Inferencia Estadística

G5: Inferencia Bayesiana

Gabriel Martos — Email: gmartos@utdt.edu
Nicolás Ferrer — Email: nicolas.ferrer.747@gmail.com

Enunciados

1. Si $X|\theta \sim \text{Poiss}(\theta)$, y además podemos asumir que $P(\theta = 2) = 1/3$ y $P(\theta = 3) = 2/3$ (esto es $\Theta = \{2, 3\}$). Dada la información $x_1 = 2$ y $x_2 = 4$, compute la probabilidad a–posteriori[1] para θ. ¿Qué inferencia puede hacer respecto del parámetro θ?

2. Una de las ventajas del enfoque Bayesiano reside en que el teorema de Bayes se puede utilizar de forma secuencial; y esto es particularmente útil cuando necesitamos 'refrescar el modelo' con información nueva. Imagina que de manera secuencial recibes la información (datos) $\mathbf{x}_1$ y luego $\mathbf{x}_2$, argumenta porque es cierto que:

$$\pi(\theta|\mathbf{x}_1, \mathbf{x}_2) \propto L(\theta|\mathbf{x}_2)\pi(\theta|\mathbf{x}_1).$$

3. Si $X|\theta \sim \text{Poiss}(\theta)$ y $\theta \sim \text{Gamma}(\alpha, \beta)$[2], esto es:

$$\pi(\theta; \alpha, \beta) = \frac{\theta^{\alpha-1} e^{-\theta/\beta}}{\Gamma(\alpha)\beta^\alpha}, \quad \alpha > 0,\ \beta > 0,\ \theta > 0,$$

donde α y β son dos hiperparámetros conocidos (elegidos por quien modela el problema).

   (a) Comprobar que la distribución aposteriori de θ es Gamma de parámetros: $\alpha_n = \sum_{i=1}^n x_i + \alpha$ y $\beta_n = \beta/(n\beta + 1)$.

   (b) ¿A dónde convergen la media y la varianza a posteriori cuando $n \to \infty$?

   (c) Imagine que la variable aleatoria X_i da cuenta de la cantidad de delitos registrados en la ciudad en el día i; y que de una muestra de 10 días se tiene que $\sum_{i=1}^{10} x_i = 140$. Justificando su elección de los parámetros α y β (encuentre algún argumento razonable para elegirlos), obtenga la distribución a posteriori de θ.

   (d) Reporte la media y varianza a posteriori.

   (e) Construya un HPD al 95% y 99% e interprete los resultados.

4. El modelo Beta–Bernoulli, asume una prior Beta para el parámetro θ. Obtener la distribución a posteriori si en vez de una prior Beta utilizáramos una distribución uniforme en el intervalo $(0, 1)$ para θ:

   (a) ¿Cómo interpretas el uso de una prior uniforme en términos prácticos?

   (b) Calcula en este contexto $E(\theta\,|\,\mathbf{x})$ y $V(\theta\,|\,\mathbf{x})$.

[1] Recuerda que por probabilidad total: $P(X_1 = 2, X_2 = 4) = \sum_{\theta\in\Theta} P(X_1 = 2, X_2 = 4|\Theta = \theta)P(\Theta = \theta)$.
[2] Bajo esta parametrización la media a priori de θ es $\alpha\beta$ y la varianza $\alpha\beta^2$

5. En clase discutimos el modelo Normal–Normal. Utiliza la fórmula de Bayes y las propiedades de los modelos conjugados para construír de forma detallada la distribución a posteriori $\pi(\theta \,|\, \mathbf{x}) = N(\mu_n, \sigma_n^2)$, que recordemos tiene parámetros:

$$\mu_n = \frac{\frac{1}{\tau_0^2}\mu_0 + \frac{n}{\sigma_0^2}\bar{x}}{\frac{1}{\tau_0^2} + \frac{n}{\sigma_0^2}} \text{ y } \sigma_n^2 = \left(\frac{1}{\tau_0^2} + \frac{n}{\sigma_0^2}\right)^{-1}.$$

   (a) ¿A dónde convergen los parámetros de la posteriori cuando $n \to \infty$?

   (b) ¿Cómo interpretas este resultado?

   (c) Determina la estructura que tendría un intervalo de confianza creíble a posteriori de probabilidad 0.95. ¿Es tu intervalo el HPD?

6. Imagina que trabajas para la consultora económica XYZ y se te encarga hacer inferencia bayesiana para el parámetro θ = "tasa de desempleo en CABA". Tomas una muestra de tamaño $n = 100$ de la población relevante y observas que la variable y = Número de desempleados en la muestra = 18. Se pide respondas a lo siguiente:

   (a) ¿Cómo propondrías elegir la prior sobre θ?

   (b) Computa la distribución a–posteriori (para tu elección de prior).

   (c) Computa la esperanza, moda y varianza aposteriori de θ.

   (d) Computa la HPD para $\alpha = 5\%$.

   (e) Un economista amigo, con una visión diametralmente opuesta a la tuya en cuanto a la situación económica actual, presenta estimaciones diferentes utilizando los mismos datos de la encuesta anterior. ¿Cómo es esto posible?

   (f) ¿Qué crees que ocurriría con la "distancia" entre tus conclusiones y la de tu amigo economista si el tamaño de la muestra fuera 10 veces más grande?

Modelos conjugados

Likelihood	Prior distribution	Posterior distribution
$y_i \sim \text{binomial}\,(n, \phi)$	$\phi \sim \text{beta}\,(\alpha, \beta)$	$\phi \sim \text{beta}\,(\sum y_i + \alpha, n - \sum y_i + \beta)$
$y_i \sim \text{Bernoulli}(\phi)$	$\phi \sim \text{beta}\,(\alpha, \beta)$	$\phi \sim \text{beta}\,(\sum_{i=1}^n y_i + \alpha, \sum_{i=1}^n (1 - y_i) + \beta)$
$y_i \sim \text{Poisson}\,(\lambda)$	$\lambda \sim \text{gamma}\,(\alpha, \beta)$	$\lambda \sim \text{gamma}\,(\alpha + \sum_{i=1}^n y_i,\ \beta + n)$
$y_i \sim \text{normal}(\mu, \sigma^2)$ σ^2 is known.	$\mu \sim \text{normal}\,(\mu_0, \sigma_0^2)$	$\mu \sim \text{normal}\left(\frac{\left(\frac{\mu_0}{\sigma_0^2} + \frac{\sum_{i=1}^n y_i}{\sigma^2}\right)}{\left(\frac{1}{\sigma_0^2} + \frac{n}{\sigma^2}\right)}, \left(\frac{1}{\sigma_0^2} + \frac{n}{\sigma^2}\right)^{-1}\right)$
$y_i \sim \text{normal}(\mu, \sigma^2)$ μ is known.	$\sigma^2 \sim$ inverse gamma(α, β)	$\sigma^2 \sim$ inverse gamma $\left(\alpha + \frac{n}{2}, \beta + \frac{\sum_{i=1}^n (y_i - \mu)^2}{2}\right)$
$y_i \sim \text{lognormal}(\mu, \sigma^2)$, μ is known	$\sigma^2 \sim$ inverse gamma(α, β) ,	$\sigma^2 \sim$ inverse gamma $\left(n/2 + \alpha, \frac{(\log(y_i) - \mu)^2}{2} + \beta\right)$
$y_i \sim \text{lognormal}\,(\mu, \sigma^2)$ σ^2 is known	$\mu \sim \text{normal}\,(\mu_0, \sigma_0^2)$	$\mu \sim \text{normal}\left(\frac{\left(\frac{\mu_0}{\sigma_0^2} + \frac{\sum_{i=1}^n \log y_i}{\sigma^2}\right)}{\left(\frac{1}{\sigma_0^2} + \frac{n}{\sigma^2}\right)}, \left(\frac{1}{\sigma_0^2} + \frac{n}{\sigma^2}\right)^{-1}\right)$

Input $\to$ $\underline{X} = \underline{x} \longrightarrow$ datos

$\to f_{\underline{X}}(\underline{x}|\theta) \longrightarrow$ modelo

$\downarrow$ para el bayesiano

$\tilde{\pi}(\theta)$ la conocemos como prior

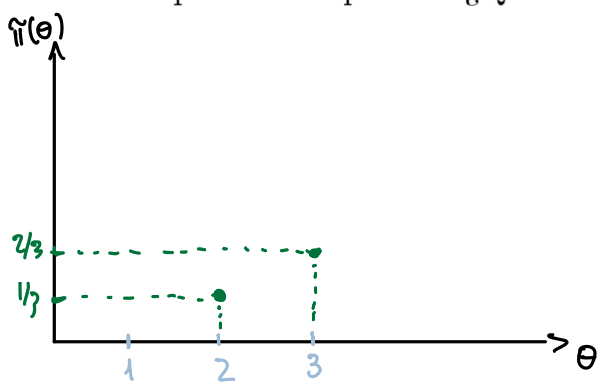
$$\tilde{\pi}(\theta|\underline{x}) = \frac{L(\theta|\underline{x})\,\tilde{\pi}(\theta)}{P(\underline{x})}$$ Regla de Bayes

estimador

$$\propto L(\theta|\underline{x})\,\tilde{\pi}(\theta)$$

(1)

1. Si $X|\theta \sim \text{Poiss}(\theta)$, y además podemos asumir que $P(\theta = 2) = 1/3$ y $P(\theta = 3) = 2/3$ (esto es $\Theta = \{2, 3\}$). Dada la información $x_1 = 2$ y $x_2 = 4$, compute la probabilidad a–posteriori[1] para θ. ¿Qué inferencia puede hacer respecto del parámetro θ?



- $\tilde{\pi}(\theta) = (1/3)^{\mathbb{1}\{\theta=2\}} \cdot (2/3)^{1-\mathbb{1}\{\theta=2\}}$
- $L(\theta|\underline{X}=\underline{x}) = f_{\underline{X}}(\underline{x}|\theta) \overset{iid}{=} \prod_{i=1}^{2} \frac{e^{-\theta}\cdot\theta^{x_i}}{(x_i)!} = \frac{e^{-\theta}\cdot\theta^{2}}{2}\cdot\frac{e^{-\theta}\cdot\theta^{4}}{24}$

$$= \frac{e^{-2\theta}\cdot\theta^{6}}{48}$$

- $P(\underline{x}) = P(\underline{x}|\theta=2)\,\tilde{\pi}(\theta=2) + P(\underline{x}|\theta=3)\,\tilde{\pi}(\theta=3)$

$$= \frac{e^{-4}\cdot 2^{6}}{48}\cdot\frac{1}{3} + \frac{e^{-6}\cdot 3^{6}}{48}\cdot\frac{2}{3}$$

$$\simeq 0.033$$

$$\Rightarrow \tilde{\pi}(\theta|\underline{x}) = \frac{\left(\frac{e^{-2\theta}\cdot\theta^{6}}{48}\right)\times\left[(1/3)^{\mathbb{1}\{\theta=2\}}\,(2/3)^{1-\mathbb{1}\{\theta=2\}}\right]}{0.033}$$

$$\simeq \left(\frac{1}{48\times 0.033}\right)\left[e^{-2\theta}\,\theta^{6}\,(1/3)^{\mathbb{1}\{\theta=2\}}\,(2/3)^{1-\mathbb{1}\{\theta=2\}}\right]$$

$$\propto e^{-2\theta}\,\theta^{6}\,(1/3)^{\mathbb{1}\{\theta=2\}}\,(2/3)^{1-\mathbb{1}\{\theta=2\}}$$

Inferencia respecto a θ: $\hat{\theta}_{MAP}$

- $\tilde{\pi}(\theta=2|\underline{x}) \propto e^{-4}\cdot 2^{6}\cdot\left(\frac{1}{3}\right) \approx 0.39$
- $\tilde{\pi}(\theta=3|\underline{x}) \propto e^{-6}\,3^{6}\left(\frac{2}{3}\right) \approx 1.2$

$\therefore \hat{\theta}_{MAP} = 3$

2. Una de las ventajas del enfoque Bayesiano reside en que el teorema de Bayes se puede utilizar de forma secuencial; y esto es particularmente útil cuando necesitamos 'refrescar el modelo' con información nueva. Imagina que de manera secuencial recibes la información (datos) $\mathbf{x}_1$ y luego $\mathbf{x}_2$, argumenta porque es cierto que:

$$\pi(\theta|\mathbf{x}_1, \mathbf{x}_2) \propto L(\theta|\mathbf{x}_2)\pi(\theta|\mathbf{x}_1).$$

3. Si $X|\theta \sim \text{Poiss}(\theta)$ y $\theta \sim \text{Gamma}(\alpha, \beta)$[2], esto es:

$$\pi(\theta; \alpha, \beta) = \frac{\theta^{\alpha-1}e^{-\theta/\beta}}{\Gamma(\alpha)\beta^\alpha}, \quad \alpha > 0,\ \beta > 0,\ \theta > 0,$$

donde α y β son dos hiperparámetros conocidos (elegidos por quien modela el problema).

(a) Comprobar que la distribución aposteriori de θ es Gamma de parámetros: $\alpha_n = \sum_{i=1}^n x_i + \alpha$ y $\beta_n = \beta/(n\beta + 1)$.

(b) ¿A dónde convergen la media y la varianza a posteriori cuando $n \to \infty$?

(c) Imagine que la variable aleatoria X_i da cuenta de la cantidad de delitos registrados en la ciudad en el día i; y que de una muestra de 10 días se tiene que $\sum_{i=1}^{10} x_i = 140$. Justificando su elección de los parámetros α y β (encuentre algún argumento razonable para elegirlos), obtenga la distribución a posteriori de θ.

(d) Reporte la media y varianza a posteriori.

(e) Construya un HPD al 95% y 99% e interprete los resultados.

$\to$ 87%

$\to$ 44%

(a) $\cdot\ L(\theta|\underline{X}) = f_{\underline{X}}(\underline{X}|\theta) \overset{iid}{=} \prod_{i=1}^n f_X(x_i|\theta) = \prod_{i=1}^n \frac{e^{-\theta}\theta^{x_i}}{(x_i)!} = \left(\prod_{i=1}^n \frac{1}{(x_i)!}\right) e^{-n\theta}\theta^{\sum^n x_i}$

$\cdot\ \pi(\theta|\alpha,\beta) = \frac{\theta^{\alpha-1}}{\Gamma(\alpha)} \cdot \frac{1}{\beta^\alpha} \cdot e^{-\frac{1}{\beta}\theta}$

$\pi(\theta|\underline{X}) \propto L(\theta|\underline{X}) . \pi(\theta|\alpha,\beta)$

$$L(\theta|\underline{X}) . \pi(\theta|\alpha,\beta) = \left(\prod_{i=1}^n \frac{1}{(x_i)!}\right) e^{-n\theta}\theta^{\sum^n x_i} \cdot \frac{\theta^{\alpha-1}}{\Gamma(\alpha)} \cdot \frac{1}{\beta^\alpha} \cdot e^{-\frac{1}{\beta}\theta}$$

$$= \left[\left(\prod_{i=1}^n \frac{1}{(x_i)!}\right)\frac{1}{\Gamma(\alpha)\beta^\alpha}\right] . \left(\theta^{\sum^n x_i + \alpha - 1}\right) . \left(e^{-n\theta - \frac{1}{\beta}\theta}\right)$$

$$\propto \theta^{\sum^n x_i + \alpha - 1} . e^{-(n + \frac{1}{\beta})\theta}$$

$\cdot\ \alpha_n = \sum_{i=1}^n x_i + \alpha$

$\cdot\ \frac{1}{\beta_n} = (n + \frac{1}{\beta}) = \frac{\beta n + 1}{\beta} \Rightarrow \beta_n = \beta\left(\frac{1}{\beta n + 1}\right)$

$\Big\}\ \pi(\theta|\underline{X}) \propto \theta^{\alpha_n - 1} e^{-\frac{1}{\beta_n}\theta} = \Gamma(\alpha_n, \beta_n)$

$= \Gamma\left(\sum^n x_i + \alpha, \beta\left(\frac{1}{\beta n + 1}\right)\right)$

(b)
- $\mathbb{E}_{\underline{x}}(\theta) = \alpha_n \cdot \beta_n = \left(\sum^{n} x_i + \alpha\right) \cdot \beta\left(\frac{1}{\beta n+1}\right) = (n\bar{x}+\alpha)\,\beta \cdot \left(\frac{1}{\beta n+1}\right)$

$$= \underbrace{\frac{\beta n \bar{x}}{\beta n+1}}_{\bar{x}} + \underbrace{\frac{\alpha\beta}{\beta n+1}}_{\to 0}$$

$$\lim_{n\to\infty} \mathbb{E}_{\underline{x}}(\theta) = \bar{x}$$

- $\mathbb{V}ar_{\underline{x}}(\theta) = \alpha_n \beta_n^2 = (n\bar{x}+\alpha)\left[\beta\left(\frac{1}{\beta n+1}\right)\right]^2$

$$\lim_{n\to\infty} \mathbb{V}ar_{\underline{x}}(\theta) = 0$$

(c) . $\sum_{i=1}^{10} x_i = 140 \Rightarrow \bar{x}_n = 14$

I) $\alpha = 2,\ \beta = 0.5 \Rightarrow \mathbb{E}_{\alpha,\beta}[\theta] = 1,\ \mathbb{V}ar_{\alpha,\beta}[\theta] = 0.5$ → a priorístico

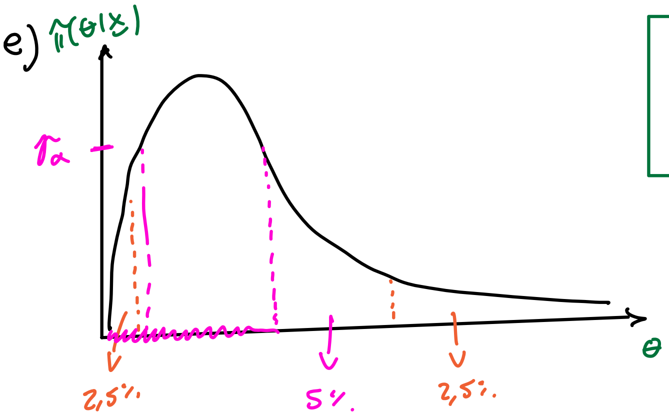
II) Empirical Bayes $\Rightarrow \mathbb{E}_{\alpha,\beta}[\theta] = \bar{x}_n = 14 \Rightarrow \alpha\beta = 14$ → infinitas combinacións
- $\alpha = 1$ y $\beta = 14$
- $\alpha = 2$ y $\beta = 7$
- $\alpha = 7$ y $\beta = 2$
- $\alpha = 14$ y $\beta = 1$

máis seguro ⇒ prior máis informativo

d)
Si $\alpha = 1$ y $\beta = 14 \Rightarrow \tilde{\pi}(\theta|\underline{x}) = \Gamma(141, 141/141) \Rightarrow \mathbb{E}_{\underline{x}}[\theta] = 14,\ \mathbb{V}ar_{\underline{x}}[\theta] = 1.39$

Si $\alpha = 14$ y $\beta = 1 \Rightarrow \tilde{\pi}(\theta|\underline{x}) = \Gamma(154, 1/11) \Rightarrow \mathbb{E}_{\underline{x}}[\theta] = 14,\ \mathbb{V}ar_{\underline{x}}[\theta] = 1.2727$

e)



$$\boxed{\text{HDI} = \{\theta \in \Theta : \tilde{\pi}(\theta|\underline{x}) \geq \nu_\alpha\} \,/\, \nu_\alpha \text{ es el valor máis alto que garantiza } P(\theta \in \text{HPD}(\nu_\alpha)|\underline{x}) = 1-\alpha}$$

. Para $\tilde{\pi}(\theta|\underline{x}) = \Gamma(154, 1/11)$

$\text{HPD}_{95\%} = [11.876, 16.295]$

$\text{HPD}_{99\%} = [11.264, 17.076]$