Clasificación y regresión logística

José R. Berrendero

Universidad Autónoma de Madrid

Repaso: distribución asintótica de los EMV

Si $\hat{\theta}$ es el EMV de $\theta \in \Theta \subset \mathbb{R}$, bajo ciertas condiciones de regularidad se verifica:

$$I(\theta)^{1/2}(\hat{\theta} - \theta) \to_d \mathrm{N}(0, 1),$$

donde $I(\theta)$ es la **información de Fisher**,

$$I(\theta) = -\mathbb{E}\left(\ell''(\theta)\right).$$

Es habitual estimar $I(\theta)$ mediante la **información observada** $\ell''(\hat{\theta})$, de manera que si n es grande:

$$\hat{\theta} \cong \mathrm{N}(\theta, -\ell''(\hat{\theta})^{-1}).$$

Si $\theta \in \Theta \subset \mathbb{R}^p$ es un vector, entonces

$$I(\theta) = -\mathbb{E}\left(H\ell(\theta)\right),$$

donde $H\ell(\theta)$ es la matriz Hessiana de $\ell(\theta)$.

Aplicación al modelo de regresión logística

Hemos visto que $\nabla \ell(\beta) = \sum_{i=1}^{n} (Y_i x_i - p_i x_i) = X'Y - X'p$, con $p = (p_1, \ldots, p_n)'$.

Por lo tanto,

$$\frac{\partial \ell(\beta)}{\partial \beta_j} = \text{constante} - \sum_{i=1}^{n} p_i x_{ij}.$$

$$\frac{\partial^2 \ell(\beta)}{\partial \beta_j \partial \beta_r} = - \sum_{i=1}^{n} p_i (1 - p_i) x_{ir} x_{ij}.$$

La fila j de la matriz Hessiana es

$$-\sum_{i=1}^{n} p_i (1 - p_i) x_{ij} x_i' = -(1, x_{1j}, \ldots, x_{nj}) \begin{pmatrix} p_1(1-p_1) & & \\ & \ddots & \\ & & p_n(1-p_n) \end{pmatrix} \begin{pmatrix} x_1' \\ \vdots \\ x_n' \end{pmatrix}.$$

Por lo tanto,

$$H\ell(\beta) = -X'WX, \quad \text{con} \quad W = \begin{pmatrix} p_1(1-p_1) & & \\ & \ddots & \\ & & p_n(1-p_n) \end{pmatrix}$$

Algoritmo de Newton-Raphson

El método de Newton para resolver $0 = f'(x_0)$ se basa en la aproximación:

$$0 = f(x_0) \approx f'(x) + f''(x)(x_0 - x),$$

es decir,

$$x_0 \approx x - [f''(x)]^{-1} f'(x).$$

Para vectores en lugar de las derivadas se usan la matrix Hessiana y el vector gradiente, lo que sugiere el algoritmo:

$$x^{(m+1)} = x^{(m)} - Hf(x^{(m)})^{-1} \nabla f(x^{(m)})$$

En el caso particular de regresión logística:

$$\beta^{(m+1)} = \beta^{(m)} + (X' W^{(m)} X)^{-1} (X' Y - X' p^{(m)}).$$

Distribución aproximada de $\hat{\beta}$

Aplicando la teoría asintótica de los EMV tenemos:

$$\hat{\beta} \cong \mathsf{N}_{k+1}(\beta, (X'\hat{W}X)^{-1})$$

donde $\hat{W} = \mathrm{diag}(\hat{p}_1(1-\hat{p}_1), \ldots, \hat{p}_n(1-\hat{p}_n))$.

Podemos escribir también

$$(X'\hat{W}X)^{-1} = \left[\sum_{i=1}^{n} \hat{p}_i(1-\hat{p}_i) x_i x_i'\right]^{-1}$$

La contribución de la observación x_i a la matriz de covarianzas de $\hat{\beta}$ es proporcional a la estimación de $\mathrm{Var}(Y_i) = p_i(1-p_i)$.