Introducción a la estadística Bayesiana con aplicaciones de estimación en áreas pequeñas usando software STAN

Ignacio Alvarez-Castro   Juan José Goyeneche

Instituto de Estadística, Facultad de Ciencias Económicas y Administración, UdelaR.

XV Congreso Latinoamericano de Sociedades de Estadística
9 al 13 de Octubre 2023
Santiago de Cali, Colombia

Probabilidad

Es usual utilizar *probabilidades* de manera informal para medir la incertidumbre o información sobre un fenómeno de interés.

Puede hacerse de manera formal

- la probabilidad puede representar creeencias *racionales*
- hay una relación entre probabilidad e información
- regla de Bayes es un método racional para actualizar la probabilidad cuando obtenemos *información nueva*

Muchas formas de interpretar una probabilidad

Todas comparten las propiedades matemáticas

Los procedimientos estadísticos que utilicen la probabilidad como medida de incertidumbre y la regla de Bayes para actualizarla, se denominan métodos Bayesianos.

Elegir estos métodos en un problema concreto atiende a:

- propiedades de estimadores
- descripción de datos observados
- predicciones de datos faltantes o futuros
- marco computacional para estimar, seleccionar y validar

La inferencia estadística se ocupa de elaborar métodos para estimar características generales de un fenómeno de interés a partir de una cantidad finita de datos observados

Datos	$y = (y_1, y_2, \ldots, y_n)$	observaciones del fenómeno bajo estudio
Parámetros	θ	características relevantes

Objetivos:

- *Explicar* características relevantes de Y (estimar θ)
- *Predecir* el valor de observaciones futuras
- *Comparar* modelizaciones alternativas

Antes de realizar observaciones, tenemos incertidumbre sobre:
los datos que vamos a observar
y los parámetros del fenómeno de estudio.

Dos componentes básicos:

Modelo para los datos	$p(y\|\theta)$	incertidumbre sobre y
Previa	$p(\theta)$	incertidumbre sobre θ

Antes de realizar observaciones, tenemos incertidumbre sobre:
los datos que vamos a observar
y los parámetros del fenómeno de estudio.

Dos componentes básicos:

Modelo para los datos	$p(y\|\theta)$	incertidumbre sobre y
Previa	$p(\theta)$	incertidumbre sobre θ

Despues de observar y,
ya no hay incertidumbre sobre los datos observados,
actualizamos la incertidumbre sobre θ mediante la regla de Bayes:

$$p(\theta|y) = \frac{p(y|\theta)p(\theta)}{p(y)} = \frac{p(y|\theta)p(\theta)}{\int p(y|\theta)p(\theta)d\theta}$$

Probabilidad de un evento raro

Queremos estimar la prevalencia de una enfermedad *rara* en una ciudad. Para esto se obtiene una muestra de 25 personas para evaluar si estan infectadas o no.

Datos: Total de personas infectadas en la muestra	$y \in \{0, 1, \ldots, 25\}$
Parámetro: prevalencia en la población	$\theta \in (0, 1)$

Un posible modelo para los datos: $y|\theta \sim \textit{binomial}(25, \theta)$ te parece apropiado?

Aproximación frecuencista

El estimador por máxima verosimilitud es la proporción muestral, $\bar{y} = y/25$, y un intervalo de confianza se puede obtener como (intervalo de Wald):

$$\bar{y} \pm 1.96\sqrt{\frac{\bar{y}(1-\bar{y})}{25}}$$

Debido a que la enfermedad es rara, esta aproximación puede tener problemas. Por ejemplo, si observamos $y = 1$:

```
pr <- 1/25
se <- sqrt((pr*(1-pr))/25)
c(pr, pr - 1.96*se, pr+1.96*se) |> round(3)


## [1]  0.040 -0.037  0.117
```

Aproximación Bayesiana: previa

Para hacer estadística Bayesiana precisamos una previa $p(\theta)$.

Conocemos que en estudios anteriores, en otras ciudades la prevalencia de esta enfermedad varía entre $(0.05, 0.2)$ y en promedio es 0.10.

Proponemos una distribución que:

- verifique los datos anteriores
- sea conveniente computacionalmente

$$\theta \sim Beta(2, 20)$$

Aplicando la regla de Bayes obtenemos la posterior $p(\theta|y)$.

Aplicando la regla de Bayes obtenemos la posterior $p(\theta|y)$.

$$\theta|y \sim Beta(2 + y, 25 + 20 - y)$$

Si observamos $y = 1$

```
pr <- 3/47
c(pr, qbeta(.025, 3, 44), qbeta(.975, 3, 44)) |> round(3)

## [1] 0.064 0.014 0.148
```

y si observamos $y = 0$????????

Un modelo estadístico Bayesiano está formado por dos componentes:

Modelo para los datos	$p(y\|\theta)$
Previa	$p(\theta)$

La estimación de θ consiste en *hallar la distribución posterior* $p(\theta|y)$

Hallar $p(\theta|y)$ mediante la regla de Bayes:

$$p(\theta|y) = \frac{p(y|\theta)p(\theta)}{\int p(y|\theta)p(\theta)d\theta}$$

- analíticamente, en problemas simples
- numéricamente: aproximando en una grilla de valores de θ
- numéricamente: mediante simulaciones

Obtener posterior

Notar que $p(y)$ es constante respecto a θ

$$p(\theta|y) = \frac{p(y|\theta)p(\theta)}{p(y)} \propto p(y|\theta)p(\theta)$$

al obtener $p(\theta|y)$ podemos evitar calcular la constante $p(y)$,

$$\begin{aligned} p(\theta|y) &= \tfrac{p(y|\theta)p(\theta)}{p(y)} \\ &\propto p(y|\theta)p(\theta) \\ &\propto \theta^{n_1}(1-\theta)^{n-n_1} \text{ si } \theta \in (0,1) \end{aligned}$$

$\theta^{n_1}(1-\theta)^{n-n_1}$ es el núcleo de una distribución *Beta* $\Rightarrow$
$\theta|y \sim Beta(n_1+1, n-n_1+1)$

Si $X \sim Beta(\alpha, \beta)$ entonces, $p(x) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1}$ si $x \in (0,1)$
Posterior, $p(\theta|y)$:

$$
\begin{aligned}
p(\theta|y) &= \frac{p(y|\theta)p(\theta)}{p(y)} \\
&\propto p(y|\theta)p(\theta) \\
&= \binom{25}{y}\theta^y(1-\theta)^{n-y}\frac{\Gamma(2+20)}{\Gamma(2)\Gamma(20)}\theta^{2-1}(1-\theta)^{20-1}
\end{aligned}
$$

$\theta|y \sim Beta(y+2, n-y+20)$

Bayes SIN información previa

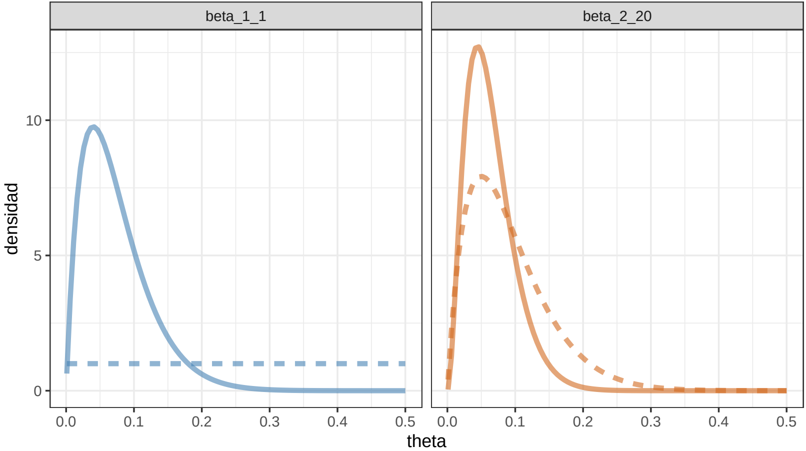
- No conocemos nada sobre la prevalencia
- Una previa que refleje la no-información
- Uniforme en el espacio paramétrico

$$\theta \sim Unif(0, 1)$$

En este caso, $Unif(0, 1)$ es también $Beta(1, 1)$, podemos suplantar en lo anterior para obtener:

$$\theta | y \sim Beta(y + 1, n - y + 1)$$

Previa y posterior



Estimar el parámetro de interés, θ, significa hallar su distribución posterior, $p(\theta|y)$.

Sin embargo, es mas práctico resumir la distribución

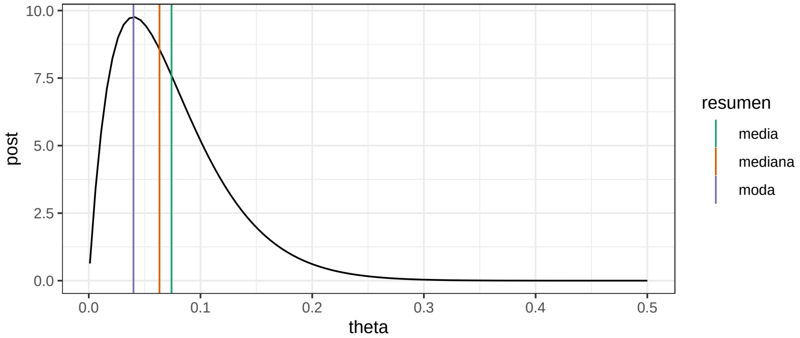
- Estimación puntual, $\hat{\theta}$.
- Estimación por intervalos.
- Probabilidades para eventos de interés.

Resumir la distribución posterior: Estimación puntual $\hat{\theta}$

Usualmente consideramos como una estimación puntual, una medida de localización de $p(\theta|y)$.

- Esperanza posterior: $E(\theta|y)$, minimiza el error cuadrático medio.
- Mediana posterior: $(\theta|y)_{0.5}$, es robusta y simple de calcular en base a valores simulados.
- Moda posterior: $MAP = \text{argmax}(p(\theta|y))$, es la versión Bayesiana de máxima verosimilitud.

En el ejemplo ...



Resumir la distribución posterior: Estimación puntual $\hat{\theta}$

En un contexto más general, se define como estimador de Bayes el valor que minimiza la pérdida esperada.

$$\hat{\theta}_{bayes} = \text{argmin}_{\theta} \left\{ \int L(\theta, \hat{\theta}) p(\theta|y) d\theta \right\}$$

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

- es una función de pérdida
- determina el estimador de Bayes