# MAT-466: Definición de GLMs

**Felipe Osorio**

fosorios.mat.utfsm.cl

Departamento de Matemática, UTFSM

**Definición 1 (Modelo lineal generalizado):**

Suponga $Y_1, \dots, Y_n$ variables aleatorias independientes cada una con densidad

$$f(y_i; \theta_i, \phi) = \exp[\phi\{y_i\theta_i - b(\theta_i)\} + c(y_i, \phi)], \qquad (1)$$

donde

$$\mathsf{E}(Y_i) = \mu_i = b'(\theta_i), \qquad \mathsf{var}(Y_i) = \phi^{-1}V(\mu_i),$$

con $V(\mu) = \mathsf{d}\,\mu/\,\mathsf{d}\,\theta$ la función de varianza[1] y $\phi^{-1} > 0$ es el parámetro de dispersión.

Los modelos lineales generalizados (GLM) son definidos por (1) y

$$g(\mu_i) = \eta_i, \qquad \eta_i = \boldsymbol{x}_i^\top \boldsymbol{\beta},$$

donde $\eta_i$ corresponde al predictor lineal, $\boldsymbol{\beta} = (\beta_1, \dots, \beta_p)^\top$, $(p < n)$ es el vector de coeficientes de regresión y $\boldsymbol{x}_i = (x_{i1}, \dots, x_{ip})^\top$ denota un vector de variables regresoras. Además, la función $g(\cdot)$ es monótona y diferenciable y se denomina función de enlace.

[1] En adelante anotaremos simplemente, $V_i = V(\mu_i)$, para $i = 1, \dots, n$.

**Definición 1 (Modelo lineal generalizado):**

Suponga $Y_1, \dots, Y_n$ variables aleatorias independientes cada una con densidad

$$f(y_i;\theta_i,\phi) = \exp[\phi\{y_i\theta_i - b(\theta_i)\} + c(y_i,\phi)], \tag{1}$$

donde

$$\mathsf{E}(Y_i) = \mu_i = b'(\theta_i), \qquad \mathsf{var}(Y_i) = \phi^{-1}V(\mu_i),$$

con $V(\mu) = \mathrm{d}\,\mu/\,\mathrm{d}\,\theta$ la función de varianza[1] y $\phi^{-1} > 0$ es el parámetro de dispersión.

Los modelos lineales generalizados (GLM) son definidos por (1) y

$$g(\mu_i) = \eta_i, \qquad \eta_i = \boldsymbol{x}_i^\top\boldsymbol{\beta},$$

donde $\eta_i$ corresponde al predictor lineal, $\boldsymbol{\beta} = (\beta_1, \dots, \beta_p)^\top$, $(p < n)$ es el vector de coeficientes de regresión y $\boldsymbol{x}_i = (x_{i1}, \dots, x_{ip})^\top$ denota un vector de variables regresoras. Además, la función $g(\cdot)$ es monótona y diferenciable y se denomina función de enlace.

[1] En adelante anotaremos simplemente, $V_i = V(\mu_i)$, para $i = 1, \dots, n$.

*Ejemplo (distribución normal):*

Considere $Y \sim \mathsf{N}(\mu, \sigma^2)$, $\mu \in \mathbb{R}$ y $\sigma^2 > 0$. De este modo

$$\begin{aligned} f(y;\mu,\sigma^2) &= (2\pi\sigma^2)^{-1/2} \exp\Big\{ -\frac{1}{2}\Big(\frac{y-\mu}{\sigma}\Big)^2 \Big\} \\ &= \exp\Big\{ \frac{1}{\sigma^2}\Big(\mu y - \frac{\mu^2}{2}\Big) - \frac{1}{2}\Big(\log 2\pi\sigma^2 + \frac{y^2}{\sigma^2}\Big)\Big\}. \end{aligned}$$

Luego, $\theta = \mu$, $b(\theta) = \theta^2/2$, $\phi = 1/\sigma^2$ y

$$c(y,\phi) = \frac{1}{2}\log(\pi/2\pi) - \frac{\phi y^2}{2}.$$

Verificamos fácilmente que $V(\mu) = 1$.

*Ejemplo (distribución Bernoulli):*

Suponga $Y \sim \text{Ber}(\mu)$, $\mu \in (0,1)$. Tenemos

$$\begin{aligned} f(y;\mu) &= \mu^y (1-\mu)^{1-y} \\ &= \exp\{y \log \mu + (1-y)\log(1-\mu)\} \\ &= \exp\left\{ y \log\left(\frac{\mu}{1-\mu}\right) + \log(1-\mu)\right\}. \end{aligned}$$

De ahí que

$$\theta = \log\left(\frac{\mu}{1-\mu}\right) \qquad \Rightarrow \qquad \mu = \frac{e^\theta}{1+e^\theta},$$

de donde obtenemos $b(\theta) = \log(1+e^\theta)$ y $\phi = 1$. Es fácil verificar

$$\begin{aligned} b'(\theta) &= \frac{1}{1+e^\theta}\frac{\mathsf{d}}{\mathsf{d}\,\theta}(1+e^\theta) = \frac{e^\theta}{1+e^\theta} = \mu \\ b''(\theta) &= \frac{\mathsf{d}}{\mathsf{d}\,\theta}\left(\frac{e^\theta}{1+e^\theta}\right) = \frac{e^\theta(1+e^\theta) - e^{2\theta}}{(1+e^\theta)^2} = \mu - \mu^2 = \mu(1-\mu) = V(\mu). \end{aligned}$$

*Ejemplo (distribución Poisson):*

Sea $Y \sim \text{Poi}(\mu)$, $\mu > 0$, con densidad

$$f(y;\mu) = \frac{e^{-\mu}\mu^y}{y!} = \exp(y\log\mu - \mu - \log y!),$$

sigue que $\theta = \log\mu$, $b(\theta) = e^\theta$, $\phi = 1$ y $c(y,\phi) = -\log y!$ Además,

$$b'(\theta) = e^\theta = \mu, \qquad b''(\theta) = e^\theta = \mu = V(\mu).$$

*Ejemplo (distribución Gama):*

Considere $Y \sim$ Gama$(\mu, \phi)$, cuya densidad es dada por

$$\begin{aligned} f(y;\mu,\phi) &= \frac{y^{-1}}{\Gamma(\phi)}\Big(\frac{\phi y}{\mu}\Big)^{\phi}\exp\Big(-\frac{\phi y}{\mu}\Big) \\ &= \exp\{\phi(-y/\mu - \log\mu) - \log\Gamma(\phi) + \phi\log(\phi y) - \log y\}. \end{aligned}$$

De ahi que $\theta = -1/\mu$, $b(\theta) = -\log(-\theta)$, y

$$c(y;\phi) = (\phi - 1)\log y + \phi\log\phi - \log\Gamma(\phi).$$

Ahora,

$$b'(\theta) = -\frac{1}{\theta} = \mu, \qquad b''(\theta) = \frac{1}{\theta}.$$

Por tanto, $V(\mu) = \mu^2$, y var$(Y) = \mu^2/\phi$.

*Ejemplo (distribución normal inversa):*

Sea $Y \sim \mathsf{IG}(\mu, \phi)$, con densidad

$$\begin{aligned} f(y;\mu,\phi) &= \Big(\frac{\phi}{2\pi y^3}\Big)^{1/2} \exp\Big\{-\frac{\phi}{2y}\Big(\frac{y-\mu}{\mu}\Big)^2\Big\} \\ &= \exp\Big\{\phi\Big(-\frac{y}{2\mu^2}+\frac{1}{\mu}\Big) - \frac{1}{2}\Big(\log(2\pi y^3/\phi) + \frac{\phi}{y}\Big)\Big\} \end{aligned}$$

Haciendo $\theta = -1/(2\mu^2)$, $b(\theta) = -(-2\theta)^{1/2}$, y

$$c(y;\phi) = \frac{1}{2}\log\Big(\frac{\phi}{2\pi y^3}\Big) - \frac{\phi}{2y}.$$

Para $Y \sim \mathsf{IG}(\mu, \phi)$ sigue que $V(\mu) = \mu^3$.

**Definición 2 (enlace canónico):**

Sea $\mu = \tau(\theta) = b'(\theta)$. La transformación inversa

$$\theta = \tau^{-1}(\mu),$$

es llamada función de enlace.

*Ejemplos:*

- Normal: $\theta = \mu$.
- Bernoulli: $\theta = \log\left(\frac{\mu}{1-\mu}\right) = \operatorname{logit}(\mu)$.
- Poisson: $\theta = \log \mu$.
- Gamma: $\theta = 1/\mu$.
- Normal inversa: $\theta = 1/\mu^2$.

# Modelo lineal generalizado

Suponga $Y \sim \text{Ber}(\mu)$, $\mu \in (0,1)$. Considere la distribución logística con función de densidad

$$f(y) = \frac{e^y}{(1+e^y)^2} \qquad \Longleftrightarrow \qquad F(y) = \frac{e^y}{1+e^y},$$

de ahí que podemos escribir la función de enlace logístico como:

$$\theta = \operatorname{logit}(\mu) = F^{-1}(\mu).$$

Lo anterior motiva las siguientes funciones de enlace:

- Enlace probit: Considere la transformación:

  $$\theta = \Phi^{-1}(\mu).$$

- Enlace complemento log-log: Considere

  $$f(y) = \exp(y - e^y), \qquad F(y) = 1 - \exp(-e^y),$$

  de ahí que

  $$\theta = \log(-\log(1-\mu)).$$

- Enlace Box-Cox: Este enlace es definido para variables aleatorias positivas

$$\theta = \frac{\mu^{\lambda} - 1}{\lambda}, \qquad \lambda \neq 0,$$

note que, para $\lambda \to 0$ obtenemos $\theta = \log \mu$.

- Enlace Aranda-Ordaz: esta función de enlace se define por la transformación:

$$\theta = \log\left\{\frac{(1-\mu)^{-\alpha} - 1}{\alpha}\right\}, \qquad \mu \in (0,1).$$

Cuando $\alpha = 1$, tenemos $\theta = \operatorname{logit}(\mu)$, mientras que $\alpha \to 0$, sigue que $\theta = \log(-\log(1-\mu))$.

Suponga $Y_1, \dots, Y_n$ variables aleatorias independientes cada una con distribución FE$(\theta_i, \phi)$. Entonces, la densidad conjunta adopta la forma:

$$f(\boldsymbol{y}; \boldsymbol{\theta}, \phi) = \exp\Big[\phi \sum_{i=1}^{n} \{y_i \theta_i - b(\theta_i)\} + \sum_{i=1}^{n} c(y_i; \phi)\Big].$$

De ahí que, la función de log-verosimilitud es dada por

$$\ell(\boldsymbol{\theta}) = \phi \sum_{i=1}^{n} \{y_i \theta_i - b(\theta_i)\} + \sum_{i=1}^{n} c(y_i; \phi).$$

Sea $\mu_i = g^{-1}(\eta_i)$, $\eta_i = \boldsymbol{x}_i^\top \boldsymbol{\beta}$, y considere $\ell(\boldsymbol{\mu}; \boldsymbol{y})$, $\ell(\boldsymbol{y}; \boldsymbol{y})$, las funciones de log-verosimilitud para el modelo de trabajo con $\mu_i = g^{-1}(\eta_i)$, $\eta_i = \boldsymbol{x}_i^\top \boldsymbol{\beta}$ y el modelo saturado $(p = n)$, respectivamente.

Recuerde que podemos escribir

$$D^*(\boldsymbol{y};\widehat{\boldsymbol{\mu}}) = \phi D(\boldsymbol{y};\widehat{\boldsymbol{\mu}}) = 2\{\ell(\boldsymbol{y};\boldsymbol{y}) - \ell(\widehat{\boldsymbol{\mu}};\boldsymbol{y})\}.$$

Suponga que $\widehat{\theta}_i = \theta_i(\widehat{\mu}_i)$ y $\widetilde{\theta}_i = \theta_i(\widetilde{\mu}_i)$ los MLE de $\boldsymbol{\theta}$ para el modelo con $p$ parámetros $(p < n)$ y el modelo saturado $(p = n)$, respectivamente. De ahí que

$$\ell(\widehat{\boldsymbol{\mu}};\boldsymbol{y}) = \phi \sum_{i=1}^{n}\{y_i\widehat{\theta}_i - b(\widehat{\theta}_i)\} + \sum_{i=1}^{n} c(y_i;\phi)$$

$$\ell(\widetilde{\boldsymbol{\mu}};\boldsymbol{y}) = \phi \sum_{i=1}^{n}\{y_i\widetilde{\theta}_i - b(\widetilde{\theta}_i)\} + \sum_{i=1}^{n} c(y_i;\phi).$$

Ahora,

$$\begin{aligned} D^*(\boldsymbol{y};\widehat{\boldsymbol{\mu}}) &= 2\{\ell(\boldsymbol{y};\boldsymbol{y}) - \ell(\widehat{\boldsymbol{\mu}};\boldsymbol{y})\} \\ &= 2\phi \sum_{i=1}^{n}[y_i(\widetilde{\theta}_i - \widehat{\theta}_i) + \{b(\widehat{\theta}_i) - b(\widetilde{\theta}_i)\}]. \end{aligned}$$

Finalmente, podemos escribir

$$D(\boldsymbol{y}; \widehat{\boldsymbol{\mu}}) = \sum_{i=1}^{n} d_i^2(y_i; \widehat{\mu}_i),$$

con

$$d_i^2(y_i; \widehat{\mu}_i) = 2\{y_i(\widetilde{\theta}_i - \widehat{\theta}_i) + (b(\widehat{\theta}_i) - b(\widetilde{\theta}_i))\},$$

es conocido como el componente del desvío (no escalado).

*Ejemplo (modelo normal):*

Tenemos $\theta_i = \mu_i$, luego $\widetilde{\theta}_i = y_i$ y $\widehat{\theta}_i = \widehat{\mu}_i$. Luego, la función desvio asume la forma:

$$\begin{aligned} D(\boldsymbol{y}; \widehat{\boldsymbol{\mu}}) &= 2\sum_{i=1}^{n} \Big\{ y_i(y_i - \widehat{\mu}_i) + \frac{\widehat{\mu}_i^2}{2} - \frac{y_i^2}{2} \Big\} \\ &= \sum_{i=1}^{n} (y_i - \widehat{\mu}_i)^2 \quad (= RSS) \end{aligned}$$

*Ejemplo (modelo Poisson):*

En este caso $\theta_i = \log \mu_i$. De ahí que,

$$D(\boldsymbol{y}; \widehat{\boldsymbol{\mu}}) = 2\sum_{i=1}^{n} \Big\{ y_i \log\Big(\frac{y_i}{\widehat{\mu}_i}\Big) - (y_i - \widehat{\mu}_i) \Big\}.$$

Para $y_i = 0$ tenemos

$$f(y_i; \mu_i) = \frac{e^{-\mu_i}\mu_i^0}{0!} = e^{-\mu_i},$$

en cuyo caso

$$d_i^2(y_i; \widehat{\mu}_i) = 2\widehat{\mu}_i.$$

*Ejemplo (modelo Gama):*

Tenemos $\theta_i = -1/\mu_i$. Así,

$$D(\boldsymbol{y}; \widehat{\boldsymbol{\mu}}) = 2 \sum_{i=1}^{n} \Big\{ -\log\Big(\frac{y_i}{\widehat{\mu}_i}\Big) + \frac{y_i - \widehat{\mu}_i}{\widehat{\mu}_i} \Big\}.$$

*Ejemplo (modelo normal inversa):*

Sabemos que $\theta_i = -1/(2\mu_i^2)$, por tanto

$$D(\boldsymbol{y}; \widehat{\boldsymbol{\mu}}) = \sum_{i=1}^{n} \frac{(y_i - \widehat{\mu}_i)^2}{y_i \widehat{\mu}_i^2}.$$