

Estadística Bayesiana  
Pregrado en Ciencia de Datos  
Universidad Externado



Andrés Martínez  
MSc Stochastic Engineering  
Universidad Externado

---

# Capítulo 1

## Primer taller

### Identificación del Problema Central y Elección de la Distribución Binomial

- Exponer el problema central de la identificación de fuentes de contaminación y por qué la distribución binomial es seleccionada para abordar este problema.
- Analizar cómo esta distribución modela adecuadamente el problema y por qué es relevante para la detección de fuentes de contaminación.

### Modelo del Paper "Bayesian Approach for Real-Time Probabilistic Contamination Source Identification y la Regla Bayesiana

Presente el modelo del paper, usando las dos perspectivas haciendo énfasis en las diferencia en cada enfoque.

### Generación de Simulación de Señales Personalizadas

Genere una simulación de señales utilizando una semilla que use como valor el último número de su cédula. Genere 200 observaciones y explique porque los autores eligieron esta variable para presentar diferentes escenarios de contaminación.

## Implementación del Algoritmo de la Regla Bayesiana

Siguiendo las pautas del paper, implemente el algoritmo de la regla bayesiana para la identificación de fuentes de contaminación en tiempo real. Observe cómo este proceso actualiza las probabilidades a medida que se obtienen más datos, inicie con 20 y termine con los 200.

## Algoritmo de la Distribución Beta Binomial

Desarrolle el algoritmo de la distribución beta binomial iniciando con 20 valores y terminando en 200. Cree una secuencia de 20 valores diferentes para los hiperparámetros  $\alpha_0$  y  $\beta_0$  (Estos se presenta en el paper, ojo son diferentes a  $\alpha$  y  $\beta$ ), explique el papel de cada uno de ellos en la creación de la distribución prior y su efecto en la distribución posterior.

## Generación de Gráficos de Comparación

- Presente los resultados de la implementación de la regla bayesiana para todos los escenarios (De 20 a 200) y diferentes valores de  $\theta$ .
- Presente los resultados de la implementación de la distribución conjugada beta-binomial para todos los escenarios (De 20 a 200), usando todos los hiperparámetros de  $\alpha_0$  y  $\beta_0$  y diferentes valores de  $\theta$ .
- Compare los gráficos y explique las posibles diferencias, considerando aspectos como la cantidad de datos, la elección de hiperparámetros y las características específicas de las simulaciones.

## Ideas

Presente un problema en donde se pueda implementar la distribución conjugada beta binomial con las siguientes características:

- Defina un valor para los hiperparámetros y las razones para que tengan esos valores.

- Cree la distribución likelihood para diferentes valores de  $\theta$ .
- Simule un vector de 300 valores, estos debe ser desbalanceados.
- Grafique la prior la likelihood y la posterior.
- Explique el resultado de la posterior.

## Entrega

Genere el documento en jupyter y entreguelo antes de acabar la clase en teams.  
Grupos de 2 personas.