Inferencia Bayesiana

Una Herramienta para el Análisis de Datos

Notas del curso, UNAM Agosto-Diciembre 2025

Autor: Imelda Trejo

11 de agosto de 2025

Resumen

Estas notas introducen los fundamentos de la inferencia bayesiana para estudiantes de maestría y doctorado en matemáticas del Centro de Ciencias Matemáticas, UNAM Morelia. Cubrimos desde los conceptos filosóficos fundamentales hasta las bases matemáticas rigurosas, incluyendo espacios muestrales, distribuciones de probabilidad, el teorema de Bayes y el concepto de intercambiabilidad.

Índice

1. Contexto Histórico

El término de inferencia Bayesiana es en honor a Thomas Bayes.

Figura 1: Esta es probablemente la únia imagen que se tiene de Thomas Bayes, imagen obtenida de https://en.wikipedia.org/wiki/Thomas_Bayes

Thomas Bayes fue un matemático y estadístico inglés que formuló el **teorema de Bayes**, una regla fundamental en estadística y aprendizaje automático moderno. Su trabajo dio origen a toda un área de la estadística: la **estadística bayesiana**, que se centra en determinar la probabilidad de un evento basándose en el conocimiento previo de condiciones que puedan estar relacionadas con dicho evento. Esto es, para dos eventos A y B con, $\mathbb{P}(B) > 0$:

$$\mathbb{P}(A \mid B) = \frac{\mathbb{P}(B \mid A)\mathbb{P}(A)}{\mathbb{P}(B)}$$

Aquí, el evento A no lo conocemos con precisión, pero creemos que este evento tiene una distribución $\mathbb{P}(A)$, que también se le conoce como *distribución a priori.* Luego, la probabilidad de A se actualiza la conocer la relación entre los eventos A con B. Siendo, $\mathbb{P}(A \mid B)$ la *distribución posterior*, distribución que resulta al incorporar los nuevos conocimientos sobre el evento A.

El teorema de Bayes proporciona un marco matemático para actualizar nuestras creencias (falta de conocimientos) sobre cantidades inciertas a medida que observamos nuevos datos—un proceso que refleja cómo progresa el conocimiento científico.

2. Introducción a la Estadística e Inferencia Estadística

2.1. ¿Qué es la estadística?

La estadística es la ciencia que **recopila, organiza, analiza e interpreta datos** para entender fenómenos y ayudar en la toma de decisiones. Podemos dividirla en dos ramas principales:

- **Estadística descriptiva**: Resume y describe los datos mediante tablas, gráficas y medidas como media, mediana y varianza.
- **Estadística inferencial**: Utiliza datos muestrales para hacer **inferencias** sobre una población más grande o proceso subyacente.

2.2. Una Distinción Fundamental: Interpretación vs. Explicación

> *"La estadística debe considerarse una interpretación de los fenómenos naturales, más que una explicación."*

Esta afirmación destaca un concepto crucial:

- La **estadística no explica por qué ocurre algo**; en cambio, **resume lo que observamos** y **cuantifica la incertidumbre** en nuestras observaciones.
- La **estadística** nos dice: *"Dadas estas observaciones, así es como se comportan los datos, y esta es la probabilidad de que este patrón sea real."*
- La **explicación** (el "por qué") típicamente proviene de una **teoría científica o modelo causal**.

Ejemplo 2.1. Patrones de influenza estacional.

- **Fenómeno**: "Los casos de gripe aumentan durante el invierno."
- **Análisis estadístico**: Analiza datos históricos, calcula aumentos promedio, estima probabilidades de brotes.
- **Explicación científica**: "El virus sobrevive mejor a temperaturas frías, y las personas pasan más tiempo en espacios cerrados en proximidad."

2.3. Por qué importa esta distinción

Confundir interpretación estadística con explicación causal lleva a:

1. Creer que correlación implica causalidad
2. Aplicar incorrectamente resultados estadísticos fuera de su contexto científico
3. Ignorar que la estadística **depende de la calidad y representatividad de los datos**

Ejemplo 2.2 (Ejemplo clásico). Las ventas de helado y los incidentes de ahogamiento muestran correlación positiva.

- **Observación estadística**: Cuando aumentan las ventas de helado, también aumentan los ahogamientos.
- **Inferencia causal incorrecta**: El helado causa ahogamientos.
- **Explicación real**: El calor es la causa común—en verano, más gente compra helado y más gente va a nadar, resultando en más ahogamientos.

2.4. Incertidumbre en Estadística

La **incertidumbre** representa **lo que no sabemos con certeza** sobre un fenómeno o valor de parámetro. En modelado matemático y análisis de datos, la incertidumbre surge de:

1. **Errores de medición** en la recolección de datos
2. **Variabilidad natural** en los fenómenos
3. **Observaciones limitadas**—observamos muestras, no poblaciones completas
4. **Limitaciones muestrales**—muestras pequeñas o no representativas
5. **Inadecuación del modelo**—los modelos son simplificaciones de la realidad
6. **Ruido en los datos**—fluctuaciones aleatorias en las mediciones

2.4.1. Cuantificación de la Incertidumbre

La estadística **cuantifica** esta incertidumbre mediante diferentes enfoques:

- **Enfoque frecuentista**: Intervalos de confianza, valores p
- **Enfoque bayesiano**: Distribuciones posteriores, intervalos creíbles

2.5. El Paradigma Bayesiano

Principio bayesiano fundamental: Cualquier cantidad que no conocemos con certeza (ya sea un parámetro, resultado futuro, o proceso no observado) se representa mediante una **distribución de probabilidad**.

Esta representación probabilística nos permite:

- Expresar grados de creencia sobre cantidades desconocidas
- Actualizar estas creencias conforme nuevos datos se vuelven disponibles
- Hacer predicciones con incertidumbre cuantificada

2.6. Introducción al análisis de datos

En el análisis de datos, es fundamental entender el tipo de variable con la que se trabaja, ya que esto determina qué métodos estadísticos son adecuados para su análisis.

1. Datos dicotómicos

- También llamados **variables binarias**.
- Solo pueden tomar **dos posibles valores o categorías**.
- Ejemplos:
  - Sexo: masculino o femenino
  - Respuesta: sí o no
  - Presencia de enfermedad: sí o no
- Son un caso particular de variable categórica, pero con solo dos opciones.

2. Datos categóricos

- Variables que toman valores en un conjunto limitado y no numérico de categorías o grupos.
- No tienen un orden natural (a menos que sean ordinales).
- Ejemplos:
  - Grupo sanguíneo: A, B, AB, O
  - Estado civil: soltero, casado, divorciado, viudo
  - Color de ojos: azul, marrón, verde
- Se usan para clasificar observaciones en grupos.

3. Datos continuos

- Variables numéricas que pueden tomar cualquier valor dentro de un rango o intervalo.
- Suelen medirse con unidades y pueden tener decimales.
- Ejemplos:
  - Peso (kg)
  - Estatura (cm)
  - Temperatura ($^{\circ}C$)
- Permiten calcular promedios, varianzas y aplicar técnicas estadísticas paramétricas.

2.7. Ejercios de cómputo

Las bases de datos se encuentran en
https://github.com/imelit/Bayesian_inference_class_2025/tree/main
Asignación

- Hacer un reporte con el resumen de los datos, son los datos continuos o categóricos? ¿Qué tipo de gráficos usaste para mostrar variablidad en edades, sexo, comorbidades, casos de Flu?
- De entre los casos reportados por Flu. Realiza una gráfica de barras del total de casos por flu estratificado por meses, partiendo de octubre a mayo, tempora de influenza, https://www.cdc.gov/flu/about/season.html.

- Visualiza la serie de tiempo de Influenza estratificado por grupos de edad (casos contra tiempo).
- ¿Cómo se correlaciona Flu contra otras comorbidades?

  Guia para el análisis:

  https://www.cdc.gov/fluview/overview/fluview-interactive.html

3. Fundamentos Matemáticos

3.1. Espacios Muestrales y Eventos

Definición 3.1 (Espacio Muestral)**.** El espacio muestral Ω es el conjunto de todos los posibles resultados de un experimento aleatorio.

Definición 3.2 (Evento)**.** Un evento A es un subconjunto del espacio muestral Ω.

Definición 3.3 (σ-álgebra)**.** Una colección $\mathcal{F}$ de subconjuntos de Ω es una σ-álgebra si:

1. $\Omega \in \mathcal{F}$
2. Si $A \in \mathcal{F}$, entonces $A^c \in \mathcal{F}$
3. Si $A_1, A_2, \ldots \in \mathcal{F}$, entonces $\bigcup_{i=1}^{\infty} A_i \in \mathcal{F}$

3.2. Particiones

Definición 3.4 (Partición)**.** Una colección de eventos $\{B_1, B_2, \ldots, B_k\}$ forma una partición de Ω si:

1. $B_i \cap B_j = \emptyset$ para $i \neq j$ (mutuamente excluyentes)
2. $\bigcup_{i=1}^{k} B_i = \Omega$ (exhaustivos)
3. $\mathbb{P}(B_i) > 0$ para todo i

Las particiones son fundamentales en análisis bayesiano para aplicar la ley de probabilidad total.

3.3. Variables Aleatorias

Definición 3.5 (Variable Aleatoria)**.** Una variable aleatoria X es una función medible $X : \Omega \to \mathbb{R}$.

Definición 3.6 (Variable Aleatoria Discreta)**.** X es discreta si toma valores en un conjunto contable.

Definición 3.7 (Variable Aleatoria Continua)**.** X es continua si existe una función $f(x) \geq 0$ tal que

$$\mathbb{P}(a < X \leq b) = \int_a^b f(x)\, dx$$

para todo $a < b$.

3.4. Funciones de Probabilidad

Para variables aleatorias discretas, la **función de masa de probabilidad (FMP)** es:

$$\mathbb{P}(X = x) = p(x), \quad \text{donde} \quad \sum_x p(x) = 1$$

Para variables aleatorias continuas, la **función de densidad de probabilidad (FDP)** satisface:

$$\int_{-\infty}^{\infty} f(x)\, dx = 1$$

3.5. Distribuciones Discretas Importantes

3.5.1. Distribución Bernoulli

Definición 3.8 (Distribución Bernoulli)**.** $X \sim \text{Bernoulli}(\theta)$ si $X \in \{0, 1\}$ con $\mathbb{P}(X = 1) = \theta$.

FMP: $p(x) = \theta^x (1-\theta)^{1-x}$ para $x \in \{0, 1\}$
Propiedades: $\mathbb{E}[X] = \theta$, $\text{Var}(X) = \theta(1-\theta)$

3.5.2. Distribución Binomial

Definición 3.9 (Distribución Binomial)**.** $X \sim \text{Binomial}(n, \theta)$ representa el número de éxitos en n ensayos independientes Bernoulli(θ).

FMP: $p(x) = \binom{n}{x} \theta^x (1-\theta)^{n-x}$ para $x \in \{0, 1, \ldots, n\}$
Propiedades: $\mathbb{E}[X] = n\theta$, $\text{Var}(X) = n\theta(1-\theta)$

3.5.3. Distribución Poisson

Definición 3.10 (Distribución Poisson). $X \sim \text{Poisson}(\lambda)$ modela eventos raros que ocurren a tasa λ.

FMP: $p(x) = \frac{e^{-\lambda}\lambda^x}{x!}$ para $x \in \{0, 1, 2, \ldots\}$
Propiedades: $\mathbb{E}[X] = \text{Var}(X) = \lambda$

3.5.4. Distribución Binomial Negativa

Definición 3.11 (Distribución Binomial Negativa). $X \sim \text{NegBin}(r, \theta)$ modela el número de fallas antes de lograr r éxitos.

FMP: $p(x) = \binom{x+r-1}{x}\theta^r(1-\theta)^x$ para $x \in \{0, 1, 2, \ldots\}$
Propiedades: $\mathbb{E}[X] = \frac{r(1-\theta)}{\theta}$, $\text{Var}(X) = \frac{r(1-\theta)}{\theta^2}$

3.6. Distribuciones Continuas Importantes

3.6.1. Distribución Beta

Definición 3.12 (Distribución Beta). $X \sim \text{Beta}(\alpha, \beta)$ con soporte en $[0, 1]$.

FDP: $f(x) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}x^{\alpha-1}(1-x)^{\beta-1}$
Propiedades: $\mathbb{E}[X] = \frac{\alpha}{\alpha+\beta}$, $\text{Var}(X) = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$
Rol bayesiano: Priori conjugada natural para verosimilitudes Bernoulli/Binomial.

3.6.2. Distribución Gamma

Definición 3.13 (Distribución Gamma). $X \sim \Gamma(\alpha, \beta)$ con forma $\alpha > 0$ y tasa $\beta > 0$.

FDP: $f(x) = \frac{\beta^\alpha}{\Gamma(\alpha)}x^{\alpha-1}e^{-\beta x}$ para $x > 0$
Propiedades: $\mathbb{E}[X] = \frac{\alpha}{\beta}$, $\text{Var}(X) = \frac{\alpha}{\beta^2}$
Rol bayesiano: Priori conjugada para verosimilitudes Poisson.

3.6.3. Distribución Normal

Definición 3.14 (Distribución Normal). $X \sim \text{Normal}(\mu, \sigma^2)$ con ubicación μ y escala σ.

FDP: $f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$
Propiedades: $\mathbb{E}[X] = \mu$, $\text{Var}(X) = \sigma^2$

3.6.4. Distribución Log-Normal

Definición 3.15 (Distribución Log-Normal)**.** $X \sim \text{LogNormal}(\mu, \sigma^2)$ si $\log(X) \sim \text{Normal}(\mu, \sigma^2)$.

FDP: $f(x) = \frac{1}{x\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(\log x - \mu)^2}{2\sigma^2}\right)$ para $x > 0$
Propiedades: $\mathbb{E}[X] = \exp\left(\mu + \frac{\sigma^2}{2}\right)$, $\text{Var}(X) = \exp(2\mu+\sigma^2)(\exp(\sigma^2)-1)$

4. La Regla de Bayes: El Fundamento

4.1. Probabilidad Condicional

Definición 4.1 (Probabilidad Condicional)**.** Para eventos A y B con $\mathbb{P}(B) > 0$:

$$\mathbb{P}(A \mid B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}$$

4.2. Teorema de Bayes

Teorema 4.1 (Regla de Bayes)**.** Para eventos A y B con $\mathbb{P}(B) > 0$:

$$\mathbb{P}(A \mid B) = \frac{\mathbb{P}(B \mid A)\mathbb{P}(A)}{\mathbb{P}(B)}$$

Forma extendida: Si $\{B_1, B_2, \ldots, B_k\}$ forma una partición de Ω:

$$\mathbb{P}(B_i \mid A) = \frac{\mathbb{P}(A \mid B_i)\mathbb{P}(B_i)}{\sum_{j=1}^{k} \mathbb{P}(A \mid B_j)\mathbb{P}(B_j)}$$

4.3. Inferencia Estadística Bayesiana

En contextos estadísticos, la regla de Bayes se convierte en:

$$\boxed{\pi(\theta \mid x) = \frac{L(x \mid \theta)\pi(\theta)}{m(x)}}$$

Donde:

- $\pi(\theta)$: **Distribución priori** (creencias antes de observar datos)
- $L(x \mid \theta)$: **Función de verosimilitud** (probabilidad de datos dado parámetro)

- $\pi(\theta \mid x)$: **Distribución posteriori** (creencias actualizadas tras observar datos)
- $m(x) = \int L(x \mid \theta)\pi(\theta)\, d\theta$: **Verosimilitud marginal** (constante normalizadora)

4.4. Proceso de Aprendizaje Bayesiano

1. **Especificación de priori**: Codificar creencias iniciales sobre θ
2. **Observación de datos**: Recopilar datos x
3. **Evaluación de verosimilitud**: Evaluar $\mathbb{P}(x \mid \theta)$ para diferentes valores de θ
4. **Cómputo de posteriori**: Aplicar regla de Bayes para actualizar creencias
5. **Decisión/Predicción**: Usar posteriori para inferencia

5. Intercambiabilidad

5.1. Definición e Intuición

Definición 5.1 (Intercambiabilidad)**.** Las variables aleatorias $X_1, X_2, \ldots, X_n$ son intercambiables si su distribución conjunta es invariante bajo permutaciones:

$$\mathbb{P}(X_1 = x_1, \ldots, X_n = x_n) = \mathbb{P}(X_{\pi(1)} = x_1, \ldots, X_{\pi(n)} = x_n)$$

para cualquier permutación π.

Intuición: No podemos distinguir entre las variables basándose en sus índices—son “simétricas” en nuestras creencias.

5.2. Teorema de De Finetti

Teorema 5.1 (De Finetti)**.** Si $X_1, X_2, \ldots$ son infinitamente intercambiables, entonces existe una medida aleatoria P tal que:

$$\mathbb{P}(X_1 \in A_1, \ldots, X_n \in A_n) = \int \prod_{i=1}^{n} P(A_i)\, d\mu(P)$$

Esto justifica el enfoque bayesiano: las observaciones intercambiables pueden modelarse como condicionalmente independientes dado un parámetro desconocido.

5.3. Implicaciones Prácticas

1. **Justifica modelos paramétricos**: La intercambiabilidad lleva naturalmente a asumir que las observaciones son i.i.d. dados los parámetros
2. **Elicitación de prioris**: Ayuda a determinar distribuciones prioris apropiadas
3. **Verificación de modelos**: Las violaciones sugieren inadecuación del modelo

6. Resumen y Próximos Pasos

Esta clase estableció los fundamentos filosóficos y matemáticos de la inferencia bayesiana:

1. **Fundamento filosófico**: La estadística interpreta más que explicar; la incertidumbre es fundamental
2. **Marco matemático**: La teoría de probabilidad proporciona el lenguaje para la incertidumbre
3. **Regla de Bayes**: El mecanismo para actualizar creencias con nueva información
4. **Intercambiabilidad**: El principio que justifica los modelos bayesianos paramétricos

Temas de la próxima clase:

- Prioris conjugadas y cómputo de posteriori
- Pruebas de hipótesis bayesianas y selección de modelos
- Métodos computacionales (MCMC, inferencia variacional)
- Modelos jerárquicos y Bayes empírico

Referencias

[1] Hoff, P. D. (2009). *A First Course in Bayesian Statistical Methods.* Springer.

[2] Kruschke, J. K. (2015). *Doing Bayesian Data Analysis: A Tutorial with R, JAGS, and Stan* (2nd ed.). Academic Press.

[3] Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., & Rubin, D. B. (2013). *Bayesian Data Analysis* (3rd ed.). CRC Press.

[4] Robert, C. P. (2007). *The Bayesian Choice: From Decision-Theoretic Foundations to Computational Implementation* (2nd ed.). Springer.