

CAPÍTULO 6. MUESTRAS GRANDES

Para leer

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 42 $X|\mu, \sigma^2 \sim \mathcal{N}(\mu, \sigma^2)$. Supongamos una distribución inicial conjugada. Entonces, por ejemplo:

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

cuando $n \rightarrow \infty$.

Un teorema central del límite bayesiano

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 $n \rightarrow \infty$. Existen resultados parecidos para la distribución a posteriori.

Teorema 9 *Sea $X_i|\theta \sim f(\cdot|\theta)$ con distribución a priori $f(\theta)$. Dados los datos \mathbf{x} , cuando $n \rightarrow \infty$,*

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

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

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

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

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

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

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

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

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

Observación 20 *En muchos casos, la media y varianza a posteriori son difíciles de calcular pero es mucho más fácil evaluar la moda.*

Ejemplo 43 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 $\hat{\theta} = \frac{x}{n}$ y $I^*(\hat{\theta}) = \frac{n^3}{x(n-x)}$. Luego

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

4

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

y se tiene la misma aproximación.

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

Ejemplo 44 Supongamos que $\alpha = \beta = 2$ y $x = 20$, $n = 30$. Hallamos las aproximaciones y estimamos $P(\theta > 0,5|data)$.

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

1. Se aproxima con

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

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

2.

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

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

3.

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

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

Ejemplo 45 Supongamos que $X|\theta \sim \mathcal{P}(\theta)$ es el n'mero de llamadas telefónicas que se reciben por minuto en una empresa grande. Se supone una distribución a priori $\theta \sim \mathcal{G}(2, 6)$. Suponiendo que se reciben 20 llamadas en 1 hora, estimar la probabilidad a posteriori de que $\theta < 1/2$.

La verosimilitud es

$$l(\theta|\mathbf{x}) = \frac{(60\theta)^{20} e^{-60\theta}}{20!} \propto \theta^{20} e^{-60\theta}$$

y luego la distribución a posteriori es

$$\begin{aligned} f(\theta|\mathbf{x}) &\propto \theta^{2-1} e^{-6\theta} \theta^{20} e^{-60\theta} \\ &\propto \theta^{22-1} e^{-66\theta} \end{aligned}$$

$$\theta|\mathbf{x} \sim \mathcal{G}(22, 66)$$

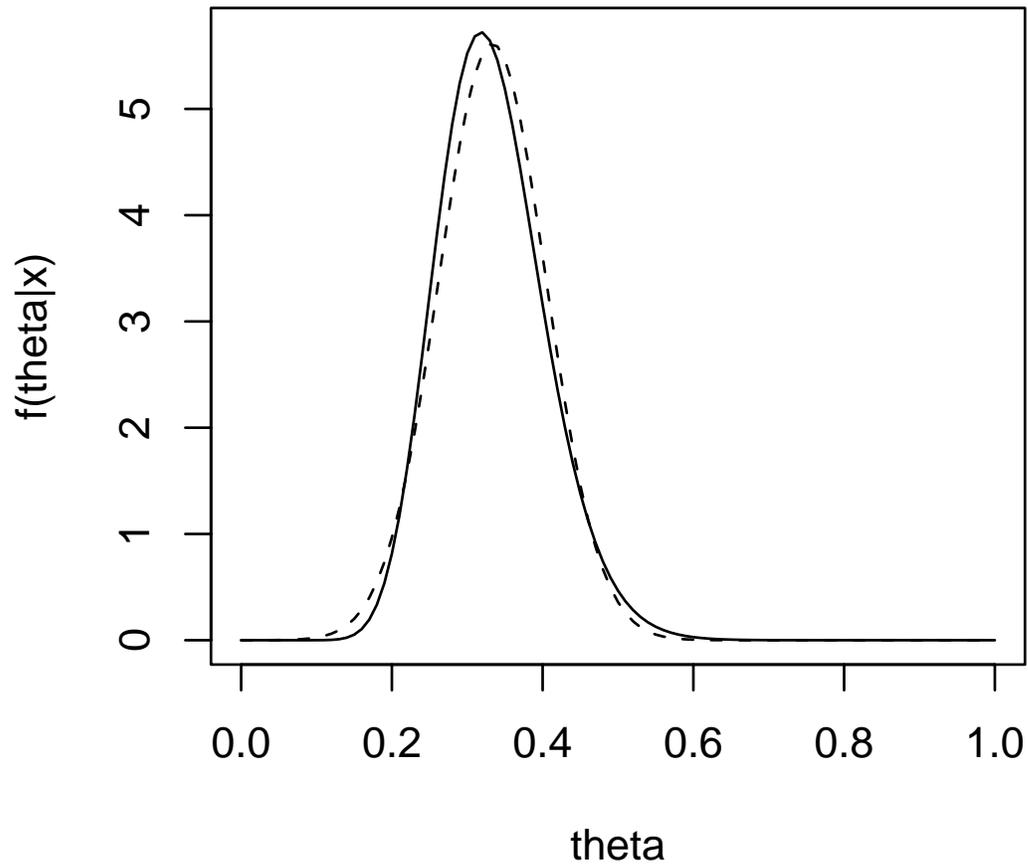
Se tiene $E[\theta|\mathbf{x}] = \frac{22}{66} = \frac{1}{3}$ y $V[\theta|\mathbf{x}] = \frac{22}{66^2} = \frac{1}{198}$.

Entonces aproximando la distribución gamma por una normal, se tiene

$$\begin{aligned} P(\theta < 1/2|\mathbf{x}) &= P\left(\sqrt{198}\left(\theta - \frac{1}{3}\right) < \sqrt{198}\left(\frac{1}{2} - \frac{1}{3}\right) \mid \mathbf{x}\right) \\ &\approx P(Z < 2,345) \quad \text{donde } Z \sim \mathcal{N}(0, 1) \\ &\approx 0,9905 \end{aligned}$$

La solución exacta que se puede calcular en R es $P(\theta < 0,5|\mathbf{x}) = 0,9825$

El siguiente gráfico muestra la verdadera densidad a posteriori y la densidad normal.



Se ve que la densidad gamma es bastante parecida a la normal.

El siguiente ejemplo demuestra la aproximación con una muestra pequeña.

Ejemplo 46

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_{exact}	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 $I^{**}(\hat{\theta}) = 5/2$. 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 θ es 0,
- Si la distribución a posteriori es impropia,
- Si el modelo no es identificable.

Ejemplo 47 *Supongamos el modelo*

$$f(x|\theta_1, \dots, \theta_k) = w_1g(x|\theta_1) + \dots + w_kg(x|\theta_k)$$

una mixtura de k densidades de la misma familia.

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

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

Ver Gelman et al (1995) para mas ejemplos.