CAPÍTULO 8. MODELOS LINEALES Y REGRESIÓN

Para leer

Gelman et al, Capítulo 8

Se supone que $\mathbf{Y}$ es un vector de n observaciones. Se define un modelo lineal para $\mathbf{y}$

$$E[\mathbf{y}|\boldsymbol{\theta}] = \mathbf{A}\boldsymbol{\theta}$$

donde $\boldsymbol{\theta}$ es un vector de p parámetros, $\mathbf{A}$ es una matriz conocida de diseño y la matriz de varianza de $\mathbf{y}$ es $\mathbf{C}$. Se define $\mathbf{C}^{-1}$ la matriz de precisión.

Observación 27 *A menudo se supone que la distribución de* $\mathbf{y}$ *es normal.*

También en muchos problemas, la matriz de varianzas tiene una forma simple, por ejemplo $\mathbf{C} = \sigma^2\mathbf{I}$.

Ejemplo 59 *Observaciones univariables*

$$\mathbf{A}^T = (1, \dots, 1), \quad \theta \text{ escalar} \quad \mathbf{C} = \frac{1}{\phi}\mathbf{I} \quad p = 1$$

Ejemplo 60 *Regresión lineal simple.*

El modelo es

$$y_i = \alpha + \beta x_i + \epsilon_i \quad i = 1, \dots, n$$

donde $\epsilon_i \sim \mathcal{N}(0, 1/\phi)$. *Entonces*

$$\mathbf{A}^T = \begin{pmatrix} 1 & 1 & \cdots & 1 \\ x_1 & x_2 & \cdots & x_n \end{pmatrix} \quad \boldsymbol{\theta} = \begin{pmatrix} \alpha \\ \beta \end{pmatrix}$$

y $\mathbf{C} = \frac{1}{\phi}\mathbf{I}$.

Es habitual escribir el modelo de otra forma:

$$y_i = \alpha' + \beta(x_i - \bar{x}) + \epsilon_i$$

donde $\alpha' = \alpha + \beta\bar{x}$.

En este caso se tiene

$$\mathbf{A}^T = \begin{pmatrix} 1 & 1 & \cdots & 1 \\ x_1 - \bar{x} & x_2 - \bar{x} & \cdots & x_n - \bar{x} \end{pmatrix} \quad \boldsymbol{\theta} = \begin{pmatrix} \alpha' \\ \beta \end{pmatrix}$$

Ejemplo 61 *El modelo de dos factores (sin réplicas) Las observaciones son y_{ij} donde $i = 1, \ldots, t$ y $j = 1, \ldots, b$. Entonces hay $n = tb$ observaciones. El modelo es*

$$E[y_{ij}|\boldsymbol{\theta}] = \mu + \alpha_i + \beta_j.$$

Suponiendo la restricción "GLIM" $\alpha_1 = \beta_1 = 0$ (para hacer el modelo identificable), en el caso de $t = 2, b = 3$ se tiene

$$E\begin{bmatrix} y_{11} \\ y_{12} \\ y_{13} \\ y_{21} \\ y_{22} \\ y_{23} \end{bmatrix} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 1 & 1 & 0 & 0 \\ 1 & 1 & 1 & 0 \\ 1 & 1 & 0 & 1 \end{pmatrix} \begin{pmatrix} \mu \\ \alpha_2 \\ \beta_2 \\ \beta_3 \end{pmatrix}$$

La distribución normal multivariante

Para hacer la inferencia se necesita conocer las propiedades de la distribución normal multivariante.

Definición 13 *Un vector* $\mathbf{y}$ *de dimensión* k *tiene una distribución normal multivariante con media* $\boldsymbol{\mu}$ *y varianza* $E[(\mathbf{y}-\boldsymbol{\mu})(\mathbf{y}-\boldsymbol{\mu})^T] = \mathbf{V}$ *si*

$$f(\mathbf{y}|\boldsymbol{\mu}, \mathbf{V}) = (2\pi)^{-\frac{k}{2}} |\mathbf{V}|^{-\frac{1}{2}} \exp\left(-\frac{1}{2}(\mathbf{y}-\boldsymbol{\mu})^T \mathbf{V}^{-1}(\mathbf{y}-\boldsymbol{\mu})\right)$$

Propiedades importantes de la distribución son las siguientes.

- Cualquier subconjunto de $\mathbf{y}$ también se distribuye como normal. Si
$$\mathbf{y} = \begin{pmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \end{pmatrix} \sim \mathcal{N}\left(\begin{matrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{matrix}, \begin{pmatrix} \mathbf{V}_1 & \mathbf{V}_{12} \\ \mathbf{V}_{21} & \mathbf{V}_2 \end{pmatrix} \right)$$
entonces $\mathbf{y}_1 \sim \mathcal{N}(\boldsymbol{\mu}_1, \mathbf{V}_1)$ y $\mathbf{y}_2 \sim \mathcal{N}(\boldsymbol{\mu}_2, \mathbf{V}_2)$.

- Si $\mathbf{y}_1 \sim \mathcal{N}(\boldsymbol{\mu}_1, \mathbf{V}_1)$ y $\mathbf{y}_2 \sim \mathcal{N}(\boldsymbol{\mu}_2, \mathbf{V}_2)$ son variables independientes con la misma dimensión entonces

$$\mathbf{y}_1 \pm \mathbf{y}_2 \sim \mathcal{N}(\boldsymbol{\mu}_1 \pm \boldsymbol{\mu}_2, \mathbf{V}_1 + \mathbf{V}_2)$$

- Si $\mathbf{z} = \mathbf{D}\mathbf{y}$ para una matriz $\mathbf{D}$ entonces

$$\mathbf{z} \sim \mathcal{N}\left(\mathbf{D}\boldsymbol{\mu}, \mathbf{D}\mathbf{V}\mathbf{D}^T\right).$$

Ejemplo 62 *El modelo lineal normal es*

$$\mathbf{y} = \mathbf{A}\boldsymbol{\theta} + \boldsymbol{\epsilon}$$

donde $\boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, \mathbf{C})$. *Entonces* $\mathbf{y}|\boldsymbol{\theta} \sim \mathcal{N}(\mathbf{A}\boldsymbol{\theta}, \mathbf{C})$.

Matrices, vectores y formas cuadráticas

Es necesario entender cómo se manipulan matrices y vectores para hacer los cálculos necesarios para la inferencia.

Unos resultados útiles son

- Para matrices o vectores $\mathbf{A}$ y $\mathbf{B}$,

$$(\mathbf{AB})^T = \mathbf{B}^T\mathbf{A}^T.$$

- Si $\mathbf{x}_i$, $(i = 1, \ldots, n)$ y $\boldsymbol{\mu}$ son vectores de dimensión $k \times 1$ y $\mathbf{V}$ es una matriz simétrica de dimensión $k \times k$,

$$\sum_{i=1}^{n} \mathbf{x}_i \mathbf{V} \boldsymbol{\mu} = n\bar{\mathbf{x}}\mathbf{V}\boldsymbol{\mu}$$

  donde $\bar{\mathbf{x}} = \frac{1}{n}\sum_{i=1}^{n} \mathbf{x}_i$

- Para vectores $\mathbf{x}$, $\boldsymbol{\mu}$ $(k \times 1)$ y matriz simétrica $\mathbf{V}$ una forma cuadrática es

$$(\mathbf{x}-\boldsymbol{\mu})^T \mathbf{V}(\mathbf{x}-\boldsymbol{\mu}).$$

Se tiene la expansión

$$\begin{aligned}(\mathbf{x}-\boldsymbol{\mu})^T \mathbf{V}(\mathbf{x}-\boldsymbol{\mu}) &= \mathbf{x}^T \mathbf{V}\mathbf{x} - \boldsymbol{\mu}^T \mathbf{V}\mathbf{x} - \mathbf{x}^T \mathbf{V}\boldsymbol{\mu} \\ &\quad + \boldsymbol{\mu}^T \mathbf{V}\boldsymbol{\mu} \\ &= \mathbf{x}^T \mathbf{V}\mathbf{x} - 2\mathbf{x}^T \mathbf{V}\boldsymbol{\mu} + \boldsymbol{\mu}^T \mathbf{V}\boldsymbol{\mu}\end{aligned}$$

Se observa que el resultado es escalar.

- Se puede expresar una forma cuadrática de otra manera:

$$(\mathbf{x}-\boldsymbol{\mu})^T \mathbf{V}(\mathbf{x}-\boldsymbol{\mu}) = tr(\mathbf{V}\mathbf{W})$$

donde $\mathbf{W} = (\mathbf{x}-\boldsymbol{\mu})(\mathbf{x}-\boldsymbol{\mu})^T$ y $tr(\cdot)$ es la traza de la matriz.

Ejemplo 63 La verosimilitud para una muestra de una distribución normal multivariable.

Sea $\mathbf{y}_1, \ldots, \mathbf{y}_n$ *una muestra de una distribución normal multivariante* $\mathcal{N}(\boldsymbol{\mu}, \mathbf{V})$. *La verosimilitud es*

$$\begin{aligned}
l(\boldsymbol{\mu}, \mathbf{V}|\mathit{datos}) &\propto |\mathbf{V}|^{-\frac{n}{2}} \exp\left(-\frac{1}{2}\sum_{i=1}^{n}(\mathbf{y}_i - \boldsymbol{\mu})^T \mathbf{V}^{-1}(\mathbf{y}_i - \boldsymbol{\mu})\right) \\
&\propto |\mathbf{V}|^{-\frac{n}{2}} \exp\left(-\frac{1}{2}\sum_{i=1}^{n}(\mathbf{y}_i - \bar{\mathbf{y}} + \bar{\mathbf{y}} - \boldsymbol{\mu})^T \mathbf{V}^{-1}\right. \\
&\qquad \left.(\mathbf{y}_i - \bar{\mathbf{y}} + \bar{\mathbf{y}} - \boldsymbol{\mu})\right) \\
&\propto |\mathbf{V}|^{-\frac{n}{2}} \exp\left(-\frac{1}{2}\left[\sum_{i=1}^{n}(\mathbf{y}_i - \bar{\mathbf{y}})^T \mathbf{V}^{-1}(\mathbf{y}_i - \bar{\mathbf{y}})\right.\right. \\
&\qquad -2\sum_{i=1}^{n}(\mathbf{y}_i - \bar{\mathbf{y}})^T \mathbf{V}^{-1}(\bar{\mathbf{y}} - \boldsymbol{\mu}) + \\
&\qquad \left.\left. n(\bar{\mathbf{y}} - \boldsymbol{\mu})^T \mathbf{V}^{-1}(\bar{\mathbf{y}} - \boldsymbol{\mu})\right]\right)
\end{aligned}$$

$$
\begin{aligned}
&= |\mathbf{V}|^{-\frac{n}{2}} \exp\left(-\frac{1}{2}\left[\sum_{i=1}^{n}(\mathbf{y}_i - \bar{\mathbf{y}})^T\mathbf{V}^{-1}(\mathbf{y}_i - \bar{\mathbf{y}})\right.\right. \\
&\qquad \left.\left. + n(\bar{\mathbf{y}} - \boldsymbol{\mu})^T\mathbf{V}^{-1}(\bar{\mathbf{y}} - \boldsymbol{\mu})\right]\right) \\
&\propto |\mathbf{V}|^{-\frac{n}{2}} \exp\left(-\frac{1}{2}\left[tr(\mathbf{V}^{-1}\mathbf{S})\right.\right. \\
&\qquad \left.\left. + n(\bar{\mathbf{y}} - \boldsymbol{\mu})^T\mathbf{V}^{-1}(\bar{\mathbf{y}} - \boldsymbol{\mu})\right]\right)
\end{aligned}
$$

donde $\mathbf{S} = \sum_{i=1}^{n}(\mathbf{y}_i - \bar{\mathbf{y}})(\mathbf{y}_i - \bar{\mathbf{y}})^T$.

La fórmula para la verosimilitud es parecida a la fórmula en el caso univariable.

Ejemplo 64 Inferencia bayesiana para la distribución normal multivariante

Se puede hacer inferencia conjugada para la distribución normal multivariable. Suponiendo $\mathbf{V}$ *conocida y con la distribución a priori*

$$\boldsymbol{\mu} \sim \mathcal{N}\left(\mathbf{m}, \frac{1}{\alpha}\mathbf{V}\right)$$

entonces, la distribución a posteriori es también normal

$$\begin{aligned} \boldsymbol{\mu}|\mathbf{y} &\sim \mathcal{N}\left(\mathbf{m}^*, \frac{1}{\alpha^* V}\right) \quad \textit{donde} \\ \alpha^* &= \alpha + n \\ \mathbf{m}^* &= \frac{\alpha\mathbf{m} + n\bar{\mathbf{y}}}{\alpha + n} \end{aligned}$$

Demostración

$$
\begin{aligned}
f(\boldsymbol{\mu}|\mathbf{y}) &\propto l(\boldsymbol{\mu}|\mathbf{y})f(\boldsymbol{\mu}) \\
&\propto \exp\left(-\frac{1}{2}\left[n(\boldsymbol{\mu}-\bar{\mathbf{y}})^T\mathbf{V}^{-1}(\boldsymbol{\mu}-\bar{\mathbf{y}})+\right.\right. \\
&\qquad \left.\left.\alpha(\boldsymbol{\mu}-\mathbf{m})^T\mathbf{V}^{-1}(\boldsymbol{\mu}-\mathbf{m})\right]\right) \\
&\propto \exp\left(-\frac{1}{2}\left[n\boldsymbol{\mu}^T\mathbf{V}^{-1}\boldsymbol{\mu}-2n\boldsymbol{\mu}^T\mathbf{V}^{-1}\bar{\mathbf{y}}+\right.\right. \\
&\qquad \left.\left.\alpha\boldsymbol{\mu}^T\mathbf{V}^{-1}\boldsymbol{\mu}-2\alpha\boldsymbol{\mu}^T\mathbf{V}^{-1}\mathbf{m}\right]\right) \\
&\propto \exp\left(-\frac{1}{2}\left[(\alpha+n)\boldsymbol{\mu}^T\mathbf{V}^{-1}\boldsymbol{\mu}-\right.\right. \\
&\qquad \left.\left.2(\alpha+n)\boldsymbol{\mu}^T\mathbf{V}^{-1}\frac{\alpha\mathbf{m}+n\bar{\mathbf{y}}}{\alpha+n}\right]\right) \\
&\propto \exp\left(-\frac{1}{2}\left[\alpha^*\boldsymbol{\mu}^T\mathbf{V}^{-1}\boldsymbol{\mu}-2\alpha^*\boldsymbol{\mu}^T\mathbf{V}^{-1}\mathbf{m}^*\right]\right) \\
&\propto \exp\left(-\frac{\alpha^*}{2}\left[(\boldsymbol{\mu}-\mathbf{m}^*)^T\mathbf{V}^{-1}(\boldsymbol{\mu}-\mathbf{m}^*)\right]\right) \\
&\propto \exp\left(-\frac{1}{2}(\boldsymbol{\mu}-\mathbf{m}^*)^T\left(\frac{1}{\alpha^*}\mathbf{V}\right)^{-1}(\boldsymbol{\mu}-\mathbf{m}^*)\right)
\end{aligned}
$$

que es el núcleo de una distribución normal. $\diamond$

Con una distribución a priori uniforme se tiene

$$\boldsymbol{\mu}|\mathbf{y} \sim \mathcal{N}\left(\bar{\mathbf{y}}, \frac{1}{n}\mathbf{V}\right)$$

y la media a posteriori de μ coincide con el EMV.

El caso de $\mathbf{V}$ desconocida.

En este caso, la distribución a priori conjugada para $\mathbf{V}$ es una distribución Wishart invertida, $\mathbf{V} \sim \mathcal{WI}(\nu, \mathbf{W})$, es decir

$$f(\mathbf{V}) \propto |\mathbf{V}|^{\frac{\nu+p+1}{2}} \exp\left(-\frac{1}{2}tr\left(\mathbf{V}^{-1}\mathbf{W}\right)\right)$$

donde $p = dim(\mathbf{V})$.

Teorema 10 *Si* $\mathbf{Y}|\boldsymbol{\mu},\mathbf{V} \sim \mathcal{N}(\boldsymbol{\mu},\mathbf{V})$ *con distribución a priori* $\boldsymbol{\mu}|\mathbf{V} \sim \mathcal{N}\left(\mathbf{m},\frac{1}{\alpha}\mathbf{V}\right)$ *y* $\mathbf{V} \sim \mathcal{WI}(\nu,\mathbf{W})$, *luego* $\boldsymbol{\mu}|\mathbf{y},\mathbf{V} \sim \mathcal{N}\left(\mathbf{m}^*,\frac{1}{\alpha^*}\mathbf{V}\right)$ *y* $\mathbf{V}|\mathbf{y} \sim \mathcal{WI}(\nu^*,\mathbf{W}^*)$ *donde*

$$\begin{aligned}
\alpha^* &= \alpha + n \\
\mathbf{m}^* &= \frac{\alpha\mathbf{m} + n\bar{\mathbf{y}}}{\alpha + n} \\
\nu^* &= \nu + n \\
\mathbf{W}^* &= \mathbf{W} + \mathbf{S} + \frac{\alpha n}{\alpha + n}(\bar{\mathbf{y}} - \mathbf{m})(\bar{\mathbf{y}} - \mathbf{m})^T
\end{aligned}$$

El resultado es similar al resultado univariable.

Inferencia para el modelo lineal con varianza conocida.

Supongamos el modelo básico $\mathbf{y}|\boldsymbol{\theta} \sim \mathcal{N}(\mathbf{A}\boldsymbol{\theta}, \mathbf{C})$ y una distribución a priori normal $\boldsymbol{\theta} \sim \mathcal{N}(\boldsymbol{\mu}, \mathbf{V})$ donde $\mathbf{A}$, $\mathbf{C}$, $\boldsymbol{\mu}$ y $\mathbf{V}$ son conocidos.

Este modelo se llama **el modelo lineal de 2 etapas**.

Tenemos el siguiente teorema

Teorema 11 *La distribución marginal de* $\mathbf{y}$ *es*

$$\mathbf{y} \sim \mathcal{N}\left(\mathbf{A}\boldsymbol{\mu}, \mathbf{C} + \mathbf{A}\mathbf{V}\mathbf{A}^T\right)$$

y la distribución a posteriori de $\boldsymbol{\theta}$ *es* $\boldsymbol{\theta}|\mathbf{y} \sim \mathcal{N}(\mathbf{B}\mathbf{b}, \mathbf{B})$ *donde*

$$\begin{aligned} \mathbf{B}^{-1} &= \mathbf{V}^{-1} + \mathbf{A}^T\mathbf{C}^{-1}\mathbf{A} \\ \mathbf{b} &= \mathbf{A}^T\mathbf{C}^{-1}\mathbf{y} + \mathbf{V}^{-1}\boldsymbol{\mu} \end{aligned}$$

Demostración

En primer lugar se calcula la distribución marginal de $\mathbf{y}$*.*

Es suficiente calcular la media y la varianza.

$$\begin{aligned} E[\mathbf{y}] &= E[E[\mathbf{y}|\boldsymbol{\theta}]] \\ &= E[\mathbf{A}\boldsymbol{\theta}] \\ &= \mathbf{A}E[\boldsymbol{\theta}] = \mathbf{A}\boldsymbol{\mu} \\ V[\mathbf{y}] &= E[V[\mathbf{y}|\boldsymbol{\theta}]] + V[E[\mathbf{y}|\boldsymbol{\theta}]] \\ &= E[\mathbf{C}] + V[\mathbf{A}\boldsymbol{\theta}] \\ &= \mathbf{C} + \mathbf{A}V[\boldsymbol{\theta}]\mathbf{A}^T \\ &= \mathbf{C} + \mathbf{AVA}^T \end{aligned}$$

La normalidad sigue imediatamente recordandoo que $\mathbf{y} = \mathbf{A}\boldsymbol{\theta} + \boldsymbol{\epsilon}$ *donde* $\boldsymbol{\theta}$ *y* $\boldsymbol{\epsilon}$ *son normales.*

Ahora se calcula la distribución a posteriori

$$
\begin{aligned}
f(\boldsymbol{\theta}|\mathbf{y}) &\propto f(\boldsymbol{\theta})f(\mathbf{y}|\boldsymbol{\theta}) \\
&\propto \exp\left(-\frac{1}{2}(\boldsymbol{\theta}-\boldsymbol{\mu})^T\mathbf{V}^{-1}(\boldsymbol{\theta}-\boldsymbol{\mu})\right) \\
&\quad \exp\left(-\frac{1}{2}(\mathbf{y}-\mathbf{A}\boldsymbol{\theta})^T\mathbf{C}^{-1}(\mathbf{y}-\mathbf{A}\boldsymbol{\theta})\right) \\
&\propto \exp\left(-\frac{1}{2}\left[\boldsymbol{\theta}^T(\mathbf{V}^{-1}+\mathbf{A}^T\mathbf{C}^{-1}\mathbf{A})\boldsymbol{\theta}-\right.\right. \\
&\quad \left.\left. 2\boldsymbol{\theta}^T(\mathbf{V}^{-1}\boldsymbol{\mu}+\mathbf{A}^T\mathbf{C}^{-1}\mathbf{y})\right]\right) \\
&\propto \exp\left(-\frac{1}{2}\left[\boldsymbol{\theta}^T\mathbf{B}^{-1}\boldsymbol{\theta}-2\boldsymbol{\theta}^T\mathbf{b}\right]\right) \\
&\propto \exp\left(-\frac{1}{2}\left[\boldsymbol{\theta}^T\mathbf{B}^{-1}\boldsymbol{\theta}-2\boldsymbol{\theta}^T\mathbf{B}^{-1}\mathbf{B}\mathbf{b}\right]\right) \\
&\propto \exp\left(-\frac{1}{2}(\boldsymbol{\theta}-\mathbf{B}\mathbf{b})^T\mathbf{B}^{-1}(\boldsymbol{\theta}-\mathbf{B}\mathbf{b})\right)
\end{aligned}
$$

que es el núcleo de una distribución normal $\mathcal{N}(\mathbf{Bb},\mathbf{B})$. $\diamond$

Relación con el EMV

El estimador de mínimos cuadrados para este problema es

$$\widehat{\boldsymbol{\theta}} = \left(\mathbf{A}^T\mathbf{C}^{-1}\mathbf{A}\right)^{-1}\mathbf{A}^T\mathbf{C}^{-1}\mathbf{y}$$

y entonces

$$\begin{aligned}\mathbf{Bb} &= \left(\mathbf{V}^{-1}+\mathbf{A}^T\mathbf{C}^{-1}\mathbf{A}\right)^{-1}\left(\mathbf{A}^T\mathbf{C}^{-1}\mathbf{y}+\mathbf{V}^{-1}\boldsymbol{\mu}\right)\\ &= \left(\mathbf{V}^{-1}+\mathbf{A}^T\mathbf{C}^{-1}\mathbf{A}\right)^{-1}\left[\left(\mathbf{A}^T\mathbf{C}^{-1}\mathbf{A}\right)\widehat{\boldsymbol{\theta}}+\mathbf{V}^{-1}\boldsymbol{\mu}\right]\end{aligned}$$

La media a posteriori es una media ponderada de la media a priori $\boldsymbol{\mu}$ y el EMV $\widehat{\boldsymbol{\theta}}$ con ponderaciones proporcionales a las matrices de precisión.

Se puede expresar la media a posteriori de otra manera.

$$\begin{aligned}\mathbf{Bb} &= \mathbf{B}\left[\mathbf{A}^T\mathbf{C}^{-1}(\mathbf{y}-\mathbf{A}\boldsymbol{\mu}) + \left(\mathbf{A}^T\mathbf{C}^{-1}\mathbf{A}+\mathbf{V}^{-1}\right)\boldsymbol{\mu}\right] \\ &= \boldsymbol{\mu} + \mathbf{B}\mathbf{A}^T\mathbf{C}^{-1}(\mathbf{y}-\mathbf{A}\boldsymbol{\mu})\end{aligned}$$

Se ha expresado la media a posteriori como la media a priori más una corrección.

La cantidad $\mathbf{y}-\mathbf{A}\boldsymbol{\mu}$ es la diferencia entre la observación $\mathbf{y}$ y su esperanza a priori $\mathbf{A}\boldsymbol{\mu}$. La cantidad $\mathbf{B}\mathbf{A}^T\mathbf{C}^{-1}$ se llama **un filtro**

Ejemplo 65 *Retomando el Ejemplo 59 con observaciones univariables y* $\mathbf{A} = (1, \ldots, 1)^T$, $\mathbf{C} = \frac{1}{\phi}\mathbf{I}$ *y la distribución a priori* $\theta \sim \mathcal{N}\left(\mu, \frac{1}{\alpha\phi}\right)$ *tenemos*

$$\begin{aligned}
\mathbf{A}^T\mathbf{C}^{-1}\mathbf{A} &= n\phi \\
\mathbf{A}^T\mathbf{C}^{-1}\mathbf{y} &= \phi\sum_{i=1}^{n} y_i = n\phi\bar{y} \\
\mathbf{A}^T\mathbf{C}^{-1}\mathbf{A} + \mathbf{V}^{-1} &= n\phi + \alpha\phi = (n+\alpha)\phi \\
\mathbf{A}^T\mathbf{C}^{-1}\mathbf{y} + \mathbf{V}^{-1}\boldsymbol{\mu} &= n\phi\bar{y} + \alpha\phi\mu \\
\mathbf{Bb} &= \frac{n\bar{y} + \alpha\mu}{n+\alpha}
\end{aligned}$$

que es el resultado que hemos visto para la media a posteriori en el Ejemplo 38.

Ejemplo 66 *Escribir el siguiente problema en términos del modelo lineal de 2 etapas y entonces calcular las distribuciones a posteriori de θ y α.*

$$y_1 \sim \mathcal{N}\left(\theta + \delta, \frac{1}{\phi}\right) \quad y_2 \sim \mathcal{N}\left(\theta - \delta, \frac{1}{\phi}\right)$$

donde y_1 y y_2 independientes dados θ y δ.

$$\theta \sim \mathcal{N}(m, v) \quad \delta \sim \mathcal{N}(0, w)$$

θ δ independientes.

¿Qué forma tiene la distribución a posteriori cuando $v \to \infty$?

Definimos $\mathbf{y} = (y_1, y_2)^T$ *y* $\boldsymbol{\theta} = (\theta, \delta)^T$ *y entonces*

$$E[\mathbf{y}|\boldsymbol{\theta}] = \mathbf{A}\boldsymbol{\theta} \quad \textit{donde} \quad \mathbf{A} = \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}$$

También $\mathbf{C} = V[\mathbf{y}|\boldsymbol{\theta}] = \frac{1}{\phi}\mathbf{I}$ *y la distribución a priori de* $\boldsymbol{\theta}$ *es*

$$\begin{aligned} \boldsymbol{\theta} &\sim \mathcal{N}(\boldsymbol{\mu}, \mathbf{V}) \quad \textit{donde} \\ \boldsymbol{\mu} &= (m, 0)^T \\ \mathbf{V} &= \begin{pmatrix} v & 0 \\ 0 & w \end{pmatrix} \end{aligned}$$

Calculamos la distribución a posteriori utilizando el resultado del teorema.

$$
\begin{aligned}
\mathbf{V}^{-1} &= \begin{pmatrix} 1/v & 0 \\ 0 & 1/w \end{pmatrix} \\
\mathbf{A}^T\mathbf{C}^{-1}\mathbf{A} &= \phi \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} \\
&= \phi \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix} \\
\mathbf{B}^{-1} &= \begin{pmatrix} \frac{1}{v} + 2\phi & 0 \\ 0 & \frac{1}{w} + 2\phi \end{pmatrix} \\
\mathbf{B} &= \begin{pmatrix} \frac{1}{\frac{1}{v}+2\phi} & 0 \\ 0 & \frac{1}{\frac{1}{w}+2\phi} \end{pmatrix} \\
\mathbf{V}^{-1}\boldsymbol{\mu} &= (m/v, 0)^T \\
\mathbf{A}^T\mathbf{C}^{-1}\mathbf{y} &= \phi(y_1 + y_2, y_1 - y_2)^T \\
\mathbf{b} &= \left(\frac{m}{v} + \phi(y_1 + y_2), \phi(y_1 - y_2)\right)^T \\
\mathbf{Bb} &= \left(\frac{\frac{m}{v} + \phi(y_1 + y_2)}{\frac{1}{v} + 2\phi}, \frac{\phi(y_1 - y_2)}{\frac{1}{w} + 2\phi}\right)^T
\end{aligned}
$$

Entonces θ y δ son independientes a posteriori con distribuciones

$$
\begin{aligned}
\theta|\mathbf{y} &\sim \mathcal{N}\left(\frac{\frac{m}{v}+\phi(y_1+y_2)}{\frac{1}{v}+2\phi}, \frac{1}{\frac{1}{v}+2\phi}\right) \\
\delta|\mathbf{y} &\sim \mathcal{N}\left(\frac{\phi(y_1-y_2)}{\frac{1}{w}+2\phi}, \frac{1}{\frac{1}{w}+2\phi}\right)
\end{aligned}
$$

La media a posteriori de θ es una media ponderada de la media a priori y el EMV con pesos proporcionales a la precisión a priori ($1/v$) y la precisión del EMV (2ϕ).

Si $v \to \infty$, tenemos

$$
\begin{aligned}
E[\theta|\mathbf{y}] &= \frac{\frac{m}{v}+\phi(y_1+y_2)}{\frac{1}{v}+2\phi} \to \frac{\phi(y_1+y_2)}{2\phi} = \bar{y} \\
V[\theta|\mathbf{y}] &= \frac{1}{\frac{1}{v}+2\phi} \to \frac{1}{2\phi}
\end{aligned}
$$

y la distribución a posteriori de θ es $\mathcal{N}\left(\bar{y}, \frac{1}{2\phi}\right)$. No cambia la distribución a posteriori de δ.

Resultados para el modelo de 2 etapas con distribución a priori uniforme

Sea la distribución a priori no informativa; $f(\boldsymbol{\theta}) \propto 1$. Esta distribución a priori equivale a poner $\mathbf{V}^{-1} = \mathbf{0}$ en el Teorema. Entonces:

$$\begin{aligned} \mathbf{B}^{-1} &= \mathbf{A}^T\mathbf{C}^{-1}\mathbf{A} \\ \Rightarrow \mathbf{B} &= \left(\mathbf{A}^T\mathbf{C}^{-1}\mathbf{A}\right)^{-1} \\ \mathbf{b} &= \mathbf{A}^T\mathbf{C}^{-1}\mathbf{y} \end{aligned}$$

y la distribución a posteriori será

$$\boldsymbol{\theta}|\mathbf{y} \sim \mathcal{N}\left(\left(\mathbf{A}^T\mathbf{C}^{-1}\mathbf{A}\right)^{-1}\mathbf{A}^T\mathbf{C}^{-1}\mathbf{y}, \left(\mathbf{A}^T\mathbf{C}^{-1}\mathbf{A}\right)^{-1}\right).$$

La media a posteriori es igual al EMV de $\boldsymbol{\theta}$.

Ejemplo 67 *Siendo $\mathbf{C} = \frac{1}{\phi}\mathbf{I}$, se tiene el modelo lineal con errores independientes y dada una distribución a priori uniforme, la distribución a posteriori es*

$$\boldsymbol{\theta}|\mathbf{y} \sim \mathcal{N}\left(\left(\mathbf{A}^T\mathbf{A}\right)^{-1}\mathbf{A}^T\mathbf{y}, \frac{1}{\phi}\left(\mathbf{A}^T\mathbf{A}\right)^{-1}\right)$$

y la media a posteriori de $\boldsymbol{\theta}$ coincide con el EMV.

Ejemplo 68 *Retomando el modelo de regresión simple del Ejemplo 60 se tiene*

$$
\begin{aligned}
\mathbf{A}^T\mathbf{A} &= \begin{pmatrix} n & 0 \\ 0 & \sum(x_i-\bar{x})^2 \end{pmatrix} \\
\mathbf{A}^T\mathbf{A}^{-1} &= \begin{pmatrix} 1/n & 0 \\ 0 & 1/\sum(x_i-\bar{x})^2 \end{pmatrix} \\
\mathbf{A}^T\mathbf{y} &= \begin{pmatrix} \sum y_i \\ \sum y_i(x_i-\bar{x}) \end{pmatrix} \\
\mathbf{A}^T\mathbf{A}^{-1}\mathbf{A}^T\mathbf{y} &= \begin{pmatrix} \frac{1}{n}\sum y_i \\ \frac{\sum y_i(x_i-\bar{x})}{\sum(x_i-\bar{x})^2} \end{pmatrix} \\
&= \begin{pmatrix} \bar{y} \\ SC(xy)/SC(xx) \end{pmatrix}
\end{aligned}
$$

donde $SC(xy) = \sum_i(x_i-\bar{x})(y_i-\bar{y})$ *y* $SC(xx) = \sum_i(x_i-\bar{x})^2$.

Las distribuciones a posteriori de α *y* β *son independientes con medias iguales a los estimadores mínimos cuadrados:*

$$\begin{aligned}\alpha|\mathbf{y} &\sim \mathcal{N}\left(\bar{y}, \frac{1}{n\phi}\right)\\ \beta|\mathbf{y} &\sim \mathcal{N}\left(\frac{SC(xy)}{SC(xx)}, \frac{1}{\phi SC(xx)}\right)\end{aligned}$$

Ejemplo 69 *Volviendo al Ejemplo 61 supongamos una distribución a priori uniforme para* $\boldsymbol{\theta} = (\mu, \alpha_2, \beta_2, \beta_3)^T$.

Tenemos

$$
\begin{aligned}
\mathbf{A}^T\mathbf{A} &= \begin{pmatrix} 6 & 3 & 2 & 2 \\ 3 & 3 & 1 & 1 \\ 2 & 1 & 2 & 0 \\ 2 & 1 & 0 & 2 \end{pmatrix} \\
(\mathbf{A}^T\mathbf{A})^{-1} &= \begin{pmatrix} 2/3 & -1/3 & -1/2 & -1/2 \\ -1/3 & 2/3 & 0 & 0 \\ -1/2 & 0 & 1 & 1/2 \\ -1/2 & 0 & 1/2 & 1 \end{pmatrix} \\
\mathbf{A}^T\mathbf{y} &= \begin{pmatrix} 6\bar{y}_{\cdot\cdot} \\ 3\bar{y}_{2\cdot} \\ 2\bar{y}_{\cdot 2} \\ 2\bar{y}_{\cdot 3} \end{pmatrix}
\end{aligned}
$$

donde $\bar{y}_{\cdot\cdot} = \frac{1}{6}\sum_i \sum_j y_{ij}$, $\bar{y}_{2\cdot} = \frac{1}{3}\sum_j y_{2j}$, *etc.*

Entonces por ejemplo, la distribución a posteriori de α_2 *es normal con media* $2\bar{y}_{2\cdot} - 2\bar{y}_{\cdot\cdot}$ *y varianza* $2/(3\phi)$.

Cuando la varianza $\mathbf{C}$ es desconocida

Consideramos sólo el caso: $\mathbf{C} = \frac{1}{\phi}\mathbf{D}$ con $\mathbf{D}$ conocida (por ejemplo $\mathbf{D} = \mathbf{I}$).

Supongamos la distribución a priori no informativa

$$f(\boldsymbol{\theta}, \phi) \propto \frac{1}{\phi}.$$

Entonces, se tiene

Teorema 12 *La distribución a posteriori es*

$$\begin{aligned}
\boldsymbol{\theta}|\phi, \mathbf{y} &\sim \mathcal{N}\left(\mathbf{Bb}, \frac{1}{\phi}\mathbf{B}\right) \quad \textit{donde} \\
\mathbf{B}^{-1} &= \mathbf{A}^T\mathbf{D}^{-1}\mathbf{A} \\
\mathbf{b} &= \mathbf{A}^T\mathbf{D}^{-1}\mathbf{y} \\
\phi|\mathbf{y} &\sim \mathcal{G}\left(\frac{n-p}{2}, \frac{\mathbf{y}^T\mathbf{D}^{-1}\mathbf{y} - \mathbf{b}^T\mathbf{Bb}}{2}\right)
\end{aligned}$$

Demostración

La demostración es similar a la del teorema 11. Se tiene

$$\begin{aligned}
f(\boldsymbol{\theta}, \phi|\mathbf{y}) &\propto \phi^{-1}\phi^{\frac{n}{2}} \exp\left(-\frac{\phi}{2}(\mathbf{y}-\mathbf{A}\boldsymbol{\theta})^T\mathbf{D}^{-1}(\mathbf{y}-\mathbf{A}\boldsymbol{\theta})\right) \\
&\propto \phi^{\frac{n}{2}-1} \exp\left(-\frac{\phi}{2}\left[(\boldsymbol{\theta}-\mathbf{Bb})^T\mathbf{B}^{-1}(\boldsymbol{\theta}-\mathbf{Bb})\right.\right. \\
&\qquad \left.\left. +\mathbf{y}^T\mathbf{D}^{-1}\mathbf{y}-(\mathbf{Bb})^T\mathbf{B}^{-1}(\mathbf{Bb})\right]\right) \\
f(\phi|\mathbf{y}) &\propto \phi^{\frac{n-p}{2}-1} \exp\left(-\frac{\phi}{2}\left[\mathbf{y}^T\mathbf{D}^{-1}\mathbf{y}-\mathbf{b}^T\mathbf{Bb}\right]\right)
\end{aligned}$$

que es el núcleo de una distribución gamma. $\diamond$

El resultado implica que la distribución a posteriori marginal de $\boldsymbol{\theta}$ será una distribución t, no centrada y multivariante. Es muy difícil tratar con esta distribución excepto en algunos casos especiales.

Ejemplo 70 *Volviendo al Ejemplo 68, tenemos*

$$\begin{aligned}
\mathbf{y}^T\mathbf{D}^{-1}\mathbf{y} &= \sum_{i=1}^n y_i^2 \\
\mathbf{b}^T\mathbf{B}\mathbf{b} &= \left(n\bar{y}, \sum_{i=1}^n (x_i - \bar{x})y_i\right)\left(\begin{array}{c} \bar{y} \\ SC(xy)/SC(xx) \end{array}\right) \\
&= n\bar{y}^2 + SC(xy)^2/SC(xx) \\
\mathbf{y}^T\mathbf{D}^{-1}\mathbf{y} - \mathbf{b}^T\mathbf{B}\mathbf{b} &= SC(yy) - SC(xy)^2/SC(xx) \\
\phi|\mathbf{y} &\sim \mathcal{G}\left(\frac{n-2}{2}, \frac{SC(yy) - SC(xy)^2/SC(xx)}{2}\right)
\end{aligned}$$

Observamos que el estimador clásico de la varianza residual es

$$s^2 = \frac{1}{n-2}\left(SC(yy) - \frac{SC(xy)^2}{SC(xx)}\right).$$

Entonces, se puede demostrar que las distribuciones marginales de α y β son independientes t no centradas y (por ejemplo) un intervalo de credibilidad para β es

$$\frac{SC(xy)}{SC(xx)} \pm \frac{s}{\sqrt{SC(xx)}} t_{n-2}(,025).$$

Este intervalo es igual al intervalo clásico de confianza.

En el caso de que la varianza es completamente desconocida se necesitan métodos numéricos para calcular las distribuciones a posteriori.