

# MGE-201: Familia exponencial

Felipe Osorio

[felipe.osorio@uv.cl](mailto:felipe.osorio@uv.cl)

## Familia exponencial

### Definición 1 (Familia exponencial):

La clase de modelos  $\{P_\theta : \theta \in \Theta\}$  se dice una familia exponencial 1-paramétrica si existen funciones  $\eta(\theta)$ ,  $b(\theta)$  y funciones real valuadas  $T$  y  $h$  tal que su densidad adopta la forma

$$f(x) = \exp[\eta(\theta)T(x) - b(\theta)]h(x), \quad (1)$$

donde  $x \in \mathcal{X} \subset \mathbb{R}^q$ .

Cuando una variable aleatoria tiene la densidad en Ecuación (1), anotamos  $X \sim \text{FE}(\theta)$ .

## Familia exponencial

*Observación:*

El parámetro  $\eta := \eta(\theta)$  es llamado **parámetro natural**, y

$$b(\eta) = \log \int \exp(\eta T(x)) h(x) dx,$$

corresponde a un factor de normalización. Además, el modelo estadístico puede ser escrito como:

$$\mathcal{P} = \{P_\eta : \eta \in \Gamma\}, \quad \Gamma := \eta(\Theta), \tag{2}$$

donde el conjunto

$$\Gamma = \{\eta : b(\eta) < \infty\},$$

se denomina **espacio paramétrico natural**.

Cuando escribimos la familia de densidades  $\{P_\eta : \eta \in \Gamma\}$  como en (2), decimos que la familia está escrita en **forma canónica**.

## Modelo estadístico

*Ejemplo:*

Sea  $P_\theta \stackrel{d}{=} \text{Poisson}(\theta)$  con parámetro de media desconocido. Entonces

$$p(x; \theta) = \frac{\theta^x e^{-\theta}}{x!} = \frac{1}{x!} \exp(x \log \theta - \theta), \quad \theta > 0,$$

con  $x \in \{0, 1, 2, \dots\}$ . Es decir,  $P_\theta$  corresponde a una **FE** (1-paramétrica) donde  $q = 1$ ,

$$\eta(\theta) = \log \theta, \quad T(x) = x, \quad b(\theta) = \theta, \quad h(x) = 1/x!$$

## Modelo estadístico

*Ejemplo:*

Sea  $P_\theta \stackrel{d}{=} \text{Bin}(n, \theta)$ ,  $\theta \in (0, 1)$  y  $x \in \{0, 1, \dots, n\}$ . De este modo,

$$\begin{aligned} p(x; \theta) &= \binom{n}{x} \theta^x (1 - \theta)^{n-x} \\ &= \binom{n}{x} \exp[x \log \theta + (n - x) \log(1 - \theta)] \\ &= \binom{n}{x} \exp \left[ x \log \left( \frac{\theta}{1 - \theta} \right) + n \log(1 - \theta) \right], \end{aligned}$$

y  $\text{Bin}(n, \theta)$  está en FE (1-paramétrica) con  $q = 1$ ,

$$\eta(\theta) = \log \left( \frac{\theta}{1 - \theta} \right), \quad T(x) = x, \quad b(\theta) = -n \log(1 - \theta), \quad h(x) = \binom{n}{x}.$$

## Modelo estadístico

### Observación:

Note que las funciones  $\eta$ ,  $b$  y  $T$  **no son únicas**. En efecto, una manera alternativa de escribir la densidad de la familia exponencial es, por ejemplo,

$$f(x) = a(\theta) \exp[\eta(\theta)T(x)]h(x),$$

donde  $h(x) \geq 0$  y  $a(\theta) > 0$ , representa un factor de normalización.

### Contraejemplo:

Sea  $\mathcal{P}$  la clase de distribuciones exponencial de dos parámetros, digamos  $\text{Exp}(1, \theta)$ , con densidad

$$f(x) = \exp[-(x - \theta)]I_{[\theta, \infty)}(x), \quad \theta \in \mathbb{R}.$$

Esta familia **no pertenece** a la clase FE.<sup>1</sup>

---

<sup>1</sup>El factor  $I_{[\theta, \infty)}(x)$  no puede ser escrito en la forma exponencial.

## Modelo estadístico

Suponga que  $X_1, \dots, X_n$  son variables aleatorias IID con distribución común  $P_\theta$  en la FE 1-paramétrica y  $\theta \in \Theta$ . De este modo,

$$\begin{aligned} f(\mathbf{x}) &= \prod_{i=1}^n \exp[\eta(\theta)T(x_i) - b(\theta)]h(x_i) \\ &= \exp\left[\eta(\theta)\sum_{i=1}^n T(x_i) - nb(\theta)\right] \prod_{i=1}^n h(x_i) \\ &= \exp[\eta(\theta)T^{(n)}(\mathbf{x}) - b^{(n)}(\theta)]\tilde{h}(\mathbf{x}) \end{aligned}$$

con

$$T^{(n)}(\mathbf{x}) = \sum_{i=1}^n T(x_i), \quad b^{(n)}(\theta) = nb(\theta), \quad \tilde{h}(\mathbf{x}) = \prod_{i=1}^n h(x_i). \quad (3)$$

Es decir, la distribución conjunta de  $\mathbf{X} = (X_1, \dots, X_n)^\top$  también pertenece a la FE 1-paramétrica.

## Familia exponencial

### Resultado 1:

Si  $X$  está en la FE 1-paramétrica. La función generadora de momentos de  $T(X)$  es dada por

$$M_T(s) = \exp[b(s + \eta) - b(\eta)],$$

para  $s$  en una vecindad de cero.

### Demostración:

Para obtener la MGF de  $T(X)$ , debemos calcular

$$\begin{aligned} M_T(s) &= \mathbb{E}\{\exp(sT(x))\} = \int_{\mathcal{X}} \exp(sT(x)) \exp(\eta T(x) - b(\eta)) h(x) dx \\ &= \int_{\mathcal{X}} \exp[(s + \eta)T(x) - b(\eta)] h(x) dx \\ &= \exp[b(s + \eta) - b(\eta)] \int_{\mathcal{X}} \exp[(s + \eta)T(x) - b(s + \eta)] h(x) dx \\ &= \exp[b(s + \eta) - b(\eta)]. \end{aligned}$$

## Modelo estadístico

La **función generadora de cumulantes** está definida como

$$K_T(s) = \log M_T(s),$$

y permite obtener los cumulantes de  $T$ ,

$$\kappa_r = \left. \frac{d^r}{ds^r} K_T(s) \right|_{s=0}.$$

En particular las primeras dos derivadas de  $K_T(s)$  asumen la forma,

$$K'_T = M'_T/M_T, \quad K''_T = [M_T M''_T - (M'_T)^2]/M_T^2,$$

y evaluando en  $s = 0$ , lleva a los primeros cumulantes:

$$\kappa_1 = E(T), \quad \kappa_2 = E(T^2) - (E(T))^2 = \text{var}(T).$$

Para la familia exponencial, sigue que la función generadora de cumulantes para  $T$  es dada por:

$$K_T(s) = b(s + \eta) - b(\eta).$$

## Familia exponencial

### Resultado 2:

Sea  $X$  una variable aleatoria perteneciente a la FE 1-paramétrica. De este modo,

$$\mathbb{E}(T) = b'(\eta), \quad \text{var}(T) = b''(\eta).$$

### Demostración:

Basta notar que

$$\mathbb{E}(T) = \frac{d}{ds} K_T(s) \Big|_{s=0} = b'(s + \eta) \Big|_{s=0} = b'(\eta),$$

$$\text{var}(T) = \frac{d^2}{ds^2} K_T(s) \Big|_{s=0} = b''(s + \eta) \Big|_{s=0} = b''(\eta).$$

## Familia exponencial

Para una muestra IID,  $X_1, \dots, X_n$ , y usando el [Resultado 2](#) sigue inmediatamente que:

$$\mathbb{E}(T^{(n)}(\mathbf{X})) = \mathbb{E} \left( \sum_{i=1}^n T(X_i) \right) = \sum_{i=1}^n \mathbb{E}(T(X_i)) = nb'(\theta)$$

$$\text{var}(T^{(n)}(\mathbf{X})) = \text{var} \left( \sum_{i=1}^n T(X_i) \right) = \sum_{i=1}^n \text{var}(T(X_i)) = nb''(\theta).$$

## Familia exponencial

Sea  $\mu = E(Y)$ , es decir, tenemos una transformación<sup>2</sup> 1-1 entre  $\theta$  y el parámetro de media

$$\mu = \mu(\theta) = b'(\theta),$$

pues  $\eta = \eta(\theta)$ . Además, conforme  $\theta$  varía en  $\Theta$ , tenemos que  $\mu(\theta)$  es conocida como **función (o superficie) de esperanza**.

Cuando  $Y \sim FE(\theta)$  con función generadora de cumulantes  $b(\cdot)$ , tenemos  $\mu(\theta) = b'(\theta)$ . Podemos parametrizar la densidad  $f(y)$  en términos de  $\mu$ . De este modo, haciendo  $\theta = \theta(\mu)$ , sigue que

$$\text{var}(Y) = b''(\theta) = \frac{d\mu}{d\theta} \Big|_{\theta=\theta(\mu)} = V(\mu),$$

donde  $V(\mu)$  es llamada **función de varianza** de  $Y$ .

---

<sup>2</sup>suave y estrictamente convexa

## Familia exponencial

### Definición 2:

La familia de distribuciones  $\{P_\theta : \theta \in \Theta\}$  con  $\Theta \subset \mathbb{R}^k$ , se dice una **familia exponencial  $k$ -paramétrica** si existen funciones  $\eta_1(\theta), \dots, \eta_k(\theta)$  y  $b(\theta)$  y funciones real-valuadas  $T_1, \dots, T_k, h$ , tal que la densidad puede ser escrita como:

$$f(\mathbf{x}; \boldsymbol{\theta}) = \exp[\boldsymbol{\eta}^\top(\boldsymbol{\theta})\mathbf{T}(\mathbf{x}) - b(\boldsymbol{\theta})]h(\mathbf{x}), \quad \mathbf{x} \in \mathcal{X} \subset \mathbb{R}^p,$$

donde  $\boldsymbol{\eta}(\boldsymbol{\theta}) = (\eta_1(\boldsymbol{\theta}), \dots, \eta_k(\boldsymbol{\theta}))^\top$  y  $\mathbf{T}(\mathbf{x}) = (T_1(\mathbf{x}), \dots, T_k(\mathbf{x}))^\top$ .

## Familia exponencial

*Ejemplo:*

Suponga  $P_\theta \stackrel{d}{=} N_1(\mu, \sigma^2)$ ,

$$\Theta = \{(\mu, \sigma^2) : \mu \in \mathbb{R}, \sigma^2 > 0\}.$$

La densidad de  $P_\theta$  puede ser escrita como

$$\begin{aligned} f(x; \boldsymbol{\theta}) &= (2\pi\sigma^2)^{-1/2} \exp \left\{ -\frac{1}{2} \left( \frac{x-\mu}{\sigma} \right)^2 \right\} \\ &= \exp \left\{ \frac{\mu x}{\sigma^2} - \frac{x^2}{2\sigma^2} - \frac{1}{2} \left( \frac{\mu^2}{\sigma^2} + \log 2\pi\sigma^2 \right) \right\}. \end{aligned}$$

De este modo,

$$\eta_1(\boldsymbol{\theta}) = \frac{\mu}{\sigma^2}, \quad \eta_2(\boldsymbol{\theta}) = -\frac{1}{2\sigma^2}, \quad T_1(x) = x, \quad T_2(x) = x^2,$$

$$\text{y } b(\boldsymbol{\theta}) = \frac{1}{2}(\mu^2/\sigma^2 + \log 2\pi\sigma^2), \quad h(x) = 1.$$

Luego,  $N_1(\mu, \sigma)$  corresponde a una **FE 2-paramétrica**.

## Familia exponencial

*Ejemplo:*

Considere la función de probabilidad

$$p(x; \boldsymbol{\theta}) = \theta_1^{I_1(x)} \theta_2^{I_2(x)} \theta_3^{I_3(x)},$$

con

$$I_j(x) = \begin{cases} 1, & x = j, \\ 0, & \text{en otro caso,} \end{cases}$$

para  $j = 1, 2, 3$ . Así podemos escribir

$$p(x; \boldsymbol{\theta}) = \exp[I_1(x) \log \theta_1 + I_2(x) \log \theta_2 + I_3(x) \log \theta_3],$$

es decir  $X$  pertenece a la **FE**.

*Sin embargo,*

$$I_1(x) + I_2(x) + I_3(x) = 1.$$

Por tanto, la familia **no** es estrictamente 3-dimensional.

## Familia exponencial

Consideremos

$$\begin{aligned} p(x; \boldsymbol{\theta}) &= \exp[I_1(x) \log \theta_1 + I_2(x) \log \theta_2 + (1 - I_1(x) - I_2(x)) \log \theta_3] \\ &= \theta_3 \exp[I_1(x)(\log \theta_1 - \log \theta_3) + I_2(x)(\log \theta_2 - \log \theta_3)]. \end{aligned}$$

Sea  $\theta_3 = 1 - \theta_1 - \theta_2$ , sigue que

$$p(x; \boldsymbol{\theta}) = \exp \left[ I_1(x) \log \left( \frac{\theta_1}{1 - \theta_1 - \theta_2} \right) + I_2(x) \log \left( \frac{\theta_2}{1 - \theta_1 - \theta_2} \right) + \log(1 - \theta_1 - \theta_2) \right].$$

Finalmente, tenemos que  $X$  pertenece a la **FE** (2-paramétrica) con

$$T_1(x) = I_1(x), \quad T_2(x) = I_2(x), \quad \eta_r = \log \left( \frac{\theta_r}{1 - \theta_1 - \theta_2} \right), \quad r = 1, 2,$$

y  $b(\boldsymbol{\theta}) = -\log(1 - \theta_1 - \theta_2)$ ,  $h(x) = 1$ .

## Familia exponencial

*Observación:*

Evidentemente, resulta fácil extender<sup>3</sup> los [Resultados 1](#) y [2](#). En efecto,

$$M_T(\mathbf{s}) = \mathbb{E}[\exp(\mathbf{s}^\top \mathbf{T})] = \exp[b(\boldsymbol{\eta} + \mathbf{s}) - b(\boldsymbol{\eta})],$$

$$K_T(\mathbf{s}) = b(\boldsymbol{\eta} + \mathbf{s}) - b(\boldsymbol{\eta}),$$

y

$$\mathbb{E}(\mathbf{T}(\mathbf{X})) = \dot{b}(\boldsymbol{\eta}) = \frac{\partial b(\boldsymbol{\eta})}{\partial \boldsymbol{\eta}}, \quad \text{Cov}(\mathbf{T}(\mathbf{X})) = \ddot{b}(\boldsymbol{\eta}) = \frac{\partial^2 b(\boldsymbol{\eta})}{\partial \boldsymbol{\eta} \partial \boldsymbol{\eta}^\top}.$$

---

<sup>3</sup>Revisar los detalles

# Familia exponencial

## Definición 3:

Se dice que  $P_{\theta, \phi}$  sigue una familia exponencial con parámetro de escala  $\phi > 0$ , si la función de densidad de un vector aleatorio  $n$ -dimensional  $\mathbf{Y} = (Y_1, \dots, Y_n)^\top$  es dada por:

$$f(\mathbf{y}; \boldsymbol{\theta}, \phi) = \exp[\phi\{\mathbf{y}^\top \boldsymbol{\theta} - b(\boldsymbol{\theta})\} + c(\mathbf{y}, \phi)],$$

donde  $b(\cdot)$  y  $c(\cdot, \cdot)$  son funciones apropiadas,  $\boldsymbol{\theta} = (\theta_1, \dots, \theta_n)^\top$  y denota el parámetro natural ( $\boldsymbol{\theta} \in \Theta \subset \mathbb{R}^n$ ). En cuyo caso escribimos  $\mathbf{Y} \sim \text{FE}(\boldsymbol{\theta}, \phi)$ .<sup>4</sup>

## Observación:

Cuando  $\phi$  es desconocido, se suele indicar que el modelo está en la familia de dispersión exponencial (Jørgensen, 1997).<sup>5</sup>

---

<sup>4</sup>El parámetro de dispersión  $\phi > 0$  usualmente es considerado un parámetro molesto.

<sup>5</sup>Jørgensen, B. (1997). *The Theory of Dispersion Models*. Chapman & Hall, London.

## Familia exponencial

*Ejemplo:*

Suponga  $\mathbf{Y} \sim N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ . Tenemos que su función de densidad puede ser escrita como

$$\begin{aligned}f(\mathbf{y}; \boldsymbol{\mu}, \boldsymbol{\Sigma}) &= (2\pi)^{-p/2} |\boldsymbol{\Sigma}|^{-1/2} \exp \left\{ -\frac{1}{2} (\mathbf{y} - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1} (\mathbf{y} - \boldsymbol{\mu}) \right\} \\&= \exp \left\{ \mathbf{y}^\top \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu} - \frac{1}{2} \boldsymbol{\mu}^\top \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu} - \frac{1}{2} \mathbf{y}^\top \boldsymbol{\Sigma}^{-1} \mathbf{y} - \frac{1}{2} \log |2\pi \boldsymbol{\Sigma}| \right\} \\&= \exp \left\{ \mathbf{y}^\top \boldsymbol{\theta} - \frac{1}{2} \boldsymbol{\theta}^\top \boldsymbol{\Sigma} \boldsymbol{\theta} - \frac{1}{2} \mathbf{y}^\top \boldsymbol{\Sigma}^{-1} \mathbf{y} - \frac{1}{2} \log |2\pi \boldsymbol{\Sigma}| \right\},\end{aligned}$$

donde  $\boldsymbol{\theta} = \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu}$  es el parámetro natural, y

$$b(\boldsymbol{\theta}) = \frac{1}{2} \boldsymbol{\theta}^\top \boldsymbol{\Sigma} \boldsymbol{\theta}, \quad c(\mathbf{y}, \boldsymbol{\Phi}) = -\frac{1}{2} (\mathbf{y}^\top \boldsymbol{\Phi} \mathbf{y} + \log |2\pi \boldsymbol{\Phi}^{-1}|),$$

con  $\boldsymbol{\Phi} = \boldsymbol{\Sigma}^{-1}$  el parámetro de dispersión.

## Familia exponencial

*Ejemplo:*

Considere  $\mathbf{Y} \sim \mathcal{N}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ , con densidad

$$\begin{aligned} f(\mathbf{y}; \boldsymbol{\mu}, \boldsymbol{\Sigma}) &= (2\pi)^{-p/2} |\boldsymbol{\Sigma}|^{-1/2} \exp \left\{ -\frac{1}{2} (\mathbf{y} - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1} (\mathbf{y} - \boldsymbol{\mu}) \right\} \\ &= \exp \left\{ \mathbf{y}^\top \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu} - \frac{1}{2} \boldsymbol{\mu}^\top \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu} - \frac{1}{2} \mathbf{y}^\top \boldsymbol{\Sigma}^{-1} \mathbf{y} - \frac{1}{2} \log |2\pi \boldsymbol{\Sigma}| \right\}. \end{aligned}$$

Usando que

$$\mathbf{y}^\top \boldsymbol{\Sigma}^{-1} \mathbf{y} = \text{tr } \boldsymbol{\Sigma}^{-1} \mathbf{y} \mathbf{y}^\top = (\text{vec } \mathbf{y} \mathbf{y}^\top)^\top \text{vec } \boldsymbol{\Sigma}^{-1},$$

obtenemos

$$f(\mathbf{y}; \boldsymbol{\mu}, \boldsymbol{\Sigma}) = \exp \left\{ \mathbf{T}_1^\top \boldsymbol{\eta}_1 + \mathbf{T}_2^\top \boldsymbol{\eta}_2 - \boldsymbol{\eta}_1^\top \boldsymbol{\Sigma} \boldsymbol{\eta}_1 - \frac{1}{2} \log |2\pi \boldsymbol{\Sigma}| \right\}.$$

Es decir, podemos escribir  $\mathbf{Y}$  como un miembro de la familia exponencial, donde

$$\mathbf{T}_1(\mathbf{Y}) = \mathbf{Y}, \quad \mathbf{T}_2(\mathbf{Y}) = \mathbf{Y} \mathbf{Y}^\top, \quad \boldsymbol{\eta}_1(\boldsymbol{\theta}) = \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu}, \quad \boldsymbol{\eta}_2(\boldsymbol{\theta}) = -\text{vec } \boldsymbol{\Sigma}^{-1}.$$