Universidad Santo Tomás

División de Ciencias Económicas y Administrativas

Departamento de Estadística

Simulación y cálculo numérico

By: **Hanwen Zhang**

hanwenzhang@usantotomas.edu.co

Índice

1. Normal con media desconocida y varianza conocida

En esta última sección de este capítulo, se considera datos que pueden ser descritos adecuadamente con la distribución normal, la cual a diferencia de las anteriores distribuciones consideradas tiene dos parámetros. En el siguiente capítulo se considera el caso general cuando ambos parámetros son desconocidos. En esta parte, se asume que la varianza teórica es conocica, y el objetivo es estimar la media teórica.

Suponga que $Y_1, \cdots, Y_n$ son variables independientes e idénticamente distribuidos con distribución $Normal(\theta, \sigma^2)$ con θ desconcido pero σ^2 conocido. De esta forma, la función de verosimilitud de los datos está dada por

$$\begin{aligned} p(\mathbf{Y} \mid \theta) &= \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{1}{2\sigma^2}(y_i - \theta)^2\right\} I_{\mathbb{R}}(y) \\ &= (2\pi\sigma^2)^{-n/2} \exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i - \theta)^2\right\} \end{aligned}$$

Como el parámetro θ puede tomar cualquier valor en los reales, es posible asignarle una distribución previa $\theta \sim Normal(\mu, \tau^2)$. Bajo este marco de referencia se tienen los siguientes resultados

Result 1. *La distribución posterior del parámetro de interés θ sigue una distribución*

$$\theta|\mathbf{Y} \sim Normal(\mu_n, \tau_n^2).$$

En donde

$$\mu_n = \frac{\frac{n}{\sigma^2}\bar{Y} + \frac{1}{\tau^2}\mu}{\frac{n}{\sigma^2} + \frac{1}{\tau^2}} \qquad y \qquad \tau_n^2 = \left(\frac{n}{\sigma^2} + \frac{1}{\tau^2}\right)^{-1} \tag{1}$$

Demostración.

$$\begin{aligned}
p(\theta \mid \mathbf{Y}) &\propto p(\mathbf{Y} \mid \theta) p(\theta \mid \mu, \tau^2) \\
&\propto \exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i - \theta)^2 - \frac{1}{2\tau^2}(\theta - \mu)^2\right\} \\
&= \exp\left\{-\frac{1}{2}\left[\frac{\sum_{i=1}^{n}(y_i - \theta)^2}{\sigma^2} + \frac{(\theta - \mu)^2}{\tau^2}\right]\right\} \\
&\propto \exp\left\{-\frac{1}{2}\left[\frac{n\theta^2}{\sigma^2} - \frac{2\theta\sum_{i=1}^{n} y_i}{\sigma^2} + \frac{\theta^2}{\tau^2} - \frac{2\theta\mu}{\tau^2}\right]\right\} \\
&= \exp\left\{-\frac{\theta^2}{2}\left[\frac{n}{\sigma^2} + \frac{1}{\tau^2}\right] + \theta\left[\frac{n\bar{y}}{\sigma^2} + \frac{\mu}{\tau^2}\right]\right\} \\
&= \exp\left\{-\frac{\theta^2}{2\tau_n^2} + \frac{\theta\mu_n}{\tau_n^2}\right\} \\
&= \exp\left\{-\frac{1}{2\tau_n^2}(\theta^2 - 2\theta\mu_n)\right\} \\
&\propto \exp\left\{-\frac{1}{2\tau_n^2}(\theta^2 - 2\theta\mu_n + \mu_n^2)\right\} \\
&= \exp\left\{-\frac{1}{2\tau_n^2}(\theta - \mu_n)^2\right\}
\end{aligned}$$

Por lo tanto, se encuentra una expresión idéntica a la función de distribución de una variable aleatoria con distribución $Normal(\mu_n, \tau_n^2)$. □

Observando la forma de μ_n, que corresponde a la estimación bayesiana del parámetro θ, podemos concluir que éste es una combinación convexa entre el estimador clásico de máxima verosimlitud $\hat{\theta}_C = \bar{y}$ y el estimador previo $\hat{\theta}_P = \mu$, puesto que:

$$\begin{aligned}
\hat{\theta}_B = \mu_n &= \frac{\frac{n}{\sigma^2}\bar{Y} + \frac{1}{\tau^2}\mu}{\frac{n}{\sigma^2} + \frac{1}{\tau^2}} \\
&= \frac{\frac{n}{\sigma^2}}{\frac{n}{\sigma^2} + \frac{1}{\tau^2}}\bar{Y} + \frac{\frac{1}{\tau^2}}{\frac{n}{\sigma^2} + \frac{1}{\tau^2}}\mu \\
&= \frac{\frac{n}{\sigma^2}}{\frac{n}{\sigma^2} + \frac{1}{\tau^2}}\hat{\theta}_C + \frac{\frac{1}{\tau^2}}{\frac{n}{\sigma^2} + \frac{1}{\tau^2}}\hat{\theta}_P
\end{aligned}$$

De donde se puede concluir que para una distribución previa fija, entre mayor sea el tamaño muestral n, más peso tendrá el estimador clásico $\hat{\theta}_C$ en el cálculo del estimador bayesiano. De la misma forma, para un conjunto fijo de datos $\mathbf{Y}$, entre menor sea la varianza previa, τ^2, más certeza tenemos sobre la información previa y por consiguiente la estimación bayesiana μ_n se acercará más a la estimación previa. En la Figura 1 se observa la función de densidad previa, función de verosimilitud y función de densidad posterior con $\mu = 5$, $\tau^2 = 0.01$, $\bar{y} = 2$, $\sigma^2 = 1$ y $n = 5, 10, 50, 200$. Podemos observar que a medida que el tamaño muestral n aumente, la función de verosimilitud (vista como la función del parámetro θ) se vuelve más concentrada alrededor del valor de $\bar{y}$, y a consecuencia, la función de densidad posterior de θ se sitúa más cercana a la función de verosimilitud, y la estimación bayesiana se acerca más a la estimación clásica $\bar{y}$.

Distribución previa no informativa para θ

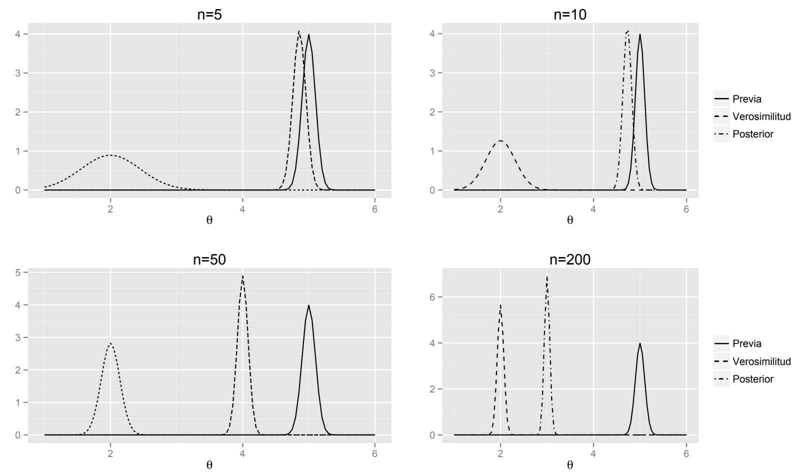

Figura 1: *Distribución previa, función de verosimilitud y distribución posterior del parámetro* θ *con* $\mu = 5$, $\tau^2 = 0.01$, $\bar{y} = 2$, $\sigma^2 = 1$ *y* $n = 5, 10, 50, 200$.

Por otro lado, nótese que en el caso en donde se desconozca el comportamiento estructural de θ, es posible hacer su distribución previa tan plana y vaga como sea posible. Para esto, basta con hacer tender al parámetro de precisión de la distribución previa hacia infinito. Es decir $\tau^2 \longrightarrow \infty$, en este caso, la distribución previa de θ corresponde a una distribución impropia, $p(\theta) \propto cte$. Se puede ver que bajo esta distribución previa, la distribución posterior tendería a una $Normal(\bar{y}, \sigma^2/n)$ (Ejercicio XXXX).

La anterior idea intuitiva de usar la distribución previa $p(\theta) \propto cte$ para representar la falta de la información previa corresponde a la previa no informativa de Jeffreys, puesto que la información de Fisher del parámetro θ en una variable con distribución normal está dada por

$$I(\theta) = 1/\sigma^2$$

De donde se puede concluir qua la previa no informativa de Jeffreys está dada por

$$p(\theta) \propto 1/\sigma \propto cte$$

Finalmente, comparemos los resultados inferenciales obtenidos con la previa no informativa de Jeffreys con el enfoque inferencial clásico en términos de la estimación puntua y el intervalo de credibilidad y de confianza.

- En cuanto a la estimación puntual, es claro que ambos enfoques conducen al mismo estimador $\hat{\theta} = \bar{Y}$.

- Con respecto al intervalo para el parámetro θ, al usar el enforque bayesiano con la previa no informativa de Jeffreys, un intervalo de credibilidad de $(1-\alpha) \times 100\,\%$ están dadas por los percentiles $\alpha/2$ y $1-\alpha/2$ de la distribución posterior de θ: $Normal(\bar{y}, \sigma^2/n)$, denotaremos estos percentiles como a y b, respectivamente. Por definición tenemos que, si $X \sim N(\bar{y}, \sigma^2/n)$,

$$\begin{aligned}
\alpha/2 &= Pr(X < a) \\
&= Pr(\frac{X-\bar{y}}{\sigma/\sqrt{n}} < \frac{a-\bar{y}}{\sigma/\sqrt{n}}) \\
&= Pr(Z < \frac{a-\bar{y}}{\sigma/\sqrt{n}})
\end{aligned}$$

  Estos es, $\frac{a-\bar{y}}{\sigma/\sqrt{n}}$ es el percentil $\alpha/2$ de la distribución normal estándar $z_{\alpha/2}$ ó equivalentemente $-z_{1-\alpha/2}$. De esta forma, tenemos que $a = bary - z_{1-\alpha/2}\sigma/\sqrt{n}$. Análogamente tenemos que $b = bary + z_{1-\alpha/2}\sigma/\sqrt{n}$, y podemos concluir que un intervalo de crediblidad de $(1-\alpha)\times 100\,\%$ está dada por $\bar{y} \pm z_{1-\alpha/2}\sigma/\sqrt{n}$, el cual coincide con el intervalo de confianza para θ usando el enfoque de la inferencia clásica (ver Zhang & Gutiérrez (2010)).

Diferentes formas de hallar la distribución previa para θ

En primer lugar, consideramos el caso cuando la información previa se encuentra en un conjunto de datos $x_1, \cdots, x_m$ que corresponden a mediciones de la variable de estudio Y en otro punto de tiempo, en otro punto geográfico, o inclusive en otra población de estudio. En este caso, podemos tomar la media de la distribución previa μ como $\bar{X}$ y la varianza de la distribución previa τ^2 como S_X^2.

En el caso de que no se dispongan de datos como información previa, sino que Ãl’sta está contenida en alguna estimaciÃşn que se haya realizado sobre θ. Por ejemplo, si se dispone de algún modelamiento estadístico que se haya hecho previamente sobre θ, podemos fácilmente obtener el valor estimado de θ y el error estándar de esta estimaciÃşn, y naturalmente, estos dos valores serían nuestros parámetros de la distribuciÃşn previa: μ y τ^2.

Finalmente, si la estimación previa de θ se presentada en forma de un intervalo, por ejemplo, si se sabe que un intervalo de confianza para θ es $(15.3,\ 24.7)$, entonces podemos usar μ como el punto medio de este intervalo, es decir, $\mu = 20$ y para escoger el valor de τ^2, se tiene en cuenta que en muchas ramas de la estadística, un intervalo de confianza se puede aproximar por $\hat{\theta} \pm 2\sqrt{var(\hat{\theta})}$. De esta forma, podemos usar $\tau^2 = \left(\frac{24.7-20}{2}\right)^2 \approx 5.5$

Distribuciones predictivas

Ahora, se discute las distribuciones predictivas previa y predictiva posterior para una observación o una nueva muestra.

Result 2. *La distribución predictiva previa para una observación y es*

$$y \sim Normal(\mu, \tau^2 + \sigma^2)$$

Demostración. De la definición de función de distribución predictiva se tiene que

$$\begin{aligned} p(Y) &= \int p(Y \mid \theta) p(\theta \mid \mu, \tau^2) \; d\theta \\ &= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{1}{2\sigma^2}(y-\theta)^2\right\} \frac{1}{\sqrt{2\pi\tau^2}} \exp\left\{-\frac{1}{2\tau^2}(\theta-\mu)^2\right\} d\theta \end{aligned}$$

(Berger 1985) desarrolló las siguientes igualdades

$$
\begin{aligned}
&\frac{1}{2}\left[\frac{(\theta-\mu)^2}{\tau^2}+\frac{(y-\theta)^2}{\sigma^2}\right]\\
&=\frac{1}{2}\left[\left(\frac{1}{\tau^2}+\frac{1}{\sigma^2}\right)\theta^2-2\left(\frac{\mu}{\tau^2}+\frac{y}{\sigma^2}\right)\theta+\left(\frac{\mu^2}{\tau^2}+\frac{y^2}{\sigma^2}\right)\right]\\
&=\frac{1}{2\tau_1^2}\left[\theta^2-2\tau_1^2\left(\frac{\mu}{\tau^2}+\frac{y}{\sigma^2}\right)\theta+\tau_1^4\left(\frac{\mu}{\tau^2}+\frac{y}{\sigma^2}\right)^2\right]+\frac{1}{2}\left(\frac{\mu^2}{\tau^2}+\frac{y^2}{\sigma^2}\right)-\frac{\tau_1^2}{2}\left(\frac{\mu}{\tau^2}+\frac{y}{\sigma^2}\right)^2\\
&=\frac{1}{2\tau_1^2}\left[\theta-\tau_1^2\left(\frac{\mu}{\tau^2}+\frac{y}{\sigma^2}\right)\right]^2+\frac{1}{2}\left[\left(\frac{1}{\sigma^2}-\frac{\tau_1^2}{\sigma^4}\right)y^2-2\frac{\mu\tau_1^2}{\tau^2\sigma^2}y+\left(\frac{\mu^2}{\tau^2}-\frac{\mu^2\tau_1^2}{\tau^4}\right)\right]\\
&=\frac{1}{2\tau_1^2}\left[\theta-\mu_1\right]^2+\frac{1}{2}\left[\frac{1}{\sigma^2+\tau^2}y^2-2\frac{\mu}{\sigma^2+\tau^2}y+\frac{\mu^2}{\sigma^2+\tau^2}\right]\\
&=\frac{1}{2\tau_1^2}\left[\theta-\mu_1\right]^2+\frac{1}{2(\sigma^2+\tau^2)}(y-\mu)^2.
\end{aligned}
$$

Entonces

$$
\begin{aligned}
p(Y)&=\int_{-\infty}^{\infty}\frac{1}{2\pi\sigma\tau}\exp\left\{-\frac{1}{2\tau_1^2}(\theta-\mu_1)^2\right\}\exp\left\{-\frac{1}{2(\tau^2+\sigma^2)}(y-\mu)^2\right\}d\theta\\
&=\frac{1}{\sqrt{2\pi\frac{\sigma^2\tau^2}{\tau_1^2}}}\exp\left\{-\frac{1}{2(\tau^2+\sigma^2)}(y-\mu)^2\right\}\int_{-\infty}^{\infty}\frac{1}{\sqrt{2\pi\tau_1^2}}\exp\left\{-\frac{1}{2\tau_1^2}(\theta-\mu_1)^2\right\}d\theta\\
&=\frac{1}{\sqrt{2\pi(\tau^2+\sigma^2)}}\exp\left\{-\frac{1}{2(\tau^2+\sigma^2)}(y-\mu)^2\right\}
\end{aligned}
$$

□

Una vez recolectados los datos $\mathbf{Y}=\{Y_1,\cdots,Y_n\}$, se obtiene la distribución predictiva posterior dada en el siguiente resultado. La demostración es similar al del resultado anterior, y se deja como ejercicio para los lectores.

Result 3. *La distribución predictiva posterior para una nueva observación* $\tilde{y}$

es

$$\tilde{y} \mid \mathbf{Y} \sim Normal(\mu_n, \tau_n^2 + \sigma^2)$$

Demostración. Se deja como ejercicio para los lectores. $\square$

En algunas situaciones, se quiere conocer el comportamiento probabilístico de más de una nueva observación, digamos $Y_1^*, \cdots, Y_{n^*}^*$, en este caso, lo ideal sería obtener la distribución conjunta predictiva posterior de la nueva muestra, $p(Y_1^*, \cdots, Y_{n^*}^* | \mathbf{Y})$. Sin embargo, esta distribución no es fácil de hallar, y procedemos a hallar la distribución predictiva posterior de la media de esta nueva muestra $\bar{Y}^*$, la cual es dada en el siguiente resultado.

Result 4. *La distribución predictiva posterior para la media muestral $\bar{Y}^*$ de una nueva muestra es*

$$\bar{Y}^* | \mathbf{Y} \sim N\left(\mu_n, \frac{\sigma^2}{n^*} + \tau_n^2\right)$$

donde μ_n y τ_n^2 fueron definidos en (1).

Demostración.

$$
\begin{aligned}
p(\bar{Y}^*|\mathbf{Y}) &= \int_{-\infty}^{\infty} p(\bar{Y}^*|\theta)p(\theta|\mathbf{Y})\ d\theta \\
&= \int_{-\infty}^{\infty} (2\pi\frac{\sigma^2}{n^*})^{-1/2} \exp\left\{-\frac{n^*}{2\sigma^2}(\bar{y}^*-\theta)^2\right\} (2\pi\tau_n^2)^{-1/2} \exp\left\{-\frac{1}{2\tau_n^2}(\theta-\mu_n)^2\right\}\ d\theta \\
&= \int_{-\infty}^{\infty} (2\pi)^{-1}(\frac{\sigma^2}{n^*}\tau_n^2)^{-1/2} \exp\left\{-\frac{1}{2}\left[\frac{(\bar{y}^*-\theta)^2}{\sigma^2/n^*}+\frac{(\theta-\mu_n)^2}{\tau_n^2}\right]\right\}\ d\theta \\
&= \underbrace{\int_{-\infty}^{\infty} (2\pi\frac{1}{n^*/\sigma^2+1/\tau_n^2})^{-1/2} \exp\left\{-\frac{1}{2}\left(\frac{n^*}{\sigma^2}+\frac{1}{\tau_n^2}\right)\left(\theta-\frac{\bar{y}^*/(\sigma^2/n^*)+\mu_n/\tau_n^2}{n^*/\sigma^2+1/\tau_n^2}\right)^2\right\}\ d\theta}_{\text{igual a 1}} \\
&\quad (2\pi)^{-1/2}(\frac{\sigma^2}{n^*}\tau_n^2)^{-1/2}(\frac{n^*}{\sigma^2}+\frac{1}{\tau_n^2})^{-1/2} \exp\left\{-\frac{1}{2(\sigma^2/n^*+\tau_n^2)}(\bar{y}^*-\mu_n)^2\right\} \\
&= (2\pi)^{-1/2}(\frac{\sigma^2}{n^*}\tau_n^2)^{-1/2}(\frac{n^*}{\sigma^2}+\frac{1}{\tau_n^2})^{-1/2} \exp\left\{-\frac{1}{2(\sigma^2/n^*+\tau_n^2)}(\bar{y}^*-\mu_n)^2\right\} \\
&= (2\pi)^{-1/2}(\frac{\sigma^2}{n^*}+\tau_n^2)^{-1/2} \exp\left\{-\frac{1}{2(\sigma^2/n^*+\tau_n^2)}(\bar{y}^*-\mu_n)^2\right\}
\end{aligned}
$$

□

Del anterior resultado, podemos ver que (1) la esperanza de la distribución de $\bar{Y}^*|\mathbf{Y}$ es igual a la esperanza de $\theta|\mathbf{Y}$, y (2) a diferencia de la varianza de $\theta|\mathbf{Y}$, la varianza de $\bar{Y}^*|\mathbf{Y}$ tiene un componente adicional: σ^2/n^*, de esta forma, tenemos tres fuentes de incertidumbre al momento de pronosticar $\bar{Y}^*$: la incertidumbre en la información previa, la incertidumbre en la muestra observada y la incertidumbre en la nueva muestra.

Referencias

Berger, J. O. (1985), *Statistical Decision Theory and Bayesian Analysis*, 2 edn, Springer.

Zhang, H. & Gutiérrez, H. A. (2010), *Teoría estadística. Aplicación y métodos.*, Universidad Santo Tomás.