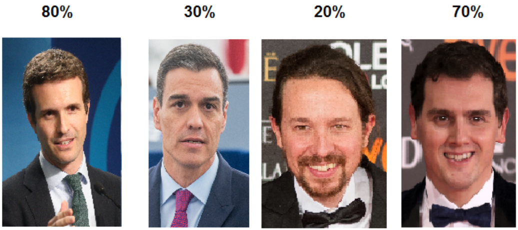
# Inferencia bayesiana

Mike Wiper
Departamento de Estadística
Universidad Carlos III de Madrid

# Objetivo

Introducir las ideas básicas de la inferencia bayesiana y las diferencias y similitudes con la inferencia clásica.

# Probabilidad subjetiva




La inferencia bayesiana se basa en la probabilidad subjetiva. Cada individuo puede tener su propia probabilidad para cualquier suceso, basada en sus propios conocimientos, experiencias, ...

¿Cuál es la probabilidad de que naciera en el año 1962?

# Probabilidad subjetiva

80% 30% 20% 70%

La inferencia bayesiana se basa en la probabilidad subjetiva. Cada individuo puede tener su propia probabilidad para cualquier suceso, basada en sus propios conocimientos, experiencias, ...

¿Cuál es la probabilidad de que naciera en el año 1962?

El único requerimiento es que seamos *coherentes*.

# Ejemplo

"Pedro es un hombre soltero de mediana edad. Hizo la carrera de matemáticas en la universidad y le gustan los sudokus. Viste traje en su trabajo y sus amigos le consideran inteligente pero algo aburrido y demasiado serio."

Ordenar los siguientes sucesos en orden du su probabilidad:

- Toca jazz
- Es contable
- Lee un periodico deportivo
- Es contable y toca jazz
- Juega al padel
- Tiene cuenta en Facebook
- Juega al padel y tiene cuenta en Facebook

# Ejemplo

¿El porcentaje de paises africanos en la ONU es por arriba o por debajo de 10 %?

# Ejemplo

¿El porcentaje de paises africanos en la ONU es por arriba o por debajo de 10 %?

Estimar el porcentaje.

# Ejemplo

¿El porcentaje de paises africanos en la ONU es por arriba o por debajo de 10 %?

Estimar el porcentaje.

¿El porcentaje de paises de la UE que utilizan el euro es por arriba o por debajo de 30 %?

# Ejemplo

¿El porcentaje de paises africanos en la ONU es por arriba o por debajo de 10 %?

Estimar el porcentaje.

¿El porcentaje de paises de la UE que utilizan el euro es por arriba o por debajo de 30 %?

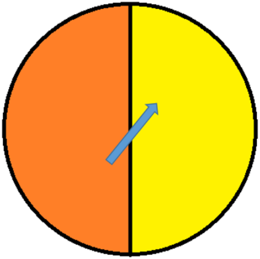
Estimar el porcentaje.

# Elicitación de probabilidades

Normalmente, los expertos a quienes queremos consultar no son estadísticos.

No saben casi nada sobre probabilidad, parámetros ...

Tenemos que preguntarles sobre cosas reales.



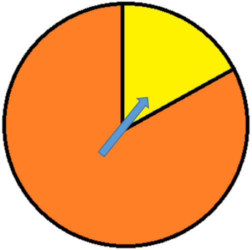
¿Qué apuesta prefieres?
Ganas 1000€ si gana Real Madrid la liga
o
Ganas 1000€ si sale amarillo

# Elicitación de probabilidades

Normalmente, los expertos a quienes queremos consultar no son estadísticos.

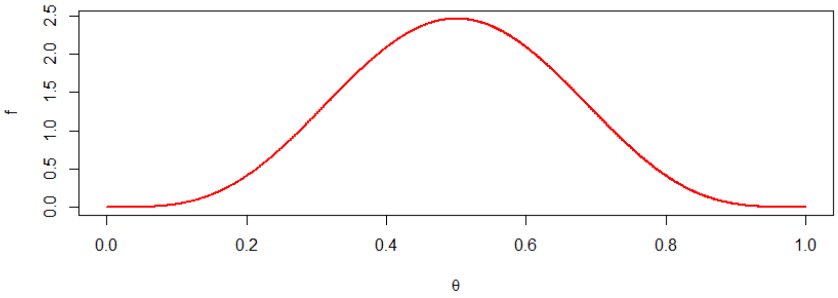
No saben casi nada sobre probabilidad, parámetros ...

Tenemos que preguntarles sobre cosas reales.



Ajustando las áreas, la probabilidad es proporcional al área cuando seas indiferente entre las dos apuestas.

# La distribución a priori




En contraste a la inferencia clásica, para un bayesiano, el parámetro, $\theta$ siempre es una variable.

Se supone que (antes de hacer cualquier experimento) una persona puede representar los conocimientos sobre $\theta$ con una distribución $f(\theta)$.

Esta distribución es la *distribución a priori*.

# Una distribución a priori para $\theta = P(\text{cruz})$

Supongamos que queremos pedir una distribución a priori de un experto para $\theta = P(\text{cruz})$.

¿Qué sabemos sobre la probabilidad?

- $0 \leq \theta \leq 1$.
- En la mayoría de experimentos, aproximadamente 50 % de los resultados son cruces.

Buscamos una distribución a priori con soporte $[0, 1]$ y centrado en 0,5.

# La distribución beta

Una variable continua $Y$ tiene una distribución beta con parámetros $\alpha, \beta > 0$ si

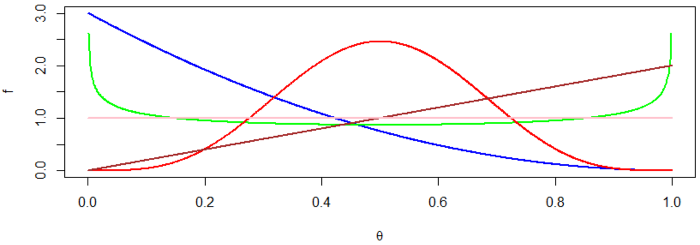
$$f_Y(y) = \frac{1}{B(\alpha, \beta)} y^{\alpha-1}(1-y)^{\beta-1} \quad \text{para } 0 < y < 1.$$

$B(\cdot, \cdot)$ es el coeficiente beta

$$B(\alpha, \beta) = \int_0^1 y^{\alpha-1}(1-y)^{\beta-1}\, dy = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}.$$

Se tiene $E[Y] = \frac{\alpha}{\alpha+\beta}$ y $V[Y] = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$.

# Distribuciones beta




$\text{Beta}\left(\frac{1}{2}, \frac{1}{2}\right)$

$\text{Beta}(1, 1)$

$\text{Beta}(5, 5)$

$\text{Beta}(1, 3)$

$\text{Beta}(2, 1)$

Cambiando los parámetros, la distribución beta toma formas muy distintas.

En nuestro caso, podemos utilizar la distribución $\text{Beta}(5, 5)$.

# La distribución a posteriori

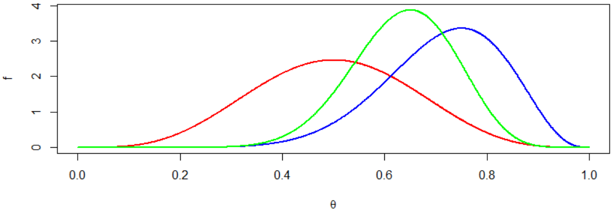
Dada una muestra de datos, modificamos nuestros creencias sobre $\theta$ mediante el teorema de Bayes:

$$f(\theta|\text{datos}) = \frac{f(\theta)f(\text{datos}|\theta)}{f(\text{datos}).}$$

En esta expresión:

- $f(\theta|\text{datos})$ es la distribución a posteriori,
- $f(\theta)$ es la distribución a priori,
- $f(\text{datos}|\theta) = l(\theta|\text{datos})$ es la función de verosimilitud,
- $f(\text{datos}) = \int f(\text{datos}|\theta)f(\theta)\,d\theta$ es la verosimilitud marginal.

# Simplificando el cálculo




En la fórmula anterior, observamos que el denominador no es una función de $\theta$. Luego, tenemos:

$$\begin{array}{rcl} f(\theta|\text{datos}) & \propto & f(\theta)f(\text{datos}|\theta) \\ \text{a posteriori} & \propto & \text{a priori} \times \text{verosimilitud} \end{array}$$

# Ejemplo

Si tiramos la moneda 12 veces y observamos 9 cruces y 3 caras, entonces:

$$
\begin{aligned}
f(\theta|\text{datos}) &\propto \frac{1}{B(5,5)}\theta^{5-1}(1-\theta)^{5-1}\left(\begin{array}{c}12\\9\end{array}\right)\theta^{9}(1-\theta)^{3} \\
&\propto \theta^{14-1}(1-\theta)^{8-1}
\end{aligned}
$$

¿Qué distribución es?

# Ejemplo

Si tiramos la moneda 12 veces y observamos 9 cruces y 3 caras, entonces:

$$\begin{aligned} f(\theta|\text{datos}) &\propto \frac{1}{B(5,5)}\theta^{5-1}(1-\theta)^{5-1}\left(\begin{array}{c}12\\9\end{array}\right)\theta^{9}(1-\theta)^{3} \\ &\propto \theta^{14-1}(1-\theta)^{8-1} \end{aligned}$$

¿Qué distribución es?

$$f(\theta|\text{datos}) = \frac{1}{B(14,8)}\theta^{14-1}(1-\theta)^{8-1}.$$

Tenemos $\theta|\text{datos} \sim \text{Beta}(14,8)$.

# Ejemplo

Si tiramos la moneda 12 veces y observamos 9 cruces y 3 caras, entonces:

$$\begin{aligned} f(\theta|\text{datos}) &\propto \frac{1}{B(5,5)}\theta^{5-1}(1-\theta)^{5-1}\left(\begin{array}{c}12\\9\end{array}\right)\theta^{9}(1-\theta)^{3} \\ &\propto \theta^{14-1}(1-\theta)^{8-1} \end{aligned}$$

¿Qué distribución es?

$$f(\theta|\text{datos}) = \frac{1}{B(14,8)}\theta^{14-1}(1-\theta)^{8-1}.$$

Tenemos $\theta|\text{datos} \sim \text{Beta}(14,8)$.

¿Sería diferente si el diseño del experimento fuese binomial negativa?

# Estimación puntual

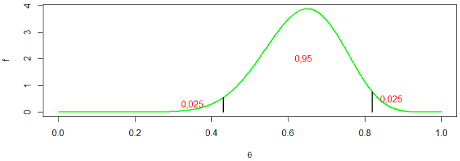
Cuando hemos calculado la distribución a posteriori, podemos utilizar la media a posteriori, la mediana a posteriori o la moda a posteriori como estimadores puntuales de $\theta$.

Este último caso es conocido como el *estimador máximo a posteriori* o MAP.

En nuestro ejemplo:

$$E[\theta|\text{datos}] = 0{,}636 \quad \text{Med}[\theta|\text{datos}] = 0{,}641 \quad \text{Mo}[\theta|\text{datos}] = 0{,}650$$

# Estimación por intervalos

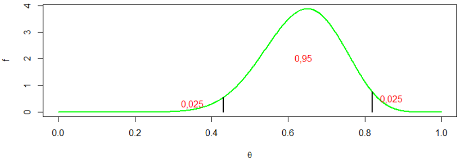



Un *intervalo de* $100(1-\alpha)\,\%$ *de credibilidad* es cualquier intervalo $(a, b)$ tal que $P\,(a < \theta < b|\text{datos}) = 1 - \alpha$.

En nuestro ejemplo, un intervalo de 95 % de credibilidad para $\theta$ es (0,430, 0,819).

¿Cómo interpretamos el intervalo?

# Estimación por intervalos




Un *intervalo de* $100(1-\alpha)\,\%$ *de credibilidad* es cualquier intervalo $(a, b)$ tal que $P\,(a < \theta < b|\text{datos}) = 1 - \alpha$.

En nuestro ejemplo, un intervalo de 95 % de credibilidad para $\theta$ es (0,430, 0,819).

¿Cómo interpretamos el intervalo?

El intervalo de credibilidad más corta se llama el *intervalo de máxima densidad a posteriori*.

En el ejemplo, el intervalo de máxima densidad a posteriori es (0,439, 0,826).

# Contrastes y selección de modelos

En teoría, hacer contrastes a la bayesiana es fácil. Dadas las distribuciones a priori, $P(H_0)$ y $P(H_1) = 1 - P(H_0)$, se calculan las probabilidades a posteriori con:

$$P(H_0|\text{datos}) \propto f(\text{datos}|H_0)P(H_0) \propto P(H_0) \int f(\text{datos}|H_0,\theta)f(\theta|H_0)\,d\theta$$

Luego, dada una *regla de decisión*, se puede decidir rechazar $H_0$ o no.

En principio, podemos utilizar las mismas técnicas para seleccionar modelos: fijadas las probabilidades a priori de cada modelo, calculamos las probabilidades a posteriori y seleccionar el modelo más probable.

# Ejemplo

Supongamos que queremos contrastar $H_0 : \theta \leq 0{,}5$ frente a $H_1 : \theta > 0{,}5$.

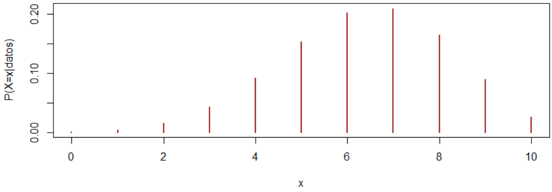
Dada la a priori Beta$(5, 5)$ empleada anteriormente, entonces $P(H_0) = \int_0^{0{,}5} f(\theta)\, d\theta = 0{,}5$ y luego, a posteriori,

$$P(H_0|\text{datos}) = \int_0^{0{,}5} \frac{1}{B(14,8)} \theta^{14-1}(1-\theta)^{8-1}\, d\theta = 0{,}0946.$$

Dada la regla de decisión de rechazar $H_0$ si su probabilidad a posteriori es menor que 0,05, no rechazariamos esta hipótesis.

¿Hay algún problema con este idea?

# Predicción




Supongamos que queremos predecir el valor de una nueva variable $Y$. Entonces,

$$f(y|\text{datos}) = \int f(y|\theta, \text{datos}) f(\theta|\text{datos})\, d\theta$$

y en muchas aplicaciones, $Y$ es condicionalmente independiente de los datos dado $\theta$, y luego:

$$f(y|\text{datos}) = \int f(y|\theta) f(\theta|\text{datos})\, d\theta.$$

# Ejemplo

Supongamos que queremos calcular la distribución del $X =$ el número de cruces en 10 más tiradas de la moneda.

Tenemos $X|\theta \sim \text{Binomial}(10, \theta)$, y $\theta|\text{datos} \sim \text{Beta}(14, 8)$. Entonces:

$$\begin{aligned} P(X = x|\text{datos}) &= \int_0^1 \binom{10}{x} \theta^x (1-\theta)^{10-x} \frac{1}{B(14,8)} \theta^{14-1}(1-\theta)^{8-1}\, d\theta \\ &= \binom{10}{x} \frac{1}{B(14,8)} \int_0^1 \theta^{14+x-1}(1-\theta)^{18-x-1}\, d\theta \end{aligned}$$

# Ejemplo

Supongamos que queremos calcular la distribución del $X$ = el número de cruces en 10 más tiradas de la moneda.

Tenemos $X|\theta \sim \text{Binomial}(10, \theta)$, y $\theta|\text{datos} \sim \text{Beta}(14, 8)$. Entonces:

$$\begin{aligned}
P(X = x|\text{datos}) &= \int_0^1 \binom{10}{x} \theta^x (1-\theta)^{10-x} \frac{1}{B(14,8)} \theta^{14-1}(1-\theta)^{8-1}\, d\theta \\
&= \binom{10}{x} \frac{1}{B(14,8)} \int_0^1 \theta^{14+x-1}(1-\theta)^{18-x-1}\, d\theta \\
&= \binom{10}{x} \frac{B(14+x, 18-x)}{B(14,8)} \int_0^1 \frac{1}{B(14+x,18-x)} \theta^{14+x-1}(1-\theta)^{18-x-1}\, d\theta \\
&= \binom{10}{x} \frac{B(14+x, 18-x)}{B(14,8)} \quad \text{para } x = 0, 1, \ldots, 10.
\end{aligned}$$

# Cálculo de la media de $X$

El cálculo directo a través de la distribución beta-binomial parece complicado. ¿Existe una manera más fácil?

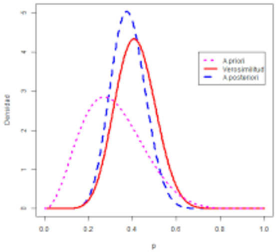
# Cálculo de la media de $X$

El cálculo directo a través de la distribución beta-binomial parece complicado. ¿Existe una manera más fácil?

$$\begin{aligned} E[X|\text{datos}] &= E[E[X|\theta]|\text{datos}] \\ &= E[10\theta|\text{datos}] \\ &= 10E[\theta|\text{datos}] \\ &= 10 \times \frac{14}{14+8} = 6{,}364. \end{aligned}$$

# Resumen y siguiente sesión

En esta clase, hemos introducido algunas de las ideas básicas de la inferencia bayesiana.




En la siguiente sesión, veremos que si utilizamos una distribución a priori beta en problemas semejantes al ejemplo de hoy, la inferencia Bayesiana es fácil de implementar.