Tarea 3: Regresión Lineal y Bayesiana

Luz Maria Salazar Manjarrez
Joksan Avendaño Caballero
Omar Garcia Ramos

Instrucciones Generales
Resuelva cuidadosamente cada apartado. Todas las demostraciones deben presentarse paso a paso, con claridad en las justificaciones matemáticas y con una breve interpretación estadística de los resultados.

1 Regresión lineal ordinaria (OLS)

1. **Derivación del estimador OLS:** Partiendo del modelo clásico

$$y = X\beta + \varepsilon, \quad \varepsilon \sim \mathcal{N}(0, \sigma^2 I),$$

demuestre que el estimador de Mínimos Cuadrados Ordinarios es

$$\hat{\beta} = (X^\top X)^{-1} X^\top y,$$

siempre que $X^\top X$ sea invertible.

2. **Propiedades del estimador:** Calcule explícitamente:

$$E[\hat{\beta}], \qquad \mathrm{Var}(\hat{\beta}).$$

Concluya que $\hat{\beta}$ es insesgado y eficiente dentro de la clase de estimadores lineales (teorema de Gauss–Markov).

Solución.

1. Consideremos el modelo lineal clásico

$$y = X\beta + \varepsilon \quad y \in \mathbb{R}^n,\ X \in \mathbb{R}^{n\times p},\ \beta \in \mathbb{R}^p, \varepsilon \sim \mathcal{N}(0, \sigma^2 I).$$

El estimador de Mínimos Cuadrados $\hat{\beta}$ se define como el vector que minimiza la suma de cuadrados de los residuos

$$L(\beta) = \|y - X\beta\|_2^2 = (y - X\beta)^\top (y - X\beta).$$

Al expandir,

$$L(\beta) = y^\top y - 2\,\beta^\top X^\top y + \beta^\top X^\top X\,\beta.$$

La condición de primer orden para un mínimo es $\nabla_\beta L(\beta) = 0$. Usando que si A es simétrica, $\frac{\partial}{\partial\beta}(\beta^\top A\beta) = 2A\beta$, se tiene:

$$\begin{aligned}
&\frac{\partial}{\partial\beta}(y^\top y) = 0,\\
&\frac{\partial}{\partial\beta}\Big(-2\,\beta^\top X^\top y\Big) = \frac{\partial}{\partial\beta}\Big(-2(X^\top y)^\top \beta\Big) = -2X^\top y,\\
&\frac{\partial}{\partial\beta}\Big(\beta^\top X^\top X\,\beta\Big) = 2X^\top X\,\beta,
\end{aligned}$$

y por tanto el gradiente es

$$\nabla_\beta L(\beta) = -2X^\top y + 2X^\top X\,\beta.$$

Igualando a 0 se tiene que

$$X^\top X\,\hat{\beta} = X^\top y. \tag{1}$$

Si X tiene rango completo ($\text{rank}(X) = p$), entonces $X^\top X$ es simétrica definida positiva e invertible, de modo que la solución de las ecuaciones normales existe y es única, dada por

$$\hat{\beta} = (X^\top X)^{-1} X^\top y.$$

Para verificar que este punto crítico es un mínimo global, el hessiano es

$$\nabla_\beta^2 L(\beta) = 2X^\top X,$$

que es definido positivo cuando X tiene rango completo; así, L es estrictamente convexa en $\mathbb{R}^p$ y el punto crítico es el único minimizador global.

La condición $X^\top(y - X\hat{\beta}) = 0$ (que se sigue de 1) muestra que el residuo $\hat{\varepsilon} = y - X\hat{\beta}$ es ortogonal a todas las columnas de X, consistente con $\hat{y} = X\hat{\beta}$ como proyección ortogonal de y sobre el subespacio columna de X.

2. Veamos que

$$\begin{aligned}
E[\hat{\beta}] &= E\Big[(X^\top X)^{-1} X^\top y\Big]\\
&= E\Big[(X^\top X)^{-1} X^\top (X\beta + \varepsilon)\Big]\\
&= (X^\top X)^{-1} X^\top X\,\beta + (X^\top X)^{-1} X^\top E[\varepsilon]\\
&= \beta + 0\\
&= \beta,
\end{aligned}$$

de modo que $\hat{\beta}$ es insesgado.

Para la varianza, usando linealidad y $\mathrm{Var}(Ay) = A\,\mathrm{Var}(y)\,A^\top$, tenemos que

$$\begin{aligned}\mathrm{Var}(\hat{\beta}) &= \mathrm{Var}\Big((X^\top X)^{-1}X^\top y\Big)\\ &= (X^\top X)^{-1}X^\top\,\mathrm{Var}(y)\,X(X^\top X)^{-1}\\ &= \sigma^2(X^\top X)^{-1}.\end{aligned}$$

Para el teorema de Gauss–Markov, considérese cualquier estimador lineal insesgado de β de la forma $\tilde{\beta} = Ay$ con $A \in \mathbb{R}^{p\times n}$ tal que $E[\tilde{\beta}] = AX\beta = \beta$ para todo β, es decir, $AX = I_p$. Sea $A_0 = (X^\top X)^{-1}X^\top$ (el de OLS) y defínase $D = A - A_0$. Entonces $DX = AX - A_0X = I_p - I_p = 0$. La varianza de $\tilde{\beta}$ es

$$\mathrm{Var}(\tilde{\beta}) = \sigma^2 AA^\top = \sigma^2(A_0 + D)(A_0 + D)^\top = \sigma^2\Big(A_0A_0^\top + DD^\top + DA_0^\top + A_0D^\top\Big).$$

Como $DA_0^\top = DX(X^\top X)^{-1} = 0$ y $A_0D^\top = 0$, se obtiene

$$\mathrm{Var}(\tilde{\beta}) - \mathrm{Var}(\hat{\beta}) = \sigma^2\left(AA^\top - A_0A_0^\top\right) = \sigma^2 DD^\top,$$

y $\sigma^2 DD^\top$ es semidefinida positiva. En consecuencia,

$$\mathrm{Var}(\tilde{\beta}) \;-\; \mathrm{Var}(\hat{\beta}) \geq 0,$$

esto es, $\hat{\beta}$ tiene varianza mínima dentro de la clase de estimadores lineales insesgados. Por tanto, bajo los supuestos del modelo, $\hat{\beta}$ es BLUE (Best Linear Unbiased Estimator), concluyendo la eficiencia de OLS en el sentido de Gauss–Markov.

■

2 Regresión lineal bayesiana (prior conjugado)

1. **Prior conjugado:** Suponga un prior conjugado:

$$\beta \mid \sigma^2 \sim \mathcal{N}(\beta_0, \sigma^2 V_0), \qquad \sigma^2 \sim \text{Inv-Gamma}(a_0, b_0).$$

2. **Distribución posterior:** Derive los parámetros posteriores (β_n, V_n, a_n, b_n) y escriba la forma explícita de la posterior conjunta

$$p(\beta, \sigma^2 \mid y).$$

3. **Distribuciones marginales:** Identifique las distribuciones marginales de β y de σ^2.

Solución.

1. Suponemos el modelo gaussiano $y \mid \beta, \sigma^2 \sim \mathcal{N}(X\beta, \sigma^2 I_n)$ y un prior conjugado
$$\beta \mid \sigma^2 \sim \mathcal{N}(\beta_0, \sigma^2 V_0), \qquad \sigma^2 \sim \text{Inv-Gamma}(a_0, b_0),$$
con V_0 simétrica definida positiva. Entonces
$$p(\beta \mid \sigma^2) = (2\pi\sigma^2)^{-p/2}|V_0|^{-1/2} \exp\Big\{ -\tfrac{1}{2\sigma^2}(\beta-\beta_0)^\top V_0^{-1}(\beta-\beta_0)\Big\}$$
y
$$p(\sigma^2) = \frac{b_0^{a_0}}{\Gamma(a_0)}(\sigma^2)^{-(a_0+1)} \exp\Big\{ -\tfrac{b_0}{\sigma^2}\Big\}.$$
La verosimilitud es
$$L(y \mid \beta, \sigma^2) \propto (\sigma^2)^{-n/2} \exp\Big\{ -\tfrac{1}{2\sigma^2}(y-X\beta)^\top (y-X\beta)\Big\}.$$

2. Multiplicando verosimilitud y prior conjunto se obtiene
$$\begin{aligned} p(\beta, \sigma^2 \mid y) &\propto L(y \mid \beta, \sigma^2)\, p(\beta \mid \sigma^2)\, p(\sigma^2) \\ &\propto (\sigma^2)^{-\frac{n}{2}} \exp\Big\{ -\tfrac{(y-X\beta)^\top (y-X\beta)}{2\sigma^2}\Big\} \\ &\cdot (\sigma^2)^{-\frac{p}{2}} \exp\Big\{ -\tfrac{(\beta-\beta_0)^\top V_0^{-1}(\beta-\beta_0)}{2\sigma^2}\Big\} \cdot (\sigma^2)^{-(a_0+1)} e^{-b_0/\sigma^2}. \end{aligned}$$
Combinando términos:
$$\begin{aligned} p(\beta, \sigma^2 \mid y) \propto (\sigma^2)^{-(n+p)/2-a_0-1} \\ \cdot \exp\left(-\frac{1}{\sigma^2}\left[b_0 + \frac{1}{2}(y-X\beta)^\top (y-X\beta) + \frac{1}{2}(\beta-\beta_0)^\top V_0^{-1}(\beta-\beta_0)\right]\right) \end{aligned}$$
Expandimos los términos cuadráticos:
$$\begin{aligned} &(y-X\beta)^\top (y-X\beta) + (\beta-\beta_0)^\top V_0^{-1}(\beta-\beta_0) \\ &= y^\top y - 2y^\top X\beta + \beta^\top X^\top X\beta + \beta^\top V_0^{-1}\beta - 2\beta_0^\top V_0^{-1}\beta + \beta_0^\top V_0^{-1}\beta_0 \\ &= \beta^\top (X^\top X + V_0^{-1})\beta - 2(y^\top X + \beta_0^\top V_0^{-1})\beta + (y^\top y + \beta_0^\top V_0^{-1}\beta_0) \end{aligned}$$
Definimos:
$$\begin{aligned} V_n^{-1} &= X^\top X + V_0^{-1} \\ \beta_n &= V_n(X^\top y + V_0^{-1}\beta_0) \end{aligned}$$
Ahora completamos el cuadrado. Queremos expresar la forma cuadrática como:
$$(\beta-\beta_n)^\top V_n^{-1}(\beta-\beta_n) + \text{constante}$$
Desarrollamos:
$$\begin{aligned} &(\beta-\beta_n)^\top V_n^{-1}(\beta-\beta_n) \\ &= \beta^\top V_n^{-1}\beta - 2\beta_n^\top V_n^{-1}\beta + \beta_n^\top V_n^{-1}\beta_n \end{aligned}$$

Comparando con nuestra expresión original, vemos que:

$$\begin{aligned} V_n^{-1} &= X^\top X + V_0^{-1} \\ -2\beta_n^\top V_n^{-1} &= -2(y^\top X + \beta_0^\top V_0^{-1}) \end{aligned}$$

Por lo tanto:

$$\beta_n^\top = (y^\top X + \beta_0^\top V_0^{-1})V_n$$

o equivalentemente:

$$\beta_n = V_n(X^\top y + V_0^{-1}\beta_0)$$

La constante que queda es:

$$\begin{aligned} &y^\top y + \beta_0^\top V_0^{-1}\beta_0 - \beta_n^\top V_n^{-1}\beta_n \\ &= y^\top y + \beta_0^\top V_0^{-1}\beta_0 - (X^\top y + V_0^{-1}\beta_0)^\top V_n(X^\top y + V_0^{-1}\beta_0) \end{aligned}$$

Por lo tanto, la distribución posterior conjunta es:

$$p(\beta, \sigma^2 \mid y) \propto (\sigma^2)^{-(n+p)/2-a_0-1} \exp\left(-\frac{1}{\sigma^2}\left[b_0 + \frac{1}{2}(\beta-\beta_n)^\top V_n^{-1}(\beta-\beta_n) + \frac{1}{2}c\right]\right)$$

donde:

$$c = y^\top y + \beta_0^\top V_0^{-1}\beta_0 - \beta_n^\top V_n^{-1}\beta_n$$

Definiendo:

$$\begin{aligned} a_n &= a_0 + \frac{n}{2} \\ b_n &= b_0 + \frac{1}{2}\left(y^\top y + \beta_0^\top V_0^{-1}\beta_0 - \beta_n^\top V_n^{-1}\beta_n\right) \end{aligned}$$

entonces

$$p(\beta, \sigma^2 \mid y) \;\propto\; (\sigma^2)^{-\frac{p}{2}-(a_n+1)} \exp\Big\{-\frac{1}{2\sigma^2}(\beta-\beta_n)^\top V_n^{-1}(\beta-\beta_n)\Big\} \exp\Big\{-\frac{b_n}{\sigma^2}\Big\},$$

Normalizando se tiene que la forma explícita de la posterior conjunta es

$$p(\beta, \sigma^2 \mid y) = (2\pi\sigma^2)^{-p/2}|V_n|^{-1/2} \exp\left(-\frac{1}{2\sigma^2}(\beta-\beta_n)^\top V_n^{-1}(\beta-\beta_n)\right) \times \frac{b_n^{a_n}}{\Gamma(a_n)}(\sigma^2)^{-a_n-1} \exp\left(-\frac{b_n}{\sigma^2}\right)$$

Es decir

$$\beta \mid \sigma^2, y \sim \mathcal{N}(\beta_n, \sigma^2 V_n), \qquad \sigma^2 \mid y \sim \text{Inv-Gamma}(a_n, b_n)$$

3. Calculando de la marginal de $\beta \mid y$:

$$\begin{aligned} p(\beta \mid y) &\propto \int_0^\infty (\sigma^2)^{-\frac{p}{2}-(a_n+1)} \exp\Big\{-\tfrac{1}{2\sigma^2}(\beta-\beta_n)^\top V_n^{-1}(\beta-\beta_n)\Big\} \exp\Big\{-\tfrac{b_n}{\sigma^2}\Big\}\, d\sigma^2 \\ &= \int_0^\infty (\sigma^2)^{-(a_n+\frac{p}{2}+1)} \exp\Big\{-\frac{b_n + \frac{1}{2}Q(\beta)}{\sigma^2}\Big\}\, d\sigma^2, \end{aligned}$$

donde $Q(\beta) := (\beta - \beta_n)^\top V_n^{-1}(\beta - \beta_n)$. Luego, denotando $\nu = 2a_n + p$ y $c = b_n + \frac{1}{2}Q(\beta)$ se tiene que

$$p(\beta \mid y) \propto \int_0^\infty (\sigma^2)^{-(\nu/2+1)} e^{-c/\sigma^2} d\sigma^2$$

Haciendo el cambio $u = 1/\sigma^2$ ($d\sigma^2 = -u^{-2}du$):

$$\int_0^\infty (\sigma^2)^{-(\nu/2+1)} e^{-c/\sigma^2} d\sigma^2 = \tfrac{1}{2}\, c^{-\nu/2}\, \Gamma(\nu/2), \quad c > 0,\ \nu > 0.$$

Luego

$$p(\beta \mid y) \ \propto\ \Big(b_n + \tfrac{1}{2}Q(\beta)\Big)^{-\frac{2a_n+p}{2}} \ =\ b_n^{-a_n-\frac{p}{2}} \Big(1 + \frac{(\beta - \beta_n)^\top V_n^{-1}(\beta - \beta_n)}{2b_n}\Big)^{-\frac{2a_n+p}{2}}.$$

Reescribiendo con $\mu = 2a_n$ y $\Sigma^{-1} = \frac{a_n}{b_n} V_n^{-1}$,

$$p(\beta \mid y) \ \propto\ \Big(1 + \frac{1}{\mu}(\beta - \beta_n)^\top \Sigma^{-1}(\beta - \beta_n)\Big)^{-\frac{\mu+p}{2}},$$

que es el kernel de una t multivariada de μ grados de libertad, media β_n y matriz de escala Σ. Así:

$$\beta \mid y \sim t_\nu\Big(\beta_n,\ \Sigma\Big), \qquad \nu = 2a_n,\ \ \Sigma = \frac{b_n}{a_n} V_n.$$

La actualización $\beta_n = V_n(V_0^{-1}\beta_0 + X^\top y)$ muestra una media posterior como promedio ponderado entre la información previa ($V_0^{-1}\beta_0$) y la información de los datos ($X^\top y$), con pesos dados por las precisiones V_0^{-1} y $X^\top X$. La matriz $V_n = (V_0^{-1} + X^\top X)^{-1}$ reduce la incertidumbre respecto a V_0 al incorporar los datos, y los parámetros a_n, b_n incrementan la evidencia efectiva sobre σ^2.

■

3 Conexión con regularización

1. **Regresión Ridge:** Muestre que si se toma un prior Normal isotrópico

$$\beta \sim \mathcal{N}(0, \tau^2 I),$$

el estimador de máxima a posteriori (MAP) es equivalente a la regresión Ridge:

$$\hat{\beta}_{\text{MAP}} = \arg\min_\beta\ \|y - X\beta\|^2 + \lambda\|\beta\|^2, \qquad \lambda = \sigma^2/\tau^2.$$

2. **Regresión Lasso:** Muestre que si en lugar de un prior Normal se utiliza un prior Laplace (doble-exponencial)

$$p(\beta_j) \propto \exp(-\lambda|\beta_j|),$$

el estimador MAP corresponde a la regresión Lasso:

$$\hat{\beta}_{\text{MAP}} = \arg\min_\beta\ \|y - X\beta\|^2 + \lambda\|\beta\|_1.$$

Solución.

1. Suponemos el modelo gaussiano $y \mid \beta, \sigma^2 \sim \mathcal{N}(X\beta, \sigma^2 I_n)$, de modo que
$$p(y \mid \beta, \sigma^2) \propto (\sigma^2)^{-n/2} \exp\Big\{-\tfrac{1}{2\sigma^2}\|y - X\beta\|^2\Big\}.$$
Tomemos como prior a $\beta \sim \mathcal{N}(0, \tau^2 I_p)$, entonces
$$p(\beta) \propto (\tau^2)^{-p/2} \exp\Big\{-\tfrac{1}{2\tau^2}\|\beta\|^2\Big\}.$$
La posterior no normalizada es entonces
$$p(\beta \mid y, \sigma^2) \propto L(y \mid \beta, \sigma^2)\, p(\beta) \propto \exp\Big\{-\tfrac{1}{2\sigma^2}\|y - X\beta\|^2 - \tfrac{1}{2\tau^2}\|\beta\|^2\Big\}.$$
El estimador MAP maximiza la log-posterior y equivalentemente, minimiza su opuesto, es decir:
$$\hat{\beta}_{\text{MAP}} = \arg\min_{\beta} \Big\{\tfrac{1}{2\sigma^2}\|y - X\beta\|^2 + \tfrac{1}{2\tau^2}\|\beta\|^2\Big\}.$$
Multiplicando por $2\sigma^2$ (al ser positivo esto no altera al minimizador):
$$\hat{\beta}_{\text{MAP}} = \arg\min_{\beta} \Big\{\|y - X\beta\|^2 + \tfrac{\sigma^2}{\tau^2}\|\beta\|^2\Big\} = \arg\min_{\beta} \Big\{\|y - X\beta\|^2 + \lambda\|\beta\|^2\Big\}$$
que es exactamente la regresión Ridge con parámetro de penalización $\lambda = \sigma^2/\tau^2$.

2. Ahora bien, supongamos un prior Laplace (doble exponencial) independiente por coordenada:
$$p(\beta) \;\propto\; \prod_{j=1}^{p} \exp\{-\lambda_0|\beta_j|\} = \exp\{-\lambda_0\|\beta\|_1\}.$$
Entonces la posterior no normalizada es
$$p(\beta \mid y, \sigma^2) \propto \exp\Big\{-\tfrac{1}{2\sigma^2}\|y - X\beta\|^2 - \lambda_0\|\beta\|_1\Big\}.$$
El MAP equivale a
$$\hat{\beta}_{\text{MAP}} = \arg\min_{\beta} \Big\{\tfrac{1}{2\sigma^2}\|y - X\beta\|^2 + \lambda_0\|\beta\|_1\Big\}.$$
Multiplicando por $2\sigma^2$ se obtiene la forma estándar de Lasso:
$$\hat{\beta}_{\text{MAP}} = \arg\min_{\beta} \Big\{\|y - X\beta\|^2 + (2\sigma^2\lambda_0)\|\beta\|_1\Big\} = \arg\min_{\beta} \Big\{\|y - X\beta\|^2 + \lambda\|\beta\|_1\Big\}$$
Es decir, el prior Laplace induce una penalización L_1; la constante de Lasso λ depende de la parametrización del prior (aquí $\lambda = 2\sigma^2\lambda_0$).

   De los dos resultados anteriores podemos decir que el prior gaussiano isotrópico contrae los coeficientes de forma cuadrática y produce la penalización ℓ_2 de Ridge (contracción suave sin volver exactamente cero los coeficientes). El prior Laplace induce la penalización ℓ_1 de Lasso y sparsity (muchos coeficientes exactamente cero, es decir, puede eliminar variables). Al ajustar λ por validación cruzada estamos, implícitamente, fijando cuánta "fuerza" le damos al prior.

■

4 Extensiones: errores no normales

1. **Modelos alternativos:** Proponga un modelo de regresión donde el error ε no siga una distribución Normal. Ejemplos:
   - $\varepsilon \sim \text{Laplace}(0, b)$ (robusto a *outliers*).
   - $\varepsilon \sim \text{Student-}t(\nu)$ (colas pesadas).
2. **Consecuencias metodológicas:** Explique cuáles serían las consecuencias sobre:
   - La forma de la verosimilitud.
   - La existencia o no de priors conjugados.
   - Los métodos de inferencia requeridos (MCMC, aproximación variacional, etc.).

Solución. Se considera el modelo de regresión lineal con errores Laplace:

$$y_i = x_i^\top \beta + \varepsilon_i, \quad \varepsilon_i \overset{\text{i.i.d.}}{\sim} \text{Laplace}(0, b),$$

entonces

$$f(\varepsilon) = \frac{1}{2b} \exp\left(-\frac{|\varepsilon|}{b}\right)$$

La verosimilitud para el modelo completo es:

$$L(\beta, b) = \prod_{i=1}^{n} \frac{1}{2b} \exp\left(-\frac{|y_i - x_i^\top \beta|}{b}\right) = (2b)^{-n} \exp\left(-\frac{1}{b} \sum_{i=1}^{n} |y_i - x_i^\top \beta|\right),$$

entonces la log-verosimilitud es:

$$\ell(\beta, b) = -n \log(2b) - \frac{1}{b} \sum_{i=1}^{n} |y_i - x_i^\top \beta|$$

Maximizar $\ell(\beta, b)$ equivale a minimizar $\sum_{i=1}^{n} |y_i - x_i^\top \beta|$ (regresión por mínimas desviaciones absolutas (LAD)) por lo que se tiene que la estimación es robusta a outliers debido al uso de valores absolutos en lugar de cuadrados.

Ahora bien, no existe una familia conjugada completa para (β, b) bajo errores Laplace, ya que la distribución Laplace no pertenece a la familia exponencial cuando ambos parámetros son desconocidos y la presencia del término $|y_i - x_i^\top \beta|$ y la forma no cuadrática impiden la conjugación algebraica.

No existe un prior $p(\beta, b)$ que combine con la verosimilitud Laplace para producir una posterior de forma cerrada conocida, por tanto se requieren métodos computacionales.

Una vía estándar es usar la representación de Laplace como mezcla de Normales con varianzas exponenciales:

$$\varepsilon_i \,\Big|\, \tau_i \sim \mathcal{N}(0, \tau_i), \qquad \tau_i \,\Big|\, \psi \sim \text{Exp}\left(\frac{\psi}{2}\right), \quad \text{con} \quad \psi = b^{-2},$$

lo cual implica marginalmente $\varepsilon_i \sim \text{Laplace}(0, b)$ (Andrews and Mallows (1974)).

Con esto, el modelo jerárquico queda

$$y \mid \beta, \tau \sim \mathcal{N}\Big(X\beta,\ \mathrm{diag}(\tau_1, \dots, \tau_n)\Big), \qquad \tau_i \mid \psi \overset{iid}{\sim} \mathrm{Exp}(\psi/2), \qquad \psi = b^{-2}.$$

Si se fija un prior gaussiano $\beta \sim \mathcal{N}(\beta_0, V_0)$ y un prior Gamma para ψ (en notación forma–tasa) $\psi \sim \mathrm{Gamma}(a_0, b_0)$, los condicionales completos para un muestreo Gibbs son cerrados (Park and Casella (2008) y West (1987)):

$$\beta \mid y, \tau \sim \mathcal{N}(\beta_n, V_n), \quad V_n = (V_0^{-1} + X^\top D^{-1} X)^{-1},\ \beta_n = V_n(V_0^{-1}\beta_0 + X^\top D^{-1} y),\ D = \mathrm{diag}(\tau),$$

luego

$$\tau_i \mid y, \beta, \psi \ \propto\ \tau_i^{-1/2} \exp\left\{-\frac{1}{2}\left(\frac{(y_i - x_i^\top \beta)^2}{\tau_i} + \psi\,\tau_i\right)\right\} \ \sim\ \mathrm{GIG}\Big(\lambda = \tfrac{1}{2},\ \chi = (y_i - x_i^\top \beta)^2,\ \psi\Big),$$

y

$$\psi \mid \tau \ \sim\ \mathrm{Gamma}\left(a_0 + n,\ b_0 + \tfrac{1}{2}\sum_{i=1}^{n} \tau_i\right), \qquad \text{y}\ \ b = \psi^{-1/2}.$$

Aquí $\mathrm{GIG}(\lambda, \chi, \psi)$ denota la Gaussiana Inversa Generalizada con densidad $f(t) \propto t^{\lambda-1} \exp\{-(\chi/t + \psi t)/2\}$ en $t > 0$. Esta construcción proporciona un algoritmo MCMC totalmente conjugado por aumentación. La implementación computacional de lo anteriro expuesto se detalla en Park and Casella (2008).

Interpretativamente, el uso de errores Laplace en regresión proporciona robustez frente a outliers pero conlleva importantes consecuencias metodológicas: pérdida de conjugación bayesiana y necesidad de métodos de inferencia computacionalmente más intensivos. ■

Referencias

Andrews, D. F. and Mallows, C. L. (1974). Scale mixtures of normal distributions. *Journal of the Royal Statistical Society: Series B*, 36(1):99–102.

Park, T. and Casella, G. (2008). The bayesian lasso. *Journal of the American Statistical Association*, 103(482):681–686.

West, M. (1987). On scale mixtures of normal distributions. *Biometrika*, 74(3):646–648.