CAPÍTULO 6. MUESTRAS GRANDES

Para leer

Bernardo y Smith (1994), Sección 5.3.
Gelman et al (1995), Capítulo 4, Secciones 4.1 – 4.3 y Apéndice B.

Si la muestra es muy grande, está claro que los valores de los parametros de la distribución a priori suelen ser poco importantes.

Ejemplo 83 $X|\mu,\sigma^2 \sim \mathcal{N}(\mu,\sigma^2)$. *Supongamos una distribución inicial conjugada. Entonces, por ejemplo:*

$$E[\mu|\mathbf{x}] = \frac{cm + n\bar{x}}{c+n} \to \bar{x}$$

cuando $n \to \infty$.

El caso discreta
Convergencia al verdadero valor de θ

El siguiente teorema demuestra la convergencia de la distribución a posteriori en el caso de que $\boldsymbol{\Theta}$ sea discreta.

Teorema 11 *Sea* $X|\boldsymbol{\theta} \sim f(\cdot|\boldsymbol{\theta})$ *donde el espacio* $\boldsymbol{\Theta} = \{\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \ldots\}$ *es contable. Supongamos que* $\boldsymbol{\theta}_v \in \boldsymbol{\Theta}$ *es el verdadero valor de* θ.

Sea la distribución a priori $P(\boldsymbol{\theta})$ *donde se tiene* $P(\boldsymbol{\theta}_i) > 0 \, \forall \, i$ *y supongamos que*

$$\int f(x|\boldsymbol{\theta}_v) \log \frac{f(x|\boldsymbol{\theta}_v)}{f(x|\boldsymbol{\theta}_i)} \, dx > 0 \quad \forall \, i \neq v. \quad \textit{Entonces}$$

$\lim_{n\to\infty} P(\boldsymbol{\theta}_v|\mathbf{x}) = 1$ *y* $\lim_{n\to\infty} P(\boldsymbol{\theta}_i|\mathbf{x}) = 0 \; \forall \; i \neq v$.

Demostración

Sea $\mathbf{x} = (x_1, \ldots, x_n)$.

$$\begin{aligned} P(\boldsymbol{\theta}_i|\mathbf{x}) &= \frac{P(\boldsymbol{\theta}_i) f(\mathbf{x}|\boldsymbol{\theta}_i)}{\sum_i P(\boldsymbol{\theta}_i) f(\mathbf{x}|\boldsymbol{\theta}_i)} \\ &= \frac{P(\boldsymbol{\theta}_i) f(\mathbf{x}|\boldsymbol{\theta}_i)/f(\mathbf{x}|\boldsymbol{\theta}_v)}{\sum_i P(\boldsymbol{\theta}_i) f(\mathbf{x}|\boldsymbol{\theta}_i)/f(\mathbf{x}|\boldsymbol{\theta}_v)} \\ &= \frac{\exp\left(\log P(\boldsymbol{\theta}_i) + S_i\right)}{\sum_i \exp\left(\log P(\boldsymbol{\theta}_i) + S_i\right)} \end{aligned}$$

donde $S_i = \sum_{j=1}^{n} \log \frac{f(x_j|\boldsymbol{\theta}_i)}{f(x_j|\boldsymbol{\theta}_v)}$.

Condicionada en $\boldsymbol{\theta}_v$, S_i es la suma de n cantidades aleatorias independientes y identicamente distribuidas y entonces, por la ley fuerte de números grandes, se tiene

$$\lim_{n\to\infty} \frac{1}{n} S_i = \int f(x|\boldsymbol{\theta}_v) \log \frac{f(x|\boldsymbol{\theta}_i)}{f(x|\boldsymbol{\theta}_v)}\, dx.$$

Esta cantidad es negativa si $i \neq v$ y cero si $i = v$. Entonces, cuando $n \to \infty$, $S_v \to 0$ y $S_i \to -\infty$ si $i \neq v$ y el resultado queda demostrado.

$\diamond$

El caso continuo

Es normal que los estimadores bayesianos se aproximen a los estimadores EMV en muestras grandes. Recordamos que la distribución de la EMV es aproximadamente normal cuando $n \to \infty$. Existen resultados parecidos para la distribución a posteriori.

Teorema 12 *Sea $X_i|\boldsymbol{\theta} \sim f(\cdot|\boldsymbol{\theta})$ con distribución a priori $f(\boldsymbol{\theta})$. Dados los datos $\mathbf{x}$, cuando $n \to \infty$,*

1 $\boldsymbol{\theta}|\mathbf{x} \approx \mathcal{N}(E[\boldsymbol{\theta}|\mathbf{x}], V[\boldsymbol{\theta}|\mathbf{x}])$, suponiendo que la media y varianza de $\boldsymbol{\theta}$ existen,

2 $\boldsymbol{\theta}|\mathbf{x} \approx \mathcal{N}(\hat{\boldsymbol{\theta}}, I(\hat{\boldsymbol{\theta}})^{-1})$ donde $\hat{\boldsymbol{\theta}}$ es la moda de la distribución final y $I(\boldsymbol{\theta})$ es **la información observada**

$$I(\boldsymbol{\theta}) = -\frac{d^2}{d\boldsymbol{\theta}^2} \log(f(\boldsymbol{\theta}|\mathbf{x})).$$

3 $\boldsymbol{\theta}|\mathbf{x} \approx \mathcal{N}(\hat{\hat{\boldsymbol{\theta}}}, I^\star(\hat{\hat{\boldsymbol{\theta}}})^{-1})$ *donde* $\hat{\hat{\boldsymbol{\theta}}}$ *es la EMV de* $\boldsymbol{\theta}$*, suponiendo que la exista y*

$$I^\star(\boldsymbol{\theta}) = -\frac{d^2}{d\boldsymbol{\theta}^2} \log(f(\mathbf{x}|\boldsymbol{\theta}))$$

4 $\boldsymbol{\theta}|\mathbf{x} \approx \mathcal{N}(\hat{\hat{\boldsymbol{\theta}}}, I^{\star\star}(\hat{\hat{\boldsymbol{\theta}}})^{-1})$ *donde*

$$I^{\star\star}(\boldsymbol{\theta}) = -nE_X\left[\frac{d^2}{d\boldsymbol{\theta}^2} \log(f(X|\boldsymbol{\theta}))\right]$$

Se demuestran los resultados del teorema mediante una expansión de Taylor de la distribución a posteriori sobre, por ejemplo $\hat{\boldsymbol{\theta}}$. Ver Bernardo y Smith (1994), Gelman et al (1995).

Normalmente, la primera aproximación será mejor que la segunda y tercera, y la ultima será la peor.

Observación 56 *En muchos casos, la media y varianza a posteriori son difíciles de calcular pero es mucho más fácil evaluar la moda.*

Ejemplo 84 Aproximación a una distribución beta. *Sea* $X|\theta \sim \mathcal{BI}(n,\theta)$ *y* $\theta \sim \mathcal{B}(\alpha,\beta)$. *Entonces,* $\theta|\mathbf{x} \sim \mathcal{B}(\alpha+x,\beta+n-x)$.

Si n *es grande, se puede aproximar la distribución a posteriori de* θ. *Comparamos las cuatro aproximaciones.*

1

$$\theta|\mathbf{x} \approx \mathcal{N}\left(\frac{\alpha+x}{\alpha+\beta+n}, \frac{(\alpha+x)(\beta+n-x)}{(\alpha+\beta+n)^2(\alpha+\beta+n+1)}\right)$$

2 *Calculamos la moda de la distribución beta.*

$$\begin{aligned}
\log f(\theta|\mathbf{x}) &= c + (\alpha + x - 1)\log(\theta) + \\
&\quad (\beta + n - x - 1)\log(1-\theta) \\
\frac{d}{d\theta} &= \frac{\alpha + x - 1}{\theta} - \frac{\beta + n - x - 1}{1-\theta} \\
\hat{\theta} &= \frac{\alpha + x - 1}{\alpha + \beta + n - 2}
\end{aligned}$$

es la moda.

Para evaluar la información observada, se tiene

$$\begin{aligned}
\frac{d^2}{d\theta^2}\log(f(\theta|\mathbf{x})) &= -\frac{\alpha + x - 1}{\theta^2} - \frac{\beta + n - x - 1}{(1-\theta)^2} \\
I(\hat{\theta}) &= \frac{\alpha + x - 1}{\hat{\theta}^2} + \frac{\beta + n - x - 1}{(1-\hat{\theta})^2} \\
&= \frac{(\alpha + \beta + n - 2)^3}{(\alpha + x - 1)(\beta + n - x - 1)}
\end{aligned}$$

$$\theta|\mathbf{x} \approx N\left(\frac{\alpha + x - 1}{\alpha + \beta + n - 2}, \frac{(\alpha + x - 1)(\beta + n - x - 1)}{(\alpha + \beta + n - 2)^3}\right)$$

3 El EMV es $\widehat{\widehat{\theta}} = \frac{x}{n}$ y $I^{\star}(\widehat{\widehat{\theta}}) = \frac{n^3}{x(n-x)}$. Luego

$$\theta|\mathbf{x} \approx \mathcal{N}\left(\frac{x}{n}, \frac{x(n-x)}{n^3}\right)$$

4

$$\begin{aligned}
-\frac{d^2}{d\theta^2}\log(f(X|\theta)) &= \frac{X}{\theta^2} + \frac{n-X}{(1-\theta)^2} \\
I^{\star\star}(\theta) &= \frac{n\theta}{\theta^2} + \frac{n(1-\theta)}{(1-\theta)^2} \\
&= \frac{n}{\theta} + \frac{n}{(1-\theta)} \\
I^{\star\star}(\widehat{\widehat{\theta}}) &= \frac{n^3}{x(n-x)}
\end{aligned}$$

y se tiene la misma aproximación.

Cuando $n >> \alpha + \beta$ las aproximaciones serán parecidas, pero en muestras pequeñas pueden dar resultados diferentes.

Ejemplo 85 *Supongamos que* $\alpha = \beta = 2$ *y* $x = 20$, $n = 30$. *Hallamos las aproximaciones y estimamos* $P(\theta > 0{,}5|data)$.

Se tiene $\theta|\mathbf{x} \sim \mathcal{B}(22, 12)$ *y mediante las tablas de la función beta incompleta, se calcula que* $P(\theta > 0{,}5|\mathbf{x}) = 0{,}95993$.

1. *Se aproxima con*

$$\theta|\mathbf{x} \approx \mathcal{N}\left(\frac{22}{34}, \frac{22 \times 12}{34^2 \times 35}\right) \approx \mathcal{N}(0{,}64706, 0{,}006525)$$

Ahora $P(\theta > 0{,}5|\mathbf{x}) \approx P(Z > -1{,}8206) = 0{,}9660$.

2.

$$\theta|\mathbf{x} \approx \mathcal{N}\left(\frac{21}{32}, \frac{21 \times 11}{32^3}\right) \approx \mathcal{N}(0{,}65625, 0{,}00705).$$

$P(\theta > 0{,}5|\mathbf{x}) \approx P(Z > -1{,}8610) = 0{,}9686$.

3.

$$\theta|\mathbf{x} \approx \mathcal{N}\left(\frac{20}{30}, \frac{20 \times 10}{30^3}\right) \approx \mathcal{N}(0{,}66667, 0{,}00741).$$

$P(\theta > 0{,}5|\mathbf{x}) \approx P(Z > -1{,}9365) = 0{,}9735$.

El siguente ejemplo demuestra la aproximación con una muestra pequeña.

Ejemplo 86 *(Berger 1985)*

Dada una muestra $\mathbf{x} = (4{,}0\ 5{,}5\ 7{,}5\ 4{,}5\ 3{,}0)$ *de una distribución Cauchy:* $\mathcal{C}(\mu, 1)$*, y una distribución inicial (impropia)* $f(\mu) \propto 1$*, queremos aproximar a la distribución a posteriori* $f(\mu|\mathbf{x})$*.*

Se puede demostrar que $E[\mu|\mathbf{x}] = 4{,}45$ *y* $V[\mu|\mathbf{x}] = 0{,}562$ *y la siguiente tabla proporciona las* α *quantiles de la distribución exacta y de la aproximación normal.*

α	2,5	25	50	75	97,5
F_{exac}	3,17	4,07	4,52	5,00	6,15
F_{norm}	3,08	4,05	4,55	5,06	6,02

También, la moda a posteriori es 4,55 *y la información esperada* $I^{\star\star}(\hat{\hat{\theta}}) = 5/2$*. Luego el método 4 da la aproximación* $\mathcal{N}(4{,}55, 0{,}4)$ *que es peor.*

Cuando no se puede aplicar el teorema

Existen algunas situaciones cuando no se puede aplicar el teorema. Por ejemplo

- Si la densidad a priori del verdadero valor de $\boldsymbol{\theta}$ es 0,
- Si la distribución a posteriori es impropia,
- Si el modelo no es identificable.

Ejemplo 87 *Supongamos el modelo*

$$f(x|\theta_1,\ldots,\theta_k) = w_1 g(x|\theta_1) + \ldots + w_k g(x|\theta_k)$$

una mixtura de k densidades de la misma familia.

Dados los datos, la verosimilitud será multimodal porque el modelo no es identifiable. Necesitamos restringir el espacio Θ, para que el modelo sea identifiable.

Por ejemplo, podemos suponer que $\theta_1 < \ldots < \theta_k$.

Ver Gelman et al (1995) para mas ejemplos.

La aproximación de Laplace

Tierney y Kadane (1996) desarrollo esta generalización de la aproximación normal para la estimación de la media a posteriori.

Se quiere estimar

$$E[g(\theta)|\mathbf{x}] = \frac{\int g(\theta) l(\theta|\mathbf{x}) f(\theta)\, d\theta}{\int l(\theta|\mathbf{x}) f(\theta)\, d\theta}$$

Supongamos que $g(\cdot)$ es no negativa.

Se expresa la esperanza en la forma

$$E[g(\theta)|\mathbf{x}] = \frac{\int \exp(-nh^*(\theta))\, d\theta}{\int \exp(-nh(\theta))\, d\theta}$$

donde

$$\begin{aligned} -nh(\theta) &= \log f(\theta) + \log l(\theta|\mathbf{x}) \\ \text{y} \quad -nh^*(\theta) &= \log g(\theta) + \log f(\theta) + \log l(\theta|\mathbf{x}). \end{aligned}$$

Dada esta expresión, se usa la expansión de Taylor de h (h^*) en torno a la moda $\widehat{\theta}$ ($\widehat{\widehat{\theta}}$).

$$\begin{aligned} -h(\widehat{\theta}) &= \max_{\theta}(-h(\theta)) \\ -h^*(\widehat{\widehat{\theta}}) &= \max_{\theta}(-h^*(\theta)) \end{aligned}$$

y se quedan con los términos cuadráticos. Se estima el integral del denominador con

$$\int \exp(-nh(\theta))\, d\theta \approx \sqrt{2\pi}\sigma n^{-1/2} \exp(-nh(\widehat{\theta}))$$

$$\text{donde} \quad \sigma = \left(\frac{d^2}{d\theta^2} h(\theta)|_{\theta=\widehat{\theta}} \right)^{-1/2}$$

y una aproximación parecida en el numerador.

Nos proporciona la estimación:

$$E[g(\theta|\mathbf{x})] \approx \left(\frac{\sigma^*}{\sigma}\right) \frac{g(\widehat{\widehat{\theta}}) f(\widehat{\widehat{\theta}}) l(\widehat{\widehat{\theta}}|\mathbf{x})}{f(\widehat{\theta}) l(\widehat{\theta}|\mathbf{x})}$$

donde

$$\sigma^* = \left(\frac{d^2}{d\theta^2} h^*(\theta)|_{\theta=\widehat{\widehat{\theta}}} \right)^{-1/2}$$

Ejemplo 88 *Volvemos al Ejemplo 84. Se tiene* $\theta|\mathbf{x} \sim \mathcal{B}(\alpha + x, \beta + n - x)$.

Observación 57 *Sin perdida de generalidad, se suponen* $0 \leq \alpha, \beta < 1$. *Si no, se puede transformar* $x \to x + [\alpha]$ *y* $n \to n + [\alpha] + [\beta]$.

Escribiendo la densidad en una forma conveniente

$$\begin{aligned} f(\theta|\mathbf{x}) &\propto \theta^{\alpha+x-1}(1-\theta)^{\beta+n-x-1} \\ &\propto \exp(-nh(\theta)) \end{aligned}$$

donde

$$\begin{aligned} h(\theta) &= -\frac{1}{n}((\alpha + x - 1)\log\theta + \\ &\quad (\beta + n - x - 1)\log(1-\theta)) \end{aligned}$$

Se puede demostrar (Ejercicio) que el estimador de Laplace de la media a posteriori será

$$\frac{(\alpha+x)^{\alpha+x+1/2}}{(\alpha+x-1)^{\alpha+x-1/2}} \frac{(\alpha+\beta+n-2)^{\alpha+\beta+n-1/2}}{(\alpha+\beta+n-1)^{\alpha+\beta+n+1/2}}.$$

Por ejemplo, si $\theta|\mathbf{x}\ \mathcal{B}(8,12)$, *poniendo* $\alpha = \beta = 0$ *y* $n = 20$ *el estimador de Laplace es*

$$E[\theta|\mathbf{x}] \approx \frac{8^{8,5}}{7^{7,5}}\frac{18^{19,5}}{19^{20,5}} \approx ,3994$$

El valor exacto de la media es $8/20 = 0{,}4$.

Utilizando la aproximación 2 del teorema de Capítulo 6, se estima la media con la moda $7/18 = ,3889$. *La aproximación de Laplace es mejor.*

Ejemplo 89 Estimación del Factor Bayes

Consideramos el caso de dos hipótesis compuestos H_0 y H_1. El factor Bayes es

$$B_1^0 = \frac{\int f(\mathbf{x}|\boldsymbol{\theta}_0, H_0) f(\boldsymbol{\theta}_0|H_0)\, d\boldsymbol{\theta}_0}{\int f(\mathbf{x}|\boldsymbol{\theta}_1, H_1) f(\boldsymbol{\theta}_1|H_1)\, d\boldsymbol{\theta}_1}$$

y el numerador y denominador son los dos funciones positivas. Entonces se puede aplicar la aproximación de Laplace. Ver Kass y Raftery (1995).

Propiedades y problemas de la aproximación de Laplace

- La aproximación es $O(1/n^2)$
- Se puede extender el método al caso multidimensional
- Si $\Theta \neq R$, se puede reparameterizar el modelo para mejorar la aproximación
- Necesitamos ser capaces de calcular los EMV.