Inferencia bayesiana

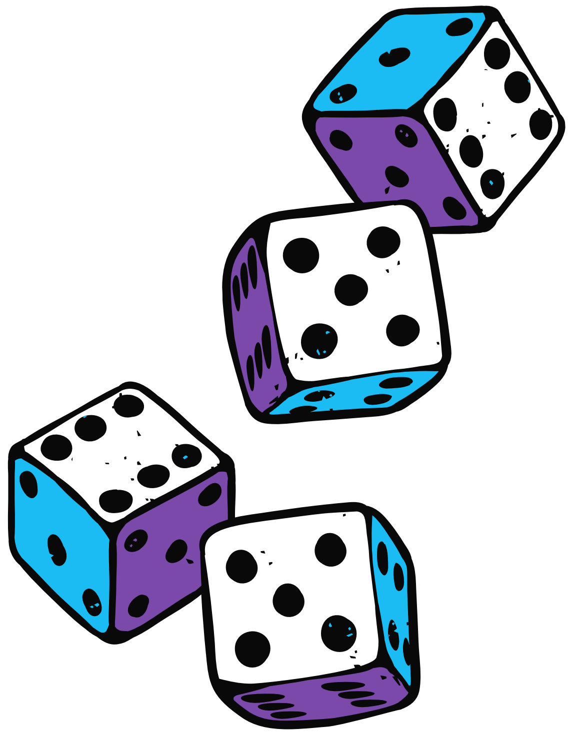
B42025

Enfoques frecuentista y bayesiano

Son dos paradigmas para interpretar la probabilidad y realizar inferencias a partir de datos.

Frecuentista

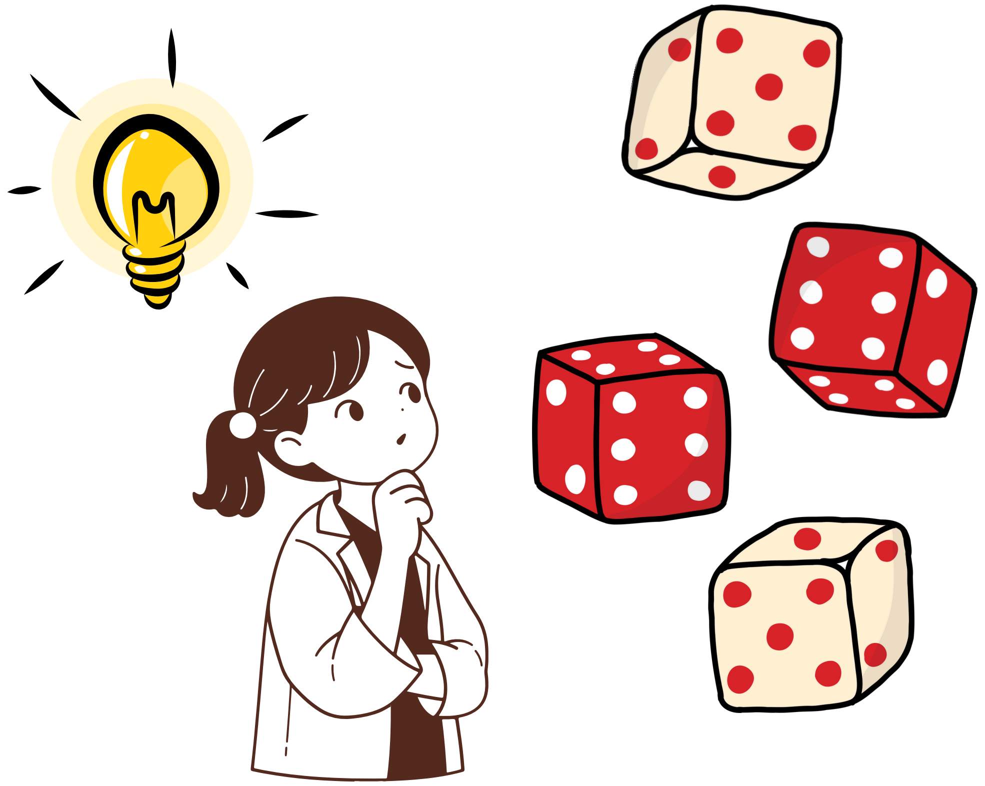
Solo usa los datos que observa



Interpreta la probabilidad como la frecuencia relativa en un número infinito de repeticiones.

Bayesiano

Combina datos con creencias previas y las va actualizando



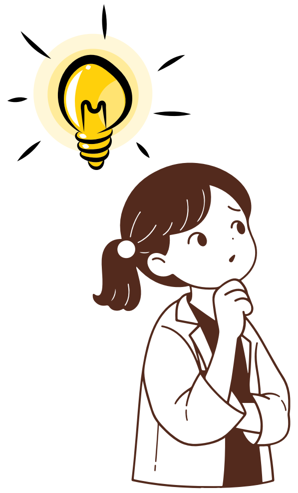
Considera la probabilidad como un grado de creencia subjetivo.

Dos caminos distintos para responder la misma pregunta estadística.

Inferencia bayesiana

Es un enfoque basado en el Teorema de Bayes, que consiste en ir actualizando nuestra creencia sobre un parámetro desconocido, a medida que vamos recolectando u observando muestras.

prior

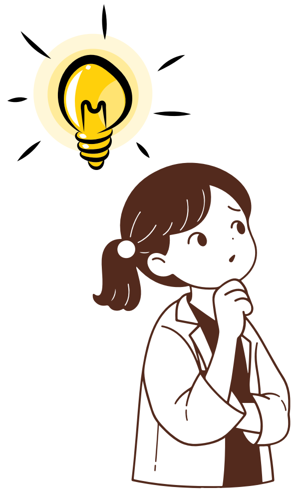


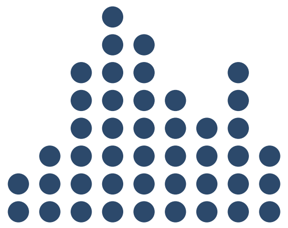
- La inferencia bayesiana requiere una **creencia a priori** acerca de los valores de los parámetros del modelo.
- Tipos de creencias a priori:
  - No informativas o planas (creencias subjetivas).
  - Informativas (experiencias previas)
- Cualquiera sea el tipo de creencia, debemos asignarle una distribución.

Inferencia bayesiana

Es un enfoque basado en el Teorema de Bayes, que consiste en ir actualizando nuestra creencia sobre un parámetro desconocido, a medida que vamos recolectando u observando muestras.

prior → **verosimilitud** → **posteriori**





A medida que tenga más muestras, estas van a pesar más que las suposiciones que haya tenido al principio.

Estimación de parámetros
Enfoques frecuentista y bayesiano

ASPECTO	FRECUENTISTA	BAYESIANO
Naturaleza del parámetro	Es fijo (pero desconocido).	Tiene una naturaleza aleatoria. Se describe con una distribución de probabilidad.
Base de inferencia	Se basa en los datos observados (sin creencias previas).	Parte de una creencia previa (prior) que se actualiza con nuevos datos (posterior).
Método de estimación	Usa estimadores puntuales (ej. máxima verosimilitud) y construye intervalos de confianza.	Usa el teorema de Bayes para combinar prior y verosimilitud y obtener la distribución posterior del parámetro.

Teorema de Bayes (repaso)

Si B_i es una partición del espacio muestral Ω y A es un evento de Ω, entonces dado un i fijo:

$$P(B_i|A) = \frac{P(A|B_i)P(B_i)}{\sum_j P(A|B_j)P(B_j)}$$

Ejercicio 01

Según el pronóstico, la probabilidad de que llueva hoy es del 50%.
Sabemos que, cuando llueve, el 90% de las personas llevan paraguas; y que cuando no llueve, solo el 20% de las personas llevan paraguas (por costumbre).

De repente, ves pasar a una persona con paraguas. Basándote en esa evidencia, actualizá tu creencia inicial sobre la probabilidad de que llueva hoy. ¿Cuál es la nueva probabilidad?

Generalización de Bayes

Dada una muestra aleatoria $\overline{X} = (X_1, \ldots, X_n)$ con valores observados $\overline{x} = (x_1, \ldots, x_n)$ y una densidad observada $\pi(\theta)$, la densidad a posteriori está dada por:

$$f_{\Theta|\overline{X}=\bar{x}}(\theta) = \frac{f_{\overline{X}|\Theta=\theta}(\bar{x})\pi(\theta)}{\int f_{\overline{X}|\Theta=\theta}(\bar{x})\pi(\theta)d\theta}$$

Generalización de Bayes

Posteriori

(creencia actualizada)

Likelihood o verosimilitud

(qué tan probable es observar los datos dado un parámetro)

Prior

(creencia inicial)

$$f_{\Theta|\overline{X}=\bar{x}}(\theta) = \frac{f_{\overline{X}|\Theta=\theta}(\bar{x})\pi(\theta)}{\int f_{\overline{X}|\Theta=\theta}(\bar{x})\pi(\theta)d\theta}$$

Factor de normalización

(para que las probabilidades posteriores sumen 1)

Generalización de Bayes

- En general, podemos afirmar que:

$$Posterior \propto Likelihood\ X\ Prior$$

- Cuando llegan nuevos datos, el posterior es el nuevo prior.
- El proceso se repite una y otra vez con nuevos datos (actualización bayesiana).

Distribuciones conjugadas

Si desconocemos la distribución a priori, ¿cómo la elegimos?

- Si no tenemos ninguna información, elegimos una distribución “razonable”, que no favorezca a ningún valor. Ejemplo: Uniforme. ¡Ojo con el soporte!
- Por simplicidad podemos elegir familias de **distribuciones conjugadas**. Como ya sabemos cuál va a ser la distribución a posteriori, solo nos preocupamos por averiguar sus parámetros.

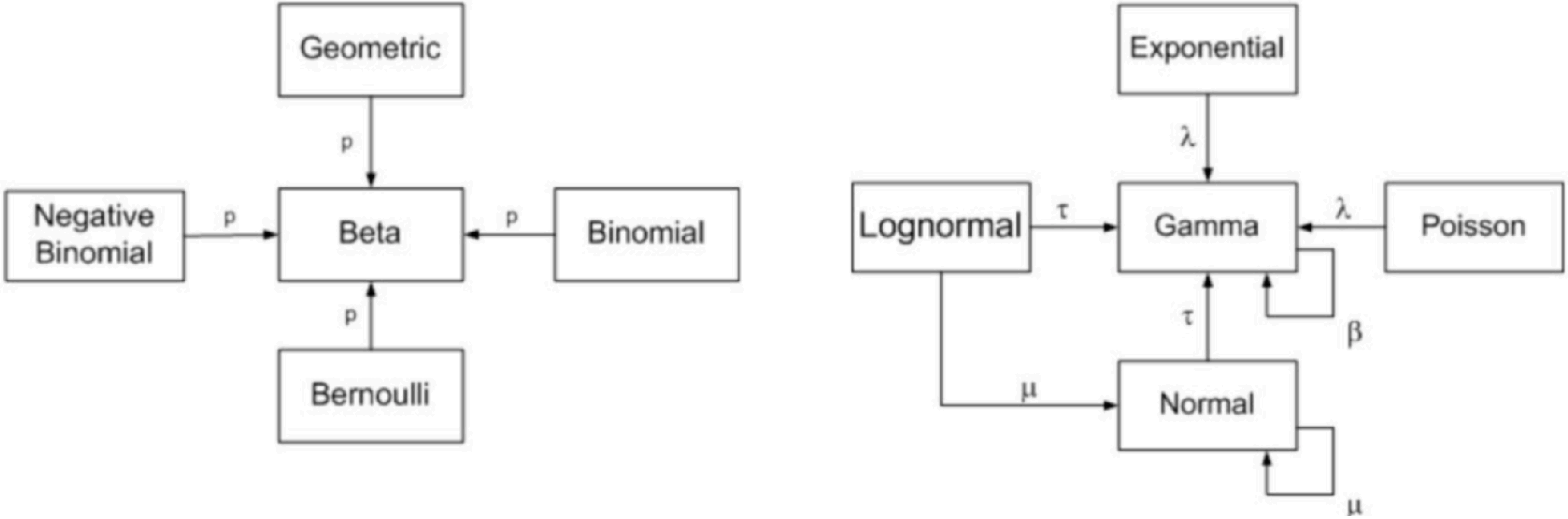
Ejemplo:

Verosimilitud	A priori	A posteriori
$X \sim \text{Binomial}(n, p)$	$p \sim \text{Beta}(\alpha, \beta)$	$p \sim \text{Beta}(\alpha', \beta')$

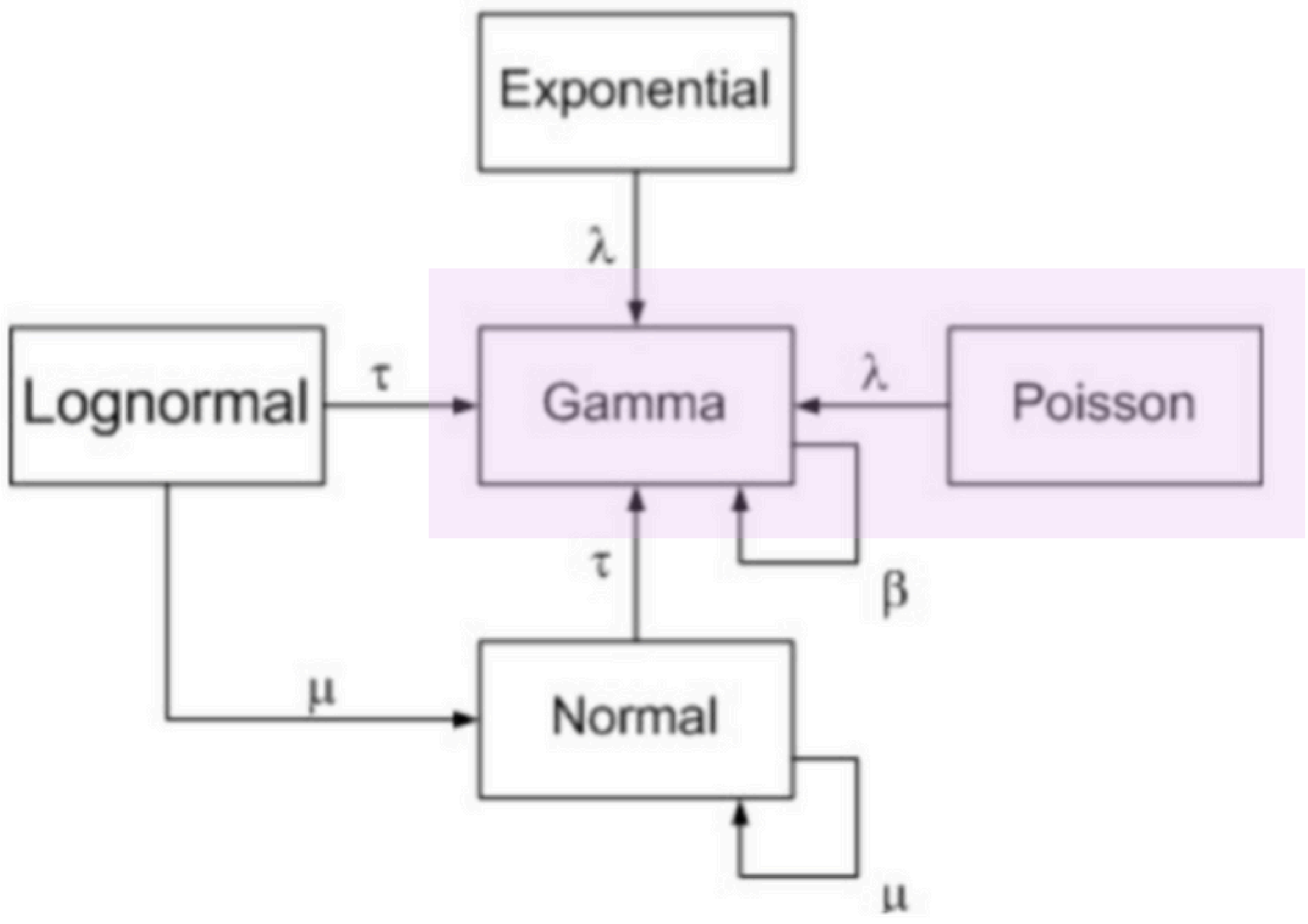
conjugadas

misma familia

Distribuciones conjugadas



Distribuciones conjugadas



- Las flechas conectan distribuciones conjugadas.
- Las flechas van desde la distribución de los datos hacia la distribución a priori/a posteriori del parámetro.
- Sobre la flecha está indicado el parámetro que estoy estimando.

Ejemplo:

Tengo datos con distribución Poisson y quiero estimar λ *.*
Gamma es la conjugada de la Poisson.
Entonces si digo que la distribución a priori de λ *es Gamma, la distribución a posteriori de* λ *también va a ser Gamma, pero con otros parámetros.*

Ejercicio 02

La cantidad de accidentes semanales en una planta industrial tiene una distribución de Poisson de media μ. En una muestra de 100 semanas se observaron las frecuencias:

Cantidad de accidentes	0	1	2	3	4	5
Frecuencia	10	29	25	17	13	6

A priori, μ tiene una distribución exponencial de media 2.
Hallar la distribución a posteriori de μ.

Ejercicio 03

La posición del impacto en un tiro al blanco (en decímetros) respecto del cero sobre el eje x es una variable aleatoria X con distribución normal de media cero y varianza $1/\theta$ donde θ representa la precisión del tirador.

A priori, la precisión θ tiene una distribución Chi-cuadrado de 8 grados de libertad.
Lucas tiró 10 veces al blanco y observó que $\sum_{x=1}^{10} x^2 = 17$

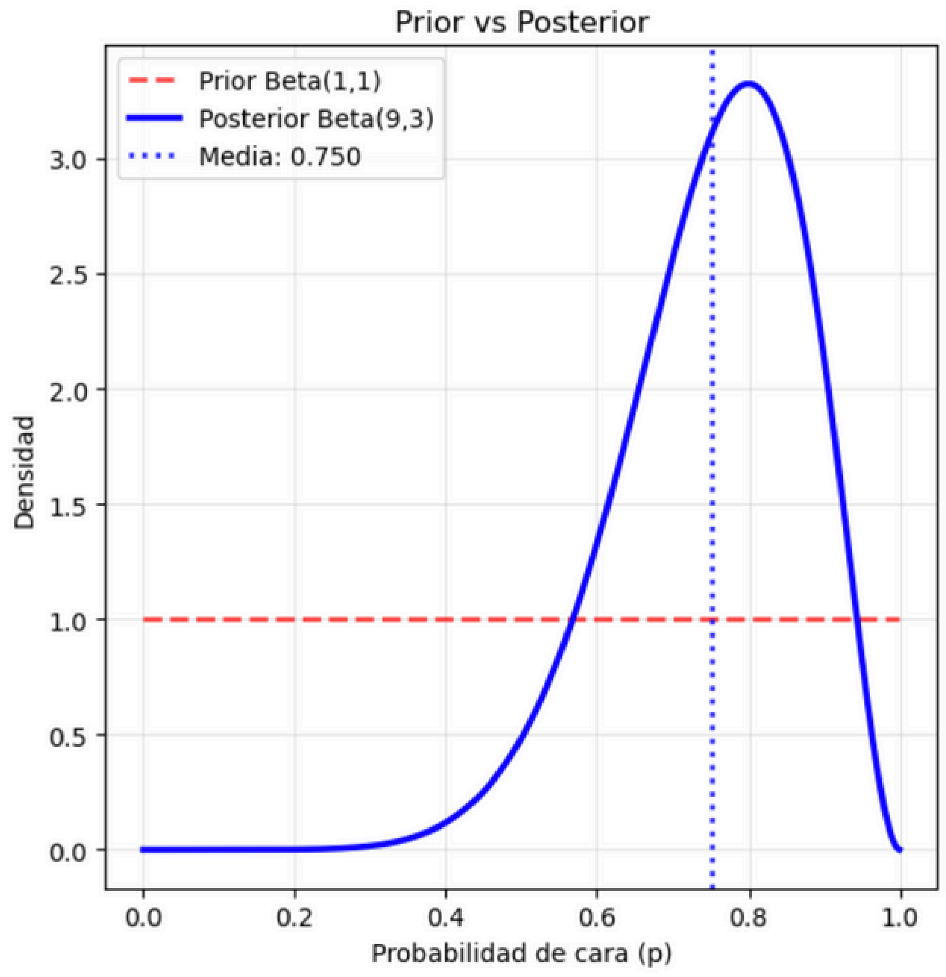
Hallar la distribución a posteriori de θ .

Ejemplo en Python

Moneda cargada

Lanzamos una moneda N veces y obtenemos C caras.

- Prior (a priori): asumimos que la probabilidad de cara (p) es uniforme entre 0 y 1.
- Likelihood (verosimilitud): la probabilidad de obtener 8 caras en 10 lanzamientos sigue una distribución binomial.
- Posterior (a posteriori): combinamos prior y likelihood. En este caso, la posterior es una Beta(1+caras, 1+sellos).



Distribuciones discretas univariadas

Nombre	Parámetros	$p(x)$	Uso
Uniforme	$N > 0$, entero	$\frac{1}{N}$, $x = 1, \ldots, N$	Fenómenos acotados cuyo comportamiento no se entiende.
Binomial	$p \in (0,1)$, $n \in \mathbb{N}_{\geq 0}$	$\binom{n}{x} p^x (1-p)^{n-x}$ $x \in \{0, 1 \ldots, n\}$	Probabilidad de x éxitos y $n - x$ fracasos.
Poisson	$\lambda > 0$	$\frac{e^{-\lambda}\lambda^x}{x!}$, $x \in \mathbb{N}_{\geq 0}$	Conteo de ocurrencias de un evento durante un periodo de tiempo.
Geométrica	$p \in (0,1)$	$(1-p)^x p$, $x \in \mathbb{N}_{\geq 0}$	Número de fracasos hasta el primer éxito.
Binomial negativa	$r > 0$, $p \in (0,1)$	$\binom{r+x-1}{x} p^r (1-p)^x$, $x \in \mathbb{N}_{\geq 0}$	Número de fracasos hasta el r-ésimo éxito.

Distribuciones continuas univariadas

Nombre	Parámetros	$f_X(x)$	Uso
Uniforme	$a < b \in \mathbb{R}$	$\frac{1}{b-a}\mathbb{I}_{(a,b)}$	Fenómenos acotados cuyo comportamiento no se entiende.
Normal	$\mu, \sigma^2 \geq 0$	$\frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-(x-\mu)^2}{2\sigma^2}}$	Promedios y errores. Hipótesis en muchas técnicas estadísticas y para aproximar otras distribuciones.
Exponencial	$\frac{1}{\lambda} = \theta > 0$	$\lambda e^{-\lambda x}$, $x > 0$	Tiempo entre eventos. Relacionada con la distribución Poisson.
Ji cuadrado	$k > 0$, grados de libertad	$\frac{(\frac{1}{2})^{\frac{k}{2}}}{\Gamma(\frac{k}{2})} x^{\frac{k}{2}-1} e^{-\frac{x}{2}}$ $x > 0$	Distribución de la varianza de una normal. Suma de cuadrados de normales.
Gamma	$\alpha > 0$, $\beta > 0$	$\frac{\beta^\alpha x^{\alpha-1} e^{-\beta x}}{\Gamma(\alpha)}$, $x > 0$	Generaliza otras distribuciones (exponencial y Ji-cuadrado).

Equivalencias

- $\mathcal{U}(0,1) \equiv \beta(1,1)$
- $\mathcal{E}(\lambda) \equiv \Gamma(1,\lambda) \equiv \mathrm{Wei}(1,\frac{1}{\lambda})$
- $\mathcal{E}(\frac{1}{2}) \equiv \chi^2_2 \equiv \Gamma(1,\frac{1}{2})$
- $\chi^2_k \equiv \Gamma(\frac{k}{2},\frac{1}{2})$ con $k \in \mathbb{N}$