



INSTITUTO TECNOLÓGICO Y DE ESTUDIOS SUPERIORES DE  
MONTERREY

INTELIGENCIA ARTIFICIAL AVANZADA PARA LA CIENCIA DE DATOS II

GRUPO 101

25 de octubre de 2024

---

## Actividad Integradora 1 - Precipitaciones máximas mensuales para el diseño de obras hidráulicas

---

*Autor:*

Luis Maximiliano López Ramírez - A00833321

*Profesor:*

Blanca R. Ruiz Hernández

## Introducción

Las precipitaciones extremas son eventos críticos que afectan la planificación y diseño de infraestructuras hidráulicas, como presas, canales, y sistemas de drenaje. El cálculo adecuado de la precipitación máxima esperada permite que estas estructuras se diseñen para soportar eventos extremos, asegurando su eficiencia y reduciendo el riesgo de desastres. Según investigaciones recientes, la elección de la distribución de probabilidad adecuada para modelar estos eventos es clave para estimar correctamente los periodos de retorno y asegurar que las infraestructuras sean resistentes y seguras (1; 2). La presente actividad se centra en analizar los datos de precipitación máxima mensual de Sonora, ajustando diferentes distribuciones de probabilidad y seleccionando la mejor opción para estimar los valores asociados a periodos de retorno de interés para el diseño hidráulico.

## Metodología

El análisis realizado para este estudio consistió en varias etapas clave, desde la recopilación y procesamiento de los datos de precipitación máxima hasta el ajuste de diferentes distribuciones de probabilidad. A continuación, se describen los pasos seguidos en detalle.

### Recopilación y Procesamiento de Datos

Los datos utilizados corresponden a los registros históricos de precipitaciones máximas mensuales en el estado de Sonora. Estos datos se obtuvieron de fuentes meteorológicas oficiales, asegurando que fueran representativos de eventos climáticos extremos que impactan en la región. Tras la recopilación, se realizó un preprocesamiento para limpiar los datos, eliminar valores atípicos no representativos y rellenar datos faltantes cuando fue necesario. Esta etapa permitió contar con un conjunto de datos robusto para el análisis posterior.

### Medidas Descriptivas y Visualización Inicial

Se calcularon medidas de tendencia central, como la media y la mediana, así como medidas de dispersión, como la desviación estándar y el rango intercuartílico. Además, se realizaron gráficos exploratorios, incluyendo histogramas y boxplots, para obtener una visión preliminar de la distribución de los datos. Estos análisis iniciales revelaron una distribución asimétrica hacia la derecha, lo que sugiere la presencia de eventos de precipitación máxima que son infrecuentes pero muy intensos.

### Ajuste a Diferentes Distribuciones de Probabilidad

Para modelar adecuadamente los datos de precipitación, se seleccionaron seis distribuciones de probabilidad conocidas por su capacidad para describir eventos extremos. A continuación se detalla el enfoque utilizado para cada una:

- **Distribución Normal:** Esta distribución se probó inicialmente para establecer una referencia, dado

que es común en muchos estudios de series temporales. Sin embargo, al evaluar los datos con pruebas de normalidad como Shapiro-Wilk y Kolmogorov-Smirnov, se observó que la distribución normal subestimaba las colas, haciendo difícil modelar eventos extremos.

- **Distribución Log-Normal:** Se seleccionó para capturar la asimetría observada en los datos. La log-normal es adecuada para datos que son positivos y exhiben una cola derecha, como es el caso de las precipitaciones. Los gráficos de densidad sugirieron que esta distribución ofrecía un mejor ajuste en comparación con la normal, especialmente en la parte alta de los valores.
- **Distribución Exponencial:** Aunque es una de las distribuciones más simples para modelar eventos aleatorios, la distribución exponencial no pudo capturar la variabilidad de los datos de manera adecuada. Su incapacidad para representar colas largas la hizo ineficiente para modelar precipitaciones máximas, ya que subestimaba significativamente los eventos más extremos.
- **Distribución Gamma:** Presentó un ajuste notable en todos los rangos de valores. La distribución Gamma es versátil, ya que puede adaptarse a diferentes formas y asimetrías, lo que la hace especialmente útil para modelar eventos climáticos como las precipitaciones. Las pruebas de bondad de ajuste mostraron que la distribución Gamma era una de las mejores opciones, ya que representaba tanto el centro de los datos como las colas, donde se encuentran los eventos extremos.
- **Distribución Weibull:** Similar a la Gamma, la Weibull es útil para modelar distribuciones asimétricas. En este caso, aunque ofreció un buen ajuste en los valores bajos y medios, tuvo dificultades para modelar los valores más extremos, lo que limitó su utilidad en este contexto.
- **Distribución Gumbel:** Conocida por su capacidad para modelar valores extremos, la Gumbel se utilizó específicamente para entender el comportamiento en las colas de la distribución. Esta distribución logró capturar adecuadamente los valores más altos de precipitación, pero mostró discrepancias en la parte central de la distribución.

### Análisis Comparativo

Para determinar cuál de estas distribuciones era la más adecuada para modelar los datos de precipitación máxima mensual, se realizaron varias comparaciones basadas en gráficos y pruebas estadísticas:

#### *Gráficos de Densidad Empírica vs. Teórica*

Se generaron gráficos de densidad para comparar la distribución empírica (basada en los datos reales) con las distribuciones teóricas ajustadas. Los resultados mostraron que la **Distribución Gamma** y la **Log-Normal** eran las más cercanas a la curva empírica. Esto sugiere que ambas distribuciones tienen la

capacidad de representar la asimetría natural de los datos, capturando tanto el centro de la distribución como las colas.

La **Distribución Exponencial** fue la que mostró el peor ajuste, ya que no lograba modelar la dispersión ni la variabilidad observada en los datos. La **Distribución Gumbel**, aunque adecuada para eventos extremos, no proporcionaba una representación precisa de los valores moderados y bajos, lo que podría ser una limitación al usarla para estimar valores de diseño en el rango completo de datos.

### *Gráficos de Probabilidad Acumulada y Excedencia*

Además de los gráficos de densidad, se crearon gráficos de probabilidad acumulada y excedencia para cada distribución. La **Distribución Gamma** mostró una excelente concordancia con la curva empírica, confirmando que esta distribución no solo capturaba los eventos más frecuentes, sino también los extremos. La **Distribución Log-Normal** también mostró buenos resultados, pero tuvo ligeras discrepancias en la cola inferior, lo que indica que podría subestimar algunos valores de precipitación moderados.

### *Pruebas de Bondad de Ajuste*

Se aplicaron pruebas estadísticas como la prueba KS para evaluar la bondad de ajuste de cada distribución. Los resultados de estas pruebas respaldaron los hallazgos visuales, mostrando que la **Distribución Gamma** tenía el estadístico KS más bajo y un p-valor alto, lo que indica que no se rechazaba la hipótesis nula de que los datos seguían una distribución Gamma. Esta conclusión fue reforzada por los gráficos de probabilidad acumulada, donde la línea teórica de la Gamma seguía de cerca a la línea empírica, sugiriendo que la probabilidad de excedencia calculada sería confiable para el diseño hidráulico.

### **Estimación de la Precipitación Máxima para Diferentes Periodos de Retorno**

Para demostrar la capacidad predictiva del modelo ajustado, se calcularon valores de precipitación máxima para periodos de retorno de 50, 100, 200 y 500 años utilizando la **Distribución Gamma**. Estos cálculos fueron realizados mediante la fórmula de probabilidad de excedencia:

$$P_{noex} = 1 - \frac{1}{T_r}$$

donde  $T_r$  es el periodo de retorno en años. Los valores obtenidos se presentan en la sección de resultados e interpretación, demostrando cómo el modelo Gamma puede ser una herramienta útil para prever eventos extremos que podrían impactar en el diseño de infraestructuras críticas.

### **Resultados e Interpretación**

La **Tabla 1** resume los valores obtenidos de los cálculos para diferentes periodos de retorno:

Los resultados confirman que la precipitación máxima esperada aumenta con el periodo de retorno. Esto es

Periodo de Retorno (años)	Precipitación Máxima (mm)
50	202.79
100	215.28
200	227.12
500	242.02

**Cuadro 1**

*Valores de precipitación máxima para distintos periodos de retorno*

coherente con la teoría, ya que eventos más extremos (y menos frecuentes) ocurren a intervalos más largos. La distribución Gamma, al ajustarse bien a la curva empírica de probabilidad de excedencia, permitió estimar de manera precisa estos valores, destacando su capacidad para modelar eventos extremos con una alta precisión.

### Discusión y Conclusiones

La selección de la **distribución Gamma** como la mejor opción para modelar los datos de precipitación máxima mensual en Sonora fue basada en su capacidad para ajustarse tanto en el centro como en las colas de la distribución, mostrando resultados consistentes en las pruebas KS y otros análisis comparativos. Esta distribución demostró ser adecuada para estimar periodos de retorno, proporcionando valores precisos que pueden ser utilizados para diseñar infraestructuras hidráulicas seguras y eficaces.

La importancia de elegir correctamente la distribución de probabilidad radica en la necesidad de prever eventos extremos de manera confiable. Un diseño basado en distribuciones que subestiman o sobrestiman las probabilidades de eventos de lluvia severa puede llevar a costos innecesarios o, peor aún, al fallo de la infraestructura. Las recomendaciones obtenidas del análisis destacan la necesidad de seguir ajustando modelos basados en datos actualizados y de aplicar revisiones periódicas para asegurar la resiliencia ante eventos extremos, que pueden cambiar con las variaciones climáticas observadas en los últimos años (1; 2; 3).

## Referencias

- Coronado-Hernández, Ó. E., Merlano-Sabalza, E., Díaz-Vergara, Z., & Coronado-Hernández, J. R. (2020). Selection of hydrological probability distributions for extreme rainfall events in the regions of Colombia. *Water*, 12(5), 1397. <https://doi.org/10.3390/w12051397>.
- Mizuki, C., & Kuzuha, Y. (2023). Frequency analysis of hydrological data for urban floods—Review of traditional methods and recent developments. *Water*, 15(13), 2490. <https://doi.org/10.3390/w15132490>.
- Mahdi, T. F., & Cenac, B. (2021). Application of extreme value analysis in urban stormwater management: Assessing rainfall patterns and return periods. *Journal of Hydrology*, 598, 126191. <https://doi.org/10.1016/j.jhydrol.2021.126191>.