse utilizando solo las funciones básicas del lenguaje y sin recurrir a librerías de Estadística.

Parte 1: Simulación

En esta primera parte, construiremos varios generadores de números aleatorios que usaremos para obtener muestras con distribución conocida sobre las que vamos a trabajar posteriormente.

1. Utilizando únicamente la función *random* de su lenguaje (la función que genera un número aleatorio uniforme entre 0 y 1), implemente una función que genere un número distribuido Bernoulli con probabilidad p.
2. Utilizando la función del punto anterior, implemente otra que genere un número binomial con los parámetros n, p.
3. Utilizando el procedimiento descrito en el capítulo 6 del Dekking (método de la función inversa o de Monte Carlo), implementar una función que permita generar un número aleatorio con distribución $Exp(\lambda)$.
4. Investigar como generar números aleatorios con distribución normal. Implementarlo.

Parte 2: Estadística descriptiva

Ahora vamos a aplicar las técnicas vistas en la materia al estudio de algunas muestras de datos.

1. Generar tres muestras de números aleatorios $Exp(0{,}5)$ de tamaño $n = 10$, $n = 30$ y $n = 200$. Para cada una, computar la media y varianza muestral. ¿Qué observa?
2. Para las tres muestras anteriores, graficar los histogramas de frecuencias relativas con anchos de banda 0,4, 0,2 y 0,1; es decir, un total de 9 histogramas. ¿Qué conclusiones puede obtener?
3. Generar una muestra de números $Bin(10, 0{,}3)$ de tamaño $n = 50$. Construir la función de distribución empírica de dicha muestra.
4. A partir de la función de distribución empírica del punto anterior, generar una nueva muestra de números aleatorios utilizando el método de simulación de la primera parte. Computar la media y varianza muestral y graficar el histograma.
5. Repetir el experimento de los dos puntos anteriores con dos muestras aleatorias más generadas con los mismos parámetros. ¿Qué conclusión saca?

Parte 3: Convergencia

El propósito de esta sección es ver en forma práctica los resultados de los teoremas de convergencia.

1. Generar cuatro muestras de números aleatorios de tamaño 100, todas con distribución binomial con $p = 0{,}40$ y $n = 10$, $n = 20$, $n = 50$ y $n = 100$ respectivamente. Graficar sus histogramas. ¿Qué observa?
2. Elija la muestra de tamaño 200 y calcule la media y desviación estándar muestral. Luego, normalice cada dato de la muestra y grafique el histograma de la muestra normalizada. Justifique lo que observa.
3. Para cada una de las muestras anteriores, calcule la media muestral. Justifique lo que observa.

Parte 4: Estadística inferencial

Para terminar, vamos a hacer inferencia con las muestras que generamos y obtener así información sobre sus distribuciones.

1. Generar dos muestras $N(100,5)$, una de tamaño $n = 10$ y otra de tamaño $n = 30$. Obtener estimaciones puntuales de su media y varianza.
2. Suponga que ya conoce el dato de que la distribución tiene varianza 5. Obtener intervalos de confianza del 95 % y 98 % para la media de ambas muestras.
3. Repita el punto anterior pero usando la varianza estimada s^2, para la muestra de tamaño adecuado.
4. Probar a nivel 0,99 la hipótesis de que la varianza sea $\sigma^2 > 5$. Calcular la probabilidad de cometer error tipo II para la hipótesis alternativa $\sigma^2 = 6$.
5. Agrupando los datos en subgrupos de longitud 0,5, probar a nivel 0,99 la hipótesis de que la muestra proviene de una distribución normal.