Tema

5

Inferencia Bayesiana en el Modelo Normal

Índice

1. Introducción

En este tema se va a estudiar la inferencia bayesiana sobre los parámetros de una distribución normal. Inicialmente supondremos uno de los dos parámetros conocidos. Este caso no es demasiado real pero facilita los cálculos y la comprensión del tema. Posteriormente abordaremos el caso real que trata cuando ambos parámetros son desconocidos.

Supongamos que $X|\mu,\sigma^2 \sim N(\mu,\sigma^2)$. Entonces, la función de densidad es proporcional a

$$f(x|\mu,\sigma^2) \propto \frac{1}{\sigma}\exp\left[-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}\right].$$

En definitiva, para reconocer el formato de una función de densidad con sus parámetros debemos fijarnos en la siguiente expresión

$$f(x|\mu,\sigma^2) \propto \frac{1}{\sigma}\exp\left[-\frac{1}{2}\frac{(x^2-2\mu x)}{\sigma^2}\right].$$

2. Inferencia en la normal de varianza conocida

2.1. La función de verosimilitud

Sea X una variable aleatoria normal de media desconocida μ y varianza conocida σ^2. Siguiendo la notación bayesiana hasta ahora vamos a denotar al parámetro desconocido, y por tanto objeto de análisis, por θ. Luego, se tendría que

$$f(x|\theta) = \frac{1}{\sqrt{2\pi}\sigma} exp\left\{-\frac{(x-\theta)^2}{2\sigma^2}\right\} \text{ para } x,\theta \in \mathbb{R}.$$

Sea $\mathbf{x}$ una muestra aleatoria simple de tamaño n de X, entonces

$$\begin{aligned} L(\theta|\mathbf{x}) &\propto \prod_{i=1}^{n} \exp\left[-\frac{(x_i-\theta)^2}{2\sigma^2}\right] \\ &= \exp\left[-\frac{\sum_{i=1}^{n}(x_i^2+\theta^2-2\theta x_i)}{2\sigma^2}\right] \\ &\propto \exp\left(-\frac{n\theta^2-2n\theta\bar{x}}{2\sigma^2}\right), \end{aligned}$$

(Nótese que se ha eliminado $\sum x_i^2$ al no depender de θ).

Luego mediante operaciones algebraicas sencillas se obtiene que

$$L(\theta|\mathbf{x}) \propto exp\left\{-\frac{n}{2\sigma^2}(\bar{x}-\theta)^2\right\} \text{ para } \theta \in \mathbb{R}.$$

2.2. Distribución a posteriori de la media

2.2.1. A priori de Jeffrey

Se tiene que $\frac{\partial \ln f(x|\theta)}{\partial\theta} = \frac{x-\theta}{\sigma^2}$. Por tanto,

$$I(\theta) = E\left[\frac{(X-\theta)^2}{\sigma^4}\right] = \frac{1}{\sigma^2}.$$

Debido a que la varianza es conocida, significa que la distribución a priori para la media es $\pi(\theta) = \sigma^{-1}$. Es decir, es constante, lo cual implica que es una distribución impropia y en este caso coincide con la no informativa completa.

Entonces, por un cálculo trivial se tiene que

$$\pi(\theta|\mathbf{x}) \propto \exp\left\{-\frac{n}{2\sigma^2}(\bar{x}-\theta)^2\right\}$$

Es decir,

$$\boxed{\theta|\mathbf{x} \sim N(\bar{x}, \sigma^2/n)}$$

2.2.2. A priori normal (Método de las conjugadas)

Supongamos que $\theta \sim N(\mu_0, \sigma_0^2)$, con μ_0 y σ_0^2 valores conocidos. Entonces, por un cálculo muy similar a la subsección anterior, se tiene que

$$\begin{aligned}\pi(\theta|\mathbf{x}) &\propto \exp\left\{-\frac{1}{2}\left(\frac{n\sigma_0^2(\bar{x}-\theta)^2+\sigma^2(\theta-\mu_0)^2}{\sigma_0^2\sigma^2}\right)\right\}\\ &= \exp\left\{-\frac{1}{2}\left[\frac{n\sigma_0^2\theta^2-2\theta\bar{x}n\sigma_0^2+\sigma^2\theta^2-2\mu_0\sigma^2\theta}{\sigma_0^2\sigma^2}\right]\right\}\\ &= \exp\left\{-\frac{1}{2}\left[\frac{(n\sigma_0^2+\sigma^2)\theta^2-2(n\sigma_0^2\bar{x}+\mu_0\sigma^2)\theta}{\sigma^2\sigma_0^2}\right]\right\}\end{aligned}$$

Es decir,

$$\boxed{\theta|\mathbf{x} \rightsquigarrow N\left(\frac{n\sigma_0^2\bar{x}+\sigma^2\mu_0}{n\sigma_0^2+\sigma^2}, \frac{\sigma_0^2\sigma^2}{n\sigma_0^2+\sigma^2}\right)}$$

2.3. Estimadores de Bayes e Intervalos de Credibilidad

Como se ha visto en la sección anterior, la distribución a posteriori va a ser normal con media μ_1 y varianza σ_1^2 donde en función de la a priori que se use dichos parámetros tendrán diferentes valores.

2.3.1. Estimador de Bayes

Como se sabe en la distribución normal la media y la mediana coinciden. Luego independientemente de la función de pérdida que se use el estimador de Bayes es para el caso de la a prioir de Jeffrey $\hat{\theta}_B = \bar{x}$ y en el caso de usar una a priori conjugada

$$\hat{\theta}_B = \frac{n\sigma_0^2\bar{x}+\sigma^2\mu_0}{n\sigma_0^2+\sigma^2}.$$

2.3.2. Intervalos de Credibilidad

En este caso el intervalo de credibilidad al $(1-\alpha)\,\%$ es muy sencillo y resulta

$$IC(\theta, 1-\alpha) = \mu_1 \pm z_{1-\frac{\alpha}{2}}\sigma_1.$$

Por ejemplo, si usamos la a priori de Jeffrey, el intervalo de credibilidad resultante es

$$IC(\theta, 1-\alpha) = \bar{x} \pm z_{1-\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}},$$

cuyo parecido con el correspondiente intervalo de confianza de la estadística frecuentista es bastante. La gran diferencia es la interpretación de cada uno de ellos. Nótese que en los intervalos de confianza la media muestral es una variable aleatoria y en el momento que se haga la realización muestral y se sustituya en la fórmula ya no proporciona la probabilidad de contener al parámetro. Sin embargo, en la estadística Bayesiana al tratar con la distribución a priori como medida de cuantificación de la incertidumbre acerca del parámetro, el intervalo de credibilidad sí proporciona la probabilidad de contener al parámetro. También es oportuno recalcar que al ser una densidad simétrica en la moda entonces el intervalo de credibilidad es HPDI.

2.4. Factor de Bayes

Supongamos que desamos realizar el siguiente contraste de hipótesis:

$$H_0 : \theta \leq \theta_0 \text{ frente a } H_1 : \theta > \theta_0.$$

Partimos de un conjunto de datos $\mathbf{x}$ procedente de la realización n pruebas independientes de una normal cuya media es θ y varianza conocida σ^2. Luego en este caso se tendría que

$$FB_{01} = \frac{P(\theta \leq \theta_0|\mathbf{x})}{P(\theta > \theta_0|\mathbf{x})} \cdot \frac{P(\theta > \theta_0)}{P(\theta \leq \theta_0)}.$$

El primer factor corresponde a las distribuciones a posteriori mientras que el segundo a las distribuciones a priori. Todas estas probabilidades se calculan a partir de la distribución normal con los parámetros en función de la a priori que se use. Es evidente que necesitaremos el **R** para llevar a cabo este cálculo. Posteriormente decidiremos la evidencia de las hipótesis en función del valor del FB_{01} según la tabla de Jeffrey.

2.5. Distribución Predictiva

Sea $\mathbf{x}$ un conjunto de datos procedentes de una muestra aleatoria simple de una distribución normal de media θ y varianza conocida σ^2. El objetivo es calcular la distribución predictiva de una nueva observación:

$$f(y|\mathbf{x}) = \int_{\Theta} f(y|\theta, \mathbf{x})\pi(\theta|\mathbf{x})d\theta.$$

Para ello vamos a calcular la función generatriz de momentos de la v.a. $y|\mathbf{x}$ y así podremos determinarla.

$$\begin{aligned} m_{y|\mathbf{x}}(t) &= \int_{-\infty}^{\infty} e^{ty} f(y|\mathbf{x}) dy \\ &= \int_{-\infty}^{\infty} e^{ty} \left[\int_{\Theta} f(y|\theta, \mathbf{x}) \pi(\theta|\mathbf{x}) d\theta \right] dy, \end{aligned}$$

aplicando ahora el Teorema de Fubini para intercambiar integrales quedaría

$$\begin{aligned} m_{y|\mathbf{x}}(t) &= \int_{\Theta} \pi(\theta|\mathbf{x}) \left[\int_{-\infty}^{\infty} e^{ty} f(y|\theta, \mathbf{x}) dy \right] d\theta \\ &= \int_{\Theta} \pi(\theta|\mathbf{x}) m_{y|\theta}(t) d\theta \\ &= \int_{\Theta} \pi(\theta|\mathbf{x}) \exp\left[\theta t + \frac{t^2\sigma^2}{2} \right] d\theta \\ &= e^{\frac{t^2\sigma^2}{2}} \int_{\Theta} \pi(\theta|\mathbf{x}) e^{\theta t} d\theta \\ &= e^{\frac{t^2\sigma^2}{2}} e^{\mu_1 t + \frac{t^2\sigma_1^2}{2}}, \end{aligned}$$

luego

$$\boxed{y|\mathbf{x} \sim N(\mu_1, \sigma^2 + \sigma_1^2)}$$

3. Inferencia Bayesiana en la normal con parámetros desconocidos

Este caso es mucho más complejo que el caso unidimensional y la solución completa del mismo sale fuera de los objetivos del curso. Nos limitaremos a calcular la distribución a posteriori y el resto de los cálculos lo haremos mediante técnicas de simulación usando **R**.

3.1. La función de verosimilitud

Sea $\mathbf{x}$ una muestra de tamaño n procedente de una normal de media μ y varianza σ^2, ambos parámetros desconocidos. Es bien conocido que

$$f(\mathbf{x}|\mu,\sigma^2) = (2\pi\sigma^2)^{-n/2} \exp\left\{-\frac{\sum_i (x_i-\mu)^2}{2\sigma^2}\right\},$$

donde $x_i \in \mathbb{R}$, $\mu \in \mathbb{R}$ y $\sigma^2 > 0$.

Para ajustarlo a la notación bayesiana vamos a denotar a μ por θ y a σ^2 por ϕ. Tras unos cálculos sencillos se obtiene la función de verosimilitud de ambos parámetros:

$$l(\theta,\phi|\mathbf{x}) = (2\pi\phi)^{-n/2} \exp\left\{-\frac{(n-1)s^2 + n(\bar{x}-\theta)^2}{2\phi}\right\},$$

donde $x_i \in \mathbb{R}$, $\mu \in \mathbb{R}$, $\phi > 0$ y $s^2 = \frac{\sum_i (x_i-\bar{x})^2}{n-1}$.

3.2. Distribución a priori y a posteriori

Definición 3.1. *Si X es una variable aleatoria e Y otra variable aleatoria no negativa, se dice que el par (X,Y) sigue una distribución Normal-Gamma Invertida de parámetros $(m,c;a,b)$, y se denota como $(X,Y) \rightsquigarrow NGaI(m,c;a,b)$ si*

1. $Y \rightsquigarrow GaI(a,b)$;
2. $X|Y = y \rightsquigarrow N(m,yc)$.

3.2.1. A priori de Jeffrey

La distribución a priori de Jeffrey se define como

$$\pi(\theta_1,\theta_2) \propto \sqrt{det(\mathcal{I}(\theta_1,\theta_2))},$$

donde $\mathcal{I}(\theta_1,\theta_2)$ es la matriz de información de Fisher verificándose que

$$\mathcal{I}_{ij}(\theta_1,\theta_2) = -E\left[\frac{\partial^2}{\partial\theta_i\partial\theta_j}\log f(X|\theta_1,\theta_2)\right].$$

Después de un cálculo sencillo se obtiene que $\pi(\theta,\phi) \propto 1/\phi^{3/2}$. Por tanto, se tiene que

$$\pi(\theta,\phi|\mathbf{x}) \propto \phi^{-(n+3)/2} \exp\left(-\frac{(n-1)s^2}{2\phi}\right) \times \exp\left(-\frac{n(\theta-\bar{x})^2}{2\phi}\right),$$

para $\theta \in \mathbb{R}$ y $\phi \in \mathbb{R}^+$. Es decir,

$$\pi(\theta, \phi|\mathbf{x}) \propto \frac{1}{\phi^{n/2+1}} \exp\left(-\frac{(n-1)s^2}{2\phi}\right) \times \frac{1}{\phi^{1/2}} \exp\left(-\frac{n(\theta-\bar{x})^2}{2\phi}\right),$$

para $\theta \in \mathbb{R}$ y $\phi \in \mathbb{R}^+$.
Luego,

$$\boxed{\theta, \phi|\mathbf{x} \sim NGaI\left(\bar{x}, \frac{1}{n}; \frac{n}{2}, \frac{(n-1)s^2}{2}\right).}$$

Esto nos permite obtener las dos marginales a posteriori. Es decir, se tiene que

$$\phi|\mathbf{x} \sim GaI\left(\frac{n}{2}, \frac{(n-1)s^2}{2}\right).$$

Utilizando las propiedades de la densidad Gamma vistas en el tema 4, se llega a que

$$\frac{(n-1)s^2}{\phi}|\mathbf{x} \sim \chi^2_n.$$

La marginal a posteriori para θ requiere de un cálculo un poco más laborioso. Se tiene que

$$\pi(\theta|\mathbf{x}) = \int_0^{+\infty} \pi(\theta, \phi|\mathbf{x}) d\phi.$$

Por tanto, después de eliminar constantes quedaría

$$\pi(\theta|\mathbf{x}) \propto \int_0^{+\infty} \frac{1}{\phi^{\frac{n+1}{2}+1}} e^{-\frac{1}{\phi}\lambda^*} d\phi,$$

donde $\lambda^* = \frac{1}{2}\left[n(\theta-\bar{x})^2 + nS^2\right]$ y $nS^2 = (n-1)s^2$. Esta útlima integral se corresponde con una gamma-invertida, por tanto se obtiene que

$$\pi(\theta|\mathbf{x}) \propto \left[1 + \frac{z^2}{n}\right]^{-\frac{n+1}{2}}, \forall z \in \mathbb{R} \text{ donde } z = \sqrt{n}\frac{(\theta-\bar{x})}{S}.$$

Lo cual implica que

$$Z|\mathbf{x} \sim t_n.$$

Nota 1. *Para obtener una muestra de la distribución a posteriori bastará con simular una muestra de $\phi|\mathbf{x}$ y para cada valor de esta muestra se simula un valor de la distribución de $\theta|\mathbf{x}, \phi$.*

Nota 2. *Los estimadores bayesianos y las regiones de credibilidad se pueden proponer para cada parámetro usando sus densidades a posteriori marginales.*

Nota 3. *Si bien puede resultar interesante el uso de la distribución predictiva en este modelo, cuyo cálculo es similar al correspondiente a la marginal de θ a posteriori, no es objetivo del temario en este curso.*

3.2.2. A priori normal-gamma-invertida

En este caso suponemos que $(\theta, \phi) \sim NGaI(m, c; a, b)$. Con cálculos similares a los anteriores, se tiene que

$$\boxed{\theta, \phi|\mathbf{x} \sim NGaI\left(\frac{m + cn\bar{x}}{nc+1}, \frac{c}{nc+1}; a + \frac{n}{2}, b + \frac{(n-1)s^2}{2} + \frac{1}{2}\frac{n(\bar{x}-m)^2}{1+nc}\right)}$$

A partir de aquí el cálculo de las densidades a posteriori marginales es similar al apartado anterior.

Ejercicios de autoaprendizaje.

1. Sea $X|\sigma^2 \sim N(\mu_0, \sigma^2)$ con valor conocido para la media. Se pide proponer una a priori de Jeffrey para la varianza y su correspondiente a posteriori.

2. Sea X una v.a. con distribución Gamma-Invertida de parámetros (a, b). Transformar la v.a. X en una distribución Chi-cuadrado y calcular sus grados de libertad.

3. Demostrar que en el caso de a posteriori normal para ambos parámetros desconocidos, la distribución a posteriori marginal de θ está relacionada con una distribución t-Student usando una distribución a priori dada por $\pi(\theta, \phi) \propto \frac{1}{\phi}$.