



Asignatura: Estadística y probabilidad 1

Facultad: Ingeniería y Ciencias Básicas

Núcleo: Matemáticas y Estadística

Proyecto 2. Aplicación de Distribuciones de Probabilidad

1. Objetivo

En este proyecto, el estudiante explorará y aplicará varios conceptos relacionados con las distribuciones de probabilidad, incluyendo la distribución normal, exponencial, gamma, Poisson y binomial, en una serie de situaciones realistas. El objetivo principal es analizar datos y tomar decisiones basadas en la teoría de probabilidad en diferentes contextos.

2. Problemas

1. Considere las siguientes variables aleatorias

a) $X \sim b(x; n, p)$. Seleccione valores para n (entre) y p .

b) $Y \sim exp(x; \lambda)$. Seleccione un valor para λ .

Para cada variable aleatoria genere una muestra aleatoria de tamaño 150 y realizar el siguiente procedimiento:

a) Construir, según el caso, un histograma de masa o densidad y, en un mismo gráfico, compararlo con la función de probabilidad teórica.

b) Calcular el promedio, los cuartiles y la desviación estándar. Comparar los valores obtenidos con los valores teóricos.

2. Suponga un sistema de almacenamiento de datos en la nube utilizado por una gran empresa de comercio electrónico. Este sistema recibe y procesa continuamente transacciones de ventas de múltiples puntos de venta en línea. El tiempo entre llegadas de transacciones tiene promedio de 20 transacciones por minuto. La eficiencia del sistema y la velocidad de las consultas dependen en gran medida de la gestión de estos tiempos entre llegadas.

a) ¿Cuál es la probabilidad de que el tiempo entre dos transacciones consecutivas sea menor o igual a 2 segundos?

b) ¿Cuál es la probabilidad de que el tiempo entre dos transacciones consecutivas sea mayor a 10 segundos?

- c) Si el sistema necesita realizar un respaldo rápido que toma exactamente 30 segundos, ¿cuál es la probabilidad de que no llegue ninguna transacción durante este período de respaldo?
 - d) ¿Cuál es el tiempo esperado entre dos transacciones consecutivas?
 - e) Si el sistema está diseñado para manejar ráfagas de datos, ¿cuál es la probabilidad de que ocurran al menos 5 transacciones en un intervalo de 10 segundos?
3. En un centro de datos de alta eficiencia, se monitorean constantemente varios parámetros críticos para asegurar el rendimiento óptimo de los sistemas. Uno de los parámetros más importantes es la temperatura de los servidores. Después de un extenso análisis, se ha determinado que la temperatura de los servidores sigue una distribución normal con una media de 35°C y una desviación estándar de $2,5^{\circ}\text{C}$.
- a) Realice las gráficas correspondientes a la función de densidad de probabilidad (pdf) y a la función de distribución acumulativa (cdf) de la temperatura de los servidores.
 - b) El sistema de enfriamiento se activa automáticamente cuando la temperatura supera los 40°C . ¿Cuál es la probabilidad de que el sistema de enfriamiento se active para un servidor elegido al azar?
 - c) Los servidores funcionan de manera óptima cuando su temperatura está entre 30°C y 38°C . ¿Qué porcentaje de tiempo se espera que los servidores operen en este rango de temperatura óptima?
 - d) El equipo de mantenimiento quiere establecer un umbral de alerta temprana, de modo que tengan tiempo de intervenir antes de que se active el sistema de enfriamiento automático. ¿A qué temperatura deberían establecer esta alerta si quieren que se active para el 10 % superior de los eventos de temperatura?
 - e) Se está implementando un nuevo sistema de monitoreo que toma el promedio de temperatura de 5 servidores seleccionados aleatoriamente cada minuto. ¿Cuál es la probabilidad de que este promedio exceda los 37°C en una lectura dada?
 - f) El centro de datos tiene 1000 servidores. En un momento dado, ¿cuál es la probabilidad de que más de 25 servidores tengan una temperatura superior a 41°C simultáneamente?
 - g) Se está diseñando un nuevo sistema de enfriamiento. Los ingenieros quieren asegurarse de que pueda manejar el 99.9 % de todos los eventos de temperatura. ¿Para qué temperatura máxima deben diseñar este sistema?

- h)* El equipo de eficiencia energética quiere implementar un sistema de enfriamiento variable que ajuste su potencia basándose en la temperatura actual. Han dividido el rango de temperaturas en cuartiles. Calcule los puntos de corte para estos cuartiles y explique cómo esto podría usarse para crear una estrategia de enfriamiento de cuatro niveles.
4. El archivo [Web_Server_Requests_G](#) contiene un conjunto de datos que proporciona un registro histórico (sintético) de la tasa de llegada de solicitudes que un servidor web ha experimentado durante un intervalo de tiempo de 5 minutos. Esta información es esencial para comprender y gestionar la carga del servidor, planificar la infraestructura, y garantizar un rendimiento óptimo en un entorno de servidor web real.
- a)* Identificar patrones de tráfico web como horas pico de actividad y momentos de menor demanda.
 - b)* Ajuste una distribución gamma a sus datos de “Arrival_Rate”, trace el histograma de sus datos junto con la función de densidad de probabilidad (pdf) de la distribución gamma ajustada y muestre los parámetros de la distribución gamma ajustada.
 - c)* ¿Cuál es la probabilidad de que el servidor reciba más de 8 solicitudes durante el próximo intervalo?
 - d)* ¿Cuál es la probabilidad de que la tasa de llegada supere un umbral crítico de, por ejemplo, 15 solicitudes por minuto, en el próximo intervalo?
 - e)* Si la capacidad máxima de procesamiento del servidor es de 20 solicitudes por intervalo, ¿Cuál es la probabilidad de superar la capacidad máxima de procesamiento del servidor en el siguiente intervalo?

3. Entrega del Proyecto

Conformar grupos de 4 estudiantes y:

- Elaborar (y compartir) en, por ejemplo “Colab”, el código que les permitió generar los datos, figuras y cálculo de indicadores.
- Elaborar (y compartir) en un documento “PDF” que incluya las respuestas a los problemas anteriores.