MAT-206: Familia exponencial

Felipe Osorio

fosorios.mat.utfsm.cl

Departamento de Matemática, UTFSM

Definición 1 (Familia exponencial):

La clase de modelos $\{\mathsf{P}_\theta : \theta \in \Theta\}$ 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

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

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

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

Observación:

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

$$b(\eta) = \log \int \exp\big(\eta T(x)\big) h(x)\, \mathrm{d}x,$$

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

$$\mathcal{P} = \{\mathsf{P}_\eta : \eta \in \Gamma\}, \qquad \Gamma := \eta(\Theta), \tag{2}$$

donde el conjunto

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

se denomina espacio paramétrico natural.

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

Ejemplo:

Sea $\mathsf{P}_\theta \stackrel{\mathsf{d}}{=} \mathsf{Poisson}(\theta)$ 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_θ corresponde a una FE (1-paramétrica) donde $q = 1$,

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

Ejemplo:

Sea $\mathsf{P}_\theta \stackrel{\mathsf{d}}{=} \mathsf{Bin}(n,\theta)$, $\theta \in (0,1)$ 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 $\mathsf{Bin}(n,\theta)$ está en FE (1-paramétrica) con $q = 1$,

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

Observación:

Note que las funciones η, b y T no son únicas. En efecto, una manera alternativa de escribir la densidad de la famila 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 Exp$(1,\theta)$, con densidad

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

Esta familia no pertenece a la clase FE.[1]

[1] 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_θ en la FE 1-paramétrica y $\theta \in \Theta$. De este modo,

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

con

$$T^{(n)}(\boldsymbol{x}) = \sum_{i=1}^{n} T(x_i), \qquad b^{(n)}(\theta) = nb(\theta), \qquad \widetilde{h}(\boldsymbol{x}) = \prod_{i=1}^{n} h(x_i). \tag{3}$$

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

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

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

La función generadora de cumulantes está definida como

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

y permite obtener los cumulantes de T,

$$\kappa_r = \frac{\mathsf{d}^r}{\mathsf{d}s^r} K_T(s)\Big|_{s=0}.$$

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

$$K_T' = M_T'/M_T, \qquad K_T'' = [M_T M_T'' - (M_T')^2]/M_T^2,$$

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

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

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

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

Resultado 2:

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

$$\mathsf{E}(T) = b'(\eta), \qquad \mathsf{var}(T) = b''(\eta).$$

Demostración:

Basta notar que

$$\begin{aligned}
\mathsf{E}(T) &= \frac{\mathsf{d}}{\mathsf{d}s} K_T(s)\Big|_{s=0} = b'(s+\eta)\big|_{s=0} = b'(\eta), \\
\mathsf{var}(T) &= \frac{\mathsf{d}^2}{\mathsf{d}s^2} K_T(s)\Big|_{s=0} = b''(s+\eta)\big|_{s=0} = b''(\eta).
\end{aligned}$$

Para una muestra IID, $X_1, \dots, X_n$, y usando el Resultado 2 sigue inmediatamente que:

$$\begin{aligned}
\mathsf{E}(T^{(n)}(\boldsymbol{X})) &= \mathsf{E}\Big(\sum_{i=1}^n T(X_i)\Big) = \sum_{i=1}^n \mathsf{E}(T(X_i)) = nb'(\theta) \\
\mathsf{var}(T^{(n)}(\boldsymbol{X})) &= \mathsf{var}\Big(\sum_{i=1}^n T(X_i)\Big) = \sum_{i=1}^n \mathsf{var}(T(X_i)) = nb''(\theta).
\end{aligned}$$

Familia exponencial

Sea $\mu = \mathsf{E}(Y)$, es decir, tenemos una transformación[2] 1-1 entre θ y el parámetro de media

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

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

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

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

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

[2] suave y estrictamente convexa

Definición 2:

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

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

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

Ejemplo:

Suponga $\mathsf{P}_\theta \stackrel{\mathsf{d}}{=} \mathsf{N}_1(\mu, \sigma^2)$,

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

La densidad de P_θ puede ser escrita como

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

De este modo,

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

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

Luego, $\mathsf{N}_1(\mu, \sigma)$ corresponde a una FE 2-paramétrica.

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.

Considere

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

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

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

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

Observación:

Evidentemente, resulta fácil extender[3] los Resultados 1 y 2. En efecto,

$$\begin{aligned} M_T(\boldsymbol{s}) &= \mathsf{E}[\exp(\boldsymbol{s}^\top \boldsymbol{T})] = \exp[b(\boldsymbol{\eta}+\boldsymbol{s}) - b(\boldsymbol{\eta})], \\ K_T(\boldsymbol{s}) &= b(\boldsymbol{\eta}+\boldsymbol{s}) - b(\boldsymbol{\eta}), \end{aligned}$$

y

$$\mathsf{E}(\boldsymbol{T}(\boldsymbol{X})) = \dot{b}(\boldsymbol{\eta}) = \frac{\partial b(\boldsymbol{\eta})}{\partial \boldsymbol{\eta}}, \qquad \operatorname{Cov}(\boldsymbol{T}(\boldsymbol{X})) = \ddot{b}(\boldsymbol{\eta}) = \frac{\partial^2 b(\boldsymbol{\eta})}{\partial \boldsymbol{\eta} \partial \boldsymbol{\eta}^\top}.$$

[3] Revisar los detalles

Definición 3:

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

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

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 $\boldsymbol{Y} \sim \mathsf{FE}(\boldsymbol{\theta}, \phi)$.[4]

Observación:

Cuando ϕ es desconocido, se suele indicar que el modelo está en la familia de dispersión exponencial (Jørgensen, 1997).[5]

[4] El parámetro de dispersión $\phi > 0$ usualmente es considerado un parámetro molesto.
[5] Jørgensen, B. (1997). *The Theory of Dispersion Models*. Chapman & Hall, London.

Ejemplo:

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

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

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

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

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

Ejemplo:

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

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

Usando que

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

obtenemos

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

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

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