

## Teoría de la Decisión

*Un meteorólogo estima, con un 95% de confianza, que la probabilidad de que el huracán no llegue a la ciudad está entre 99% y 100%. Muy feliz con su precisión y su modelo, aconseja que la evacuación de la ciudad no es necesaria. Desafortunadamente, el huracán llega a la ciudad produciendo una grave inundación.*

*Un meteorólogo estima, con un 95% de confianza, que la probabilidad de que el huracán no llegue a la ciudad está entre 99% y 100%. Muy feliz con su precisión y su modelo, aconseja que la evacuación de la ciudad no es necesaria. Desafortunadamente, el huracán llega a la ciudad produciendo una grave inundación.*

*"I would rather be vaguely right than very wrong."*

*Con el clima, las personas tienden a notar un error más que otro. Cuando llueve sin estar anunciado, se tiende a insultar al servicio meteorológico mientras que la ausencia de lluvia a pesar del pronóstico se toma con buena cara.*

*Con el clima, las personas tienden a notar un error más que otro. Cuando llueve sin estar anunciado, se tiende a insultar al servicio meteorológico mientras que la ausencia de lluvia a pesar del pronóstico se toma con buena cara. El Weather Channel exagera ligeramente la probabilidad de lluvia cuando es poco probable que ocurra: dicen que es de 20% cuando en realidad es de 5% o 10%*

En la estadística bayesiana, la distribución *a posteriori* es la base de todas inferencia: combina el conocimiento *a priori* con la información provista por los datos. Contiene todo lo que se sabe y no se sabe sobre un parámetro desconocido.

En la estadística bayesiana, la distribución *a posteriori* es la base de todas inferencia: combina el conocimiento *a priori* con la información provista por los datos. Contiene todo lo que se sabe y no se sabe sobre un parámetro desconocido.

La respuesta a los problemas es toda la distribución *a posteriori* de los parámetros (y de otras cantidades de interés).

No obstante, puede ser de utilidad (o incluso necesario) tomar decisiones concretas o resumir la distribución *a posteriori*.



No obstante, puede ser de utilidad (o incluso necesario) tomar decisiones concretas o resumir la distribución *a posteriori*.

$a$  es la acción que tomamos (intervenir o no intervenir quirúrgicamente a una persona) o la respuesta que damos (ganancia de una campaña de marketing).

No obstante, puede ser de utilidad (o incluso necesario) tomar decisiones concretas o resumir la distribución *a posteriori*.

$a$  es la acción que tomamos (intervenir o no intervenir quirúrgicamente a una persona) o la respuesta que damos (ganancia de una campaña de marketing).

Puede ser una estimación puntual  $\hat{\theta}$ : dada una inferencia sobre la ganancia de una campaña de marketing, es necesario informar un valor puntual (quizás con un intervalo)

Tratamos a los parámetros sobre los que realizamos inferencias como variables aleatorias. Una muestra de la distribución a *posteriori* es una posible realización del verdadero valor del parámetro.

Tratamos a los parámetros sobre los que realizamos inferencias como variables aleatorias. Una muestra de la distribución *a posteriori* es una posible realización del verdadero valor del parámetro.

Al dar una respuesta (o resumir la información *a posteriori*), podemos incurrir en un error (grande o chico) según se den los eventos posibles.

Tratamos a los parámetros sobre los que realizamos inferencias como variables aleatorias. Una muestra de la distribución *a posteriori* es una posible realización del verdadero valor del parámetro.

Al dar una respuesta (o resumir la información *a posteriori*), podemos incurrir en un error (grande o chico) según se den los eventos posibles.

¿Qué es un error? ¿Cómo definimos si el error es grande o chico?  
¿Cómo definimos si el error es relevante o no?

# Funciones de pérdida

$$L(\theta, \hat{\theta}) = f(\theta, \hat{\theta})$$

es una **función de pérdida**, qué tanto pierdo por usar  $\hat{\theta}$  para estimar  $\theta$ .

Por ejemplo:

$$L_2 = (\theta - \hat{\theta})^2$$

$$L_1 = |\theta - \hat{\theta}|$$

$$L_{0/1} = \begin{cases} 0 & \text{si } \hat{\theta} = \theta \\ 1 & \text{si } \hat{\theta} \neq \theta \end{cases}$$

$$L(\theta, \hat{\theta}) = \begin{cases} (\theta - \hat{\theta})^2 & \hat{\theta} < \theta \\ c(\theta - \hat{\theta})^2 & \hat{\theta} \geq \theta, \quad 0 < c < 1 \end{cases}$$

Buscamos elegir  $\hat{\theta}$  de manera tal que minimice  $L$ . El problema es que no conocemos  $\theta$  y por lo tanto no podemos calcular  $L(\theta, \hat{\theta})$ .



Buscamos elegir  $\hat{\theta}$  de manera tal que minimice  $L$ . El problema es que no conocemos  $\theta$  y por lo tanto no podemos calcular  $L(\theta, \hat{\theta})$ .

¿Sabemos algo sobre  $\theta$  que nos pueda ayudar? Conocemos su distribución *a posteriori*

Buscamos elegir  $\hat{\theta}$  de manera tal que minimice  $L$ . El problema es que no conocemos  $\theta$  y por lo tanto no podemos calcular  $L(\theta, \hat{\theta})$ .

¿Sabemos algo sobre  $\theta$  que nos pueda ayudar? Conocemos su distribución *a posteriori*

Podemos promediar  $L$  para los valores posibles de  $\theta$  (ponderando según la distribución *a posteriori*)

El riesgo *a posteriori* (*posterior risk* o *posterior expected loss*) es la pérdida esperada ponderada por los valores de  $\theta$  (y su distribución *a posteriori*).

El riesgo *a posteriori* (*posterior risk* o *posterior expected loss*) es la pérdida esperada ponderada por los valores de  $\theta$  (y su distribución *a posteriori*).

$$R(\hat{\theta}) = \mathbb{E}_{\theta|y}[L(\theta, \hat{\theta})] = \int L(\theta, \hat{\theta})p(\theta | y)d\theta$$

El riesgo *a posteriori* (*posterior risk* o *posterior expected loss*) es la pérdida esperada ponderada por los valores de  $\theta$  (y su distribución *a posteriori*).

$$R(\hat{\theta}) = \mathbb{E}_{\theta|y}[L(\theta, \hat{\theta})] = \int L(\theta, \hat{\theta})p(\theta | y)d\theta$$

Es una función de los posibles valores que puede tomar  $\hat{\theta}$ .

El riesgo *a posteriori* (*posterior risk* o *posterior expected loss*) es la pérdida esperada ponderada por los valores de  $\theta$  (y su distribución *a posteriori*).

$$R(\hat{\theta}) = \mathbb{E}_{\theta|y}[L(\theta, \hat{\theta})] = \int L(\theta, \hat{\theta})p(\theta | y)d\theta$$

Es una función de los posibles valores que puede tomar  $\hat{\theta}$ .

Podemos obtener el  $\hat{\theta}$  que minimice  $R(\hat{\theta})$ . Es decir, buscamos un valor (un estimador) que minimice la pérdida esperada al usarlo para resumir  $p(\theta | y)$ :

$$\hat{\theta} = \arg \min_{\hat{\theta}} [R(\hat{\theta})]$$

Simulemos...

Simulemos...

Supongamos que  $\theta \mid y \sim \text{Beta}(2, 9)$



Simulemos...

Supongamos que  $\theta \mid y \sim \text{Beta}(2, 9)$

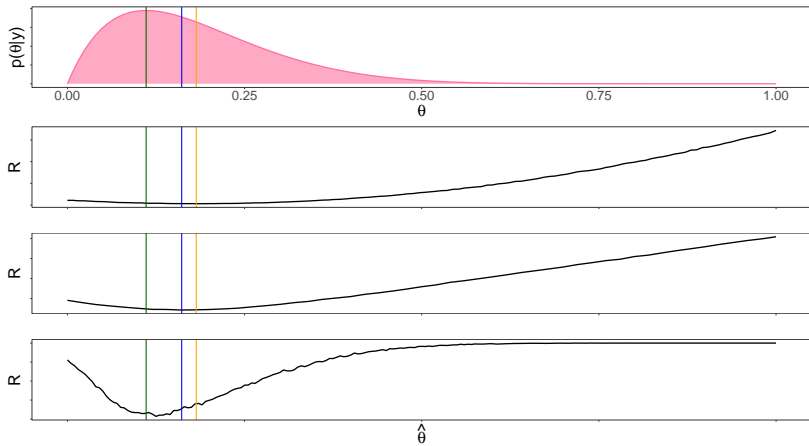
- ▶  $R$  es una función de  $\hat{\theta}$  (los distintos valores que podemos usar para resumir  $p(\theta \mid y)$ )
- ▶ Para distintos valores de  $\hat{\theta}$  voy a tomar muestras de  $p(\theta \mid y)$  y calcular la pérdida  $L$
- ▶ Para cada valor de  $\hat{\theta}$  voy a calcular la pérdida promedio (ya va a estar ponderada por la probabilidad *a posteriori* de  $\theta$ )

```
L_2 <- function(theta,theta_hat) (theta-theta_hat)^2

loss <- data.frame(theta_hat = double(),
                    theta = double(),
                    L = double())

for(theta_hat in seq(0,1,0.008)){
  theta <- rbeta(2000, shape1 = 2, shape2 = 9)
  L <- L_2(theta,theta_hat)
  loss <- bind_rows(loss,data.frame(theta_hat,theta,L))
}

expected.loss <- loss |>
  group_by(theta_hat) |>
  summarise(loss.mean = mean(L))
```



Si deseamos resumir la distribución *a posteriori* con un único valor (¡perdiendo información!), puede usarse:

Si deseamos resumir la distribución *a posteriori* con un único valor (¡perdiendo información!), puede usarse:

- ▶ La **media**: minimiza la pérdida cuadrática esperada *a posteriori*
- ▶ La **mediana**: minimiza la pérdida absoluta esperada *a posteriori*
- ▶ La **moda** (también llamado MAP por *maximum a posteriori* o estimador generalizado de máxima verosimilitud): minimiza la pérdida 0/1 esperada *a posteriori*

Una prueba más formal para el caso de la **media**...

Una prueba más formal para el caso de la **media**...

Sea la pérdida cuadrática  $L(\theta, \hat{\theta}) = (\theta - \hat{\theta})^2$ , el riesgo (*posterior expected loss*) es:

Una prueba más formal para el caso de la **media**...

Sea la pérdida cuadrática  $L(\theta, \hat{\theta}) = (\theta - \hat{\theta})^2$ , el riesgo (*posterior expected loss*) es:

$$\mathbb{E}_{\theta|y}[L(\theta, \hat{\theta})] = \mathbb{E}_{\theta|y}[\theta^2] - 2\hat{\theta}\mathbb{E}_{\theta|y}[\theta] + \hat{\theta}^2$$



Una prueba más formal para el caso de la **media**...

Sea la pérdida cuadrática  $L(\theta, \hat{\theta}) = (\theta - \hat{\theta})^2$ , el riesgo (*posterior expected loss*) es:

$$\mathbb{E}_{\theta|y}[L(\theta, \hat{\theta})] = \mathbb{E}_{\theta|y}[\theta^2] - 2\hat{\theta}\mathbb{E}_{\theta|y}[\theta] + \hat{\theta}^2$$

derivando respecto a  $\hat{\theta}$  e igualando a cero se obtiene que  
 $\arg \min_{\hat{\theta}} [R(\hat{\theta})] = \mathbb{E}_{\theta|y}[\theta] = \mathbb{E}[p(\theta | y)]$

Una prueba más formal para el caso de la **mediana**...

Una prueba más formal para el caso de la **mediana**...

Sea la pérdida absoluta  $L(\theta, \hat{\theta}) = |\theta - \hat{\theta}|$ , el riesgo (*posterior expected loss*) es:

Una prueba más formal para el caso de la **mediana**...

Sea la pérdida absoluta  $L(\theta, \hat{\theta}) = |\theta - \hat{\theta}|$ , el riesgo (*posterior expected loss*) es:

$$\mathbb{E}_{\theta|y}[L(\theta, \hat{\theta})] = \int_{-\infty}^{\infty} |\theta - \hat{\theta}| p(\theta | y) d\theta$$

Una prueba más formal para el caso de la **mediana**...

Sea la pérdida absoluta  $L(\theta, \hat{\theta}) = |\theta - \hat{\theta}|$ , el riesgo (*posterior expected loss*) es:

$$\mathbb{E}_{\theta|y}[L(\theta, \hat{\theta})] = \int_{-\infty}^{\infty} |\theta - \hat{\theta}| p(\theta | y) d\theta$$

$$\int_{-\infty}^{\hat{\theta}} (\hat{\theta} - \theta) p(\theta | y) d\theta + \int_{\hat{\theta}}^{\infty} (\theta - \hat{\theta}) p(\theta | y) d\theta$$

Una prueba más formal para el caso de la **mediana**...

Sea la pérdida absoluta  $L(\theta, \hat{\theta}) = |\theta - \hat{\theta}|$ , el riesgo (*posterior expected loss*) es:

$$\mathbb{E}_{\theta|y}[L(\theta, \hat{\theta})] = \int_{-\infty}^{\infty} |\theta - \hat{\theta}| p(\theta | y) d\theta$$

$$\int_{-\infty}^{\hat{\theta}} (\hat{\theta} - \theta) p(\theta | y) d\theta + \int_{\hat{\theta}}^{\infty} (\theta - \hat{\theta}) p(\theta | y) d\theta$$

Para derivar, se utiliza la regla integral de Leibniz:

Una prueba más formal para el caso de la **mediana**...

Sea la pérdida absoluta  $L(\theta, \hat{\theta}) = |\theta - \hat{\theta}|$ , el riesgo (*posterior expected loss*) es:

$$\mathbb{E}_{\theta|y}[L(\theta, \hat{\theta})] = \int_{-\infty}^{\infty} |\theta - \hat{\theta}| p(\theta | y) d\theta$$

$$\int_{-\infty}^{\hat{\theta}} (\hat{\theta} - \theta) p(\theta | y) d\theta + \int_{\hat{\theta}}^{\infty} (\theta - \hat{\theta}) p(\theta | y) d\theta$$

Para derivar, se utiliza la regla integral de Leibniz:

$$\frac{d}{d\hat{\theta}} \int_{-\infty}^{\hat{\theta}} g(\hat{\theta}, \theta) d\theta = g(\hat{\theta}, \hat{\theta}) + \int_{-\infty}^{\hat{\theta}} \frac{\partial}{\partial \hat{\theta}} g(\hat{\theta}, \theta) d\theta$$

Una prueba más formal para el caso de la **mediana**...

Sea la pérdida absoluta  $L(\theta, \hat{\theta}) = |\theta - \hat{\theta}|$ , el riesgo (*posterior expected loss*) es:

$$\mathbb{E}_{\theta|y}[L(\theta, \hat{\theta})] = \int_{-\infty}^{\infty} |\theta - \hat{\theta}| p(\theta | y) d\theta$$

$$\int_{-\infty}^{\hat{\theta}} (\hat{\theta} - \theta) p(\theta | y) d\theta + \int_{\hat{\theta}}^{\infty} (\theta - \hat{\theta}) p(\theta | y) d\theta$$

Para derivar, se utiliza la regla integral de Leibniz:

$$\frac{d}{d\hat{\theta}} \int_{-\infty}^{\hat{\theta}} g(\hat{\theta}, \theta) d\theta = g(\hat{\theta}, \hat{\theta}) + \int_{-\infty}^{\hat{\theta}} \frac{\partial}{\partial \hat{\theta}} g(\hat{\theta}, \theta) d\theta$$

Se puede probar que  $\int_{-\infty}^{\hat{\theta}} p(\theta | y) d\theta = \frac{1}{2}$ , por lo que el  $\hat{\theta}$  que minimiza la expresión es la mediana.



# Intervalos de Credibilidad

También llamados: intervalos de probabilidad, intervalo de confianza bayesiano, región de credibilidad. Es una región del dominio del parámetro que tiene alta probabilidad de contenerlo. Se utiliza para resumir el posterior.

# Intervalos de Credibilidad

También llamados: intervalos de probabilidad, intervalo de confianza bayesiano, región de credibilidad. Es una región del dominio del parámetro que tiene alta probabilidad de contenerlo. Se utiliza para resumir el posterior.

Un intervalo de credibilidad es una región  $C$  tal que la probabilidad de que contenga al parámetro sea al menos  $1 - \alpha$ :

# Intervalos de Credibilidad

También llamados: intervalos de probabilidad, intervalo de confianza bayesiano, región de credibilidad. Es una región del dominio del parámetro que tiene alta probabilidad de contenerlo. Se utiliza para resumir el posterior.

Un intervalo de credibilidad es una región  $C$  tal que la probabilidad de que contenga al parámetro sea al menos  $1 - \alpha$ :

$$p(\theta \in C \mid y) = \int_C p(\theta \mid y) d\theta = 1 - \alpha$$

en el caso discreto es  $(\geq 1 - \alpha)$

# Intervalos de Credibilidad

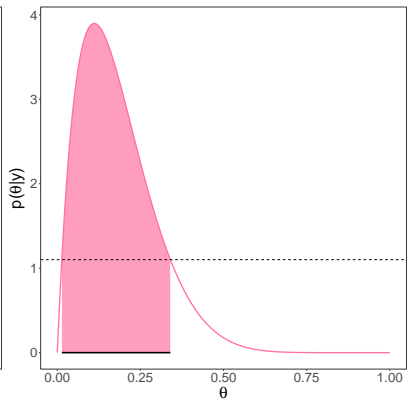
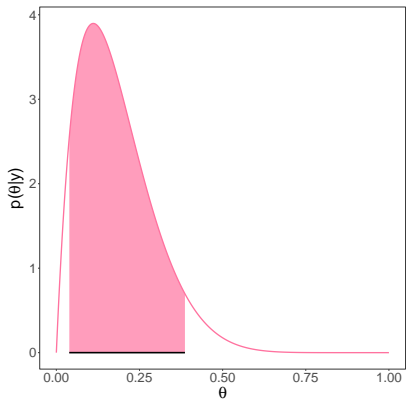
También llamados: intervalos de probabilidad, intervalo de confianza bayesiano, región de credibilidad. Es una región del dominio del parámetro que tiene alta probabilidad de contenerlo. Se utiliza para resumir el posterior.

Un intervalo de credibilidad es una región  $C$  tal que la probabilidad de que contenga al parámetro sea al menos  $1 - \alpha$ :

$$p(\theta \in C \mid y) = \int_C p(\theta \mid y) d\theta = 1 - \alpha$$

en el caso discreto es  $(\geq 1 - \alpha)$

Decimos: *la probabilidad de que  $\theta$  esté contenido en  $C$ , dados los datos (y el modelo) es de  $1 - \alpha$*



En el análisis de datos bayesiano, es habitual resumir los hallazgos reportando:

En el análisis de datos bayesiano, es habitual resumir los hallazgos reportando:

- ▶ Un gráfico de la distribución *a posteriori*
- ▶ Algún medida de centralidad de la distribución *a posteriori*
- ▶ Percentiles relevantes de la distribución *a posteriori*
- ▶ Probabilidades *a posteriori* de interés  $p(\theta > c \mid y)$  para algún  $c$  interesante, por ejemplo  $c = 0$

# Simulaciones

Para interpretar los resultados de la inferencia bayesiana podemos simplemente realizar simulaciones a partir del *posterior* y estimar probabilidades contando.



# Simulaciones

Para interpretar los resultados de la inferencia bayesiana podemos simplemente realizar simulaciones a partir del *posterior* y estimar probabilidades contando.

- ▶ los parámetros
- ▶ funciones de los parámetros
- ▶ la variable respuesta (predicciones)

Un nuevo ejemplo del modelo beta–Binomial: partimos de  $\text{Beta}(2, 2)$ , observamos 4 caras en 6 tiradas y nuestra creencia *a posteriori* pasa a ser  $\text{Beta}(6, 4)$ .

## Parámetros

¿Cuál es la probabilidad *a posteriori* de que  $\pi$  sea mayor a 0.50? ¿y de que sea mayor a 0.80?

```
muestras_pi <- rbeta(2000,6,4)
mean(muestras_pi > 0.5)
mean(muestras_pi > 0.8)
```

## Funciones de los parámetros

¿Cuál es la distribución *a posteriori* de la chance de obtener cara  $\frac{\pi}{1-\pi}$ ?

```
muestras_pi <- rbeta(2000,6,4)
muestras_odds <- muestras_pi/(1-muestras_pi)
```

## Predicciones

Si se arroja la moneda 11 veces más ¿cuál es la distribución de probabilidad de la cantidad de caras? (es la distribución predictiva *a posteriori*)

```
muestras_pi <- rbeta(2000,6,4)
y_new <- rbinom(2000,11,muestras_pi)
```

Toda cantidad que dependa de los parámetros tiene una distribución *a posteriori*: una incertidumbre asociada.