Modelos Estadísticos I	(Fecha: 18/05/2021)

Tarea #7

Estudiante: Roberto Vásquez Martínez — *NUA:* 424662

Antes de comenzar con la solución de los problemas haremos una aclaración sobre que entendemos por la **familia exponencial** en el contexto de los **modelos lineales generalizados**.

Definition 0.1 *Decimos que una distribución con densidad f pertenece a la* ***familia exponencial sobredispersa*** *o a la familia exponencial en el contexto de modelos lineales generalizados si f esta parametrizada por θ y un parámetro de dispersión φ y podemos factorizar la densidad de la siguiente forma*

$$f(y;\theta,\varphi) = \exp\left\{\frac{T(y)S(\theta)-b(\theta)}{a(\varphi)} + c(y,\varphi)\right\}, \tag{1.1}$$

donde $T(y), S(\theta), a(\varphi), b(\theta), c(y,\varphi)$ son funciones conocidas.

Problema 1

Si la variable aleatoria Y tiene distribución Gamma con un parámetro de escala θ, que es el parámetro de interés, y un parámetro de forma conocido φ, entonces la función de densidad es

$$f(y;\theta) = \frac{y^{\varphi-1}\theta^{\varphi}e^{-y\theta}}{\Gamma(\varphi)}.$$

Pruebe que esta distribución pertenece a la familia exponencial y encuentre el parámetro natural. También encuentre $E(Y)$ y $V(Y)$.

(Solución)

Veremos que $Y \sim \Gamma(\theta,\varphi)$ con parámetro de escala $\theta > 0$ y forma $\varphi > 0$, con φ conocido pertenece a la familia exponencial.

Notamos las siguientes igualdades

$$\begin{aligned} y^{\varphi-1} &= \exp\{(\varphi-1)\log y\} \\ \theta^{\varphi} &= \exp\{\varphi\log\theta\} \\ (\Gamma(\varphi))^{-1} &= \exp\{-\log(\Gamma(\varphi))\}. \end{aligned}$$

Por lo que

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

Consideramos

$$\begin{aligned}
T(y) =& y\\
S(\theta) =& \theta\\
b(\theta) =& \varphi \log \theta\\
a(\varphi) =& -1\\
c(y, \varphi) =& (\varphi - 1) \log y - \log(\Gamma(\varphi)),
\end{aligned}$$

y observamos que b solo es función de θ pues φ es conocido.

Por lo tanto

$$f(y; \theta) = \exp\left\{ \frac{T(y)S(\theta) - b(\theta)}{a(\varphi)} + c(y, \varphi) \right\},$$

y de esto concluimos que la distribución con parámetro de forma conocido pertenece a la familia exponencial con parámetro natural $b(\theta) = \varphi \log \theta$.

Por las relaciones vistas en clase se tiene que cuando φ es conocido y S y T son la función identidad obtenemos

$$\mathbb{E}[Y] = b'(\theta) = \frac{\varphi}{\theta},$$

y

$$\begin{aligned}
V[Y] =& b''(\theta)a(\varphi) = \left(-\frac{\varphi}{\theta^2}\right)(-1)\\
=& \frac{\varphi}{\theta^2},
\end{aligned}$$

que es lo que queríamos hallar.

$\square$

Problema 2

Probar las siguientes funciones de densidad pertenecen a la familia exponencial.

Antes de empezar recordaremos la definición de la familia exponencial

(a) Distribución Pareto: $f(y;\theta) = \theta y^{-\theta-1}$.

(Solución)

Notamos las siguientes identidades

$$\begin{aligned}\theta =& \exp\{\log\theta\}\\ y^{-\theta-1} =& \exp\{(-\theta-1)\log y\},\end{aligned}$$

por lo que

$$f(y;\theta) = \exp\{\log\theta + (-\theta-1)\log y\}$$

Por otro lado, recordamos que una distribución pertenece a la familia exponencial si podemos factorizar la densidad g correspondiente de la siguiente forma

$$g(y;\theta,\varphi) = \exp\left\{\frac{T(y)\theta - b(\theta)}{a(\varphi)} + c(y,\varphi)\right\},$$

donde $T(y)$ es una estadística suficiente.

Notamos que

$$f(y;\theta) = \exp\left\{\frac{\log(y)\theta - \log(\theta)}{-1} - \log(y)\right\},$$

luego si

$$\begin{aligned}T(y) =& \log(y)\\ S(\theta) =& \theta\\ a(\varphi) =& -1\\ b(\theta) =& \log(\theta)\\ c(y,\varphi) =& -\log(y),\end{aligned}$$

entonces

$$f(y,\theta) = \exp\left\{\frac{T(y)S(\theta) - b(\theta)}{a(\varphi)} + c(y,\varphi)\right\},$$

en este caso el parámetro de dispersión $\varphi = 1$, se tiene así que la distribución Pareto con densidad f pertenece a la familia exponencial.

(b) Distribución exponencial: $f(y;\theta) = \theta e^{-y\theta}$.

(Solución)

Notamos que

$$f(y;\theta) = \exp\left\{\frac{y\theta - \log\theta}{-1}\right\},$$

luego si $T(y) = y$, $S(\theta)$, $a(\varphi) = -1$, $b(\theta) = \log(\theta)$ y $c(y,\varphi) = 0$ entonces

$$f(y;\theta) = \exp\left\{\frac{T(y)S(\theta) - b(\theta)}{a(\varphi)} + c(y,\varphi)\right\},$$

por lo que la distribución exponencial pertenece a la familia exponencial.

(c) Binomial Negativa. $f(y;\theta) = \binom{y+r-1}{r-1}\theta^r(1-\theta)^y$ con r conocido.

(Solución)

Observamos que

$$\begin{aligned}
\binom{y+r-1}{r-1} &= \exp\left\{\log\binom{y+r-1}{r-1}\right\} \\
\theta^r &= \exp\{r\log\theta\} \\
(1-\theta)^y &= \exp\{y\log(1-\theta)\},
\end{aligned}$$

luego

$$f(y;\theta) = \exp\left\{y\log(1-\theta) + r\log(\theta) + \log\binom{y+r-1}{r-1}\right\}.$$

Como r es conocido consideramos

$$\begin{aligned}
T(y) &= y \\
S(\theta) &= \log(1-\theta) \\
b(\theta) &= -r\log(\theta) \\
a(\varphi) &= 1 \\
c(y,\varphi) &= \log\binom{y+r-1}{r-1}.
\end{aligned}$$

Por lo tanto

$$f(y;\theta) = \exp\left\{\frac{T(y)S(\theta) - b(\theta)}{a(\varphi)} + c(y,\varphi)\right\},$$

por lo que la distribución binonomial negativa con r conocido pertenece a la familia exponencial.

□

Problema 3

Para la distribución Pareto encuentre la estadística score U y la información $\mathcal{I} = V(U)$. Verifique que $\mathbb{E}(U) = 0$.

(Solución)

Notamos que la estadística score U queda definida de la siguiente forma

$$U(\theta) = \frac{\partial l}{\partial \theta},$$

donde l es la log-verosimilitud.

En el caso de la distribución Pareto, por el inciso a) del problema anterior la log-verosimilud es

$$l(\theta) = \log f(y;\theta) = \log\theta - \theta\log y - \log y,$$

derivando respecto a θ se tiene que

$$U(\theta) = \frac{1}{\theta} - \log y.$$

Sea Y una variable aleatoria Pareto con densidad la especificada en el inciso a) de el Problema 2, sea $Z := \log Y$, entonces por el Teorema de Cambio de Variable la densidad de Z viene dada por

$$\begin{aligned} g(z;\theta) =& f(e^z;\theta)e^z \\ =& \theta e^{-z(\theta+1)}e^z \\ =& \theta e^{-\theta z}, \end{aligned}$$

luego Z se distribuye exponencial con media $1/\theta$, se sigue de esto que

$$\mathbb{E}[\log Y] = \frac{1}{\theta},$$

por lo que

$$\mathbb{E}[U] = \frac{1}{\theta} - \mathbb{E}[\log Y] = 0.$$

Finalmente hallaremos $\mathcal{I} = V(U)$, como $\mathbb{E}[U] = 0$ entonces

$$\mathcal{I} = V(U) = \mathbb{E}[U^2] = \mathbb{E}\left[\left(\frac{1}{\theta} - \log Y\right)^2\right]$$

Simplificando la expresión anterior tenemos

$$\begin{aligned} \mathcal{I} =& \frac{1}{\theta^2} - 2\cdot\frac{1}{\theta}\cdot\mathbb{E}[\log Y] + \mathbb{E}[(\log Y)^2] \\ =& \mathbb{E}[(\log Y)^2] - \frac{1}{\theta^2} \\ =& \mathbb{E}[(\log Y)^2] - (\mathbb{E}[\log Y])^2 \\ =& V[\log Y]. \end{aligned}$$

Como $\log Y$ se distribuye exponencial con media $1/\theta$ entonces $V[\log Y] = \frac{1}{\theta^2}$, de lo que concluimos

$$\mathcal{I} = \frac{1}{\theta^2},$$

que es lo que queríamos hallar.

$\square$

Problema 4

(a)

(Solución)
Notamos que $Y_i \sim \text{Ber}(\pi_i)$ con soporte $\{0,1\}$ si f_i es su función de probabilidades entonces

$$\begin{aligned} f_i(y_i;\pi_i) =& \pi_i^{y_i}(1-\pi_i)^{1-y_i} \\ =& \exp\{y_i \log \pi_i + (1-y_i)\log(1-\pi_i)\} \\ =& \exp\left\{ y_i \log\left(\frac{\pi_i}{1-\pi_i}\right) + \log(1-\pi_i)\right\}, \end{aligned}$$

consideramos

$$\begin{aligned} T(y_i) =& y_i \\ S(\pi_i) =& \log\left(\frac{\pi_i}{1-\pi_i}\right) \\ a(\varphi) =& 1 \\ b(\pi_i) =& -\log(1-\pi_i) \\ c(y_i,\varphi) =& 0. \end{aligned}$$

Por lo tanto

$$f_i(y_i;\pi) = \exp\left\{ \frac{T(y_i)S(\theta) - b(\theta)}{a(y_i,\varphi)} + c(y_i,\varphi)\right\},$$

por lo que la distribución de Y_i pertenece a la familia exponencial, en este caso el parámetro de dispersión es $\varphi = 1$.

(b)

(Solución)

El parámetro natural bajo la parametrización de la familia exponencial que estamos considerando es $S(\pi_i)$, por lo que el parámetro natural tiene la forma

$$S(\pi_i) = \log\left(\frac{\pi_i}{1-\pi_i}\right).$$

(c)

(Solución)

Al tener $Y_i \sim \text{Ber}(\pi_i)$ con soporte $\{0,1\}$ tenemos que

$$\mathbb{E}[Y_i] = 1\cdot\pi_i + 0\cdot(1-\pi_i) = \pi_i,$$

que es lo que queríamos ver.

(d)

(Solución)

Estamos considerando la liga canónica

$$g(\pi) = S(\pi) = x'\beta.$$

Notamos que si

$$h(t) = \frac{e^t}{1+e^t},$$

entonces

$$\begin{aligned} h(g(\pi)) =& \frac{\exp\left\{\log\left(\frac{\pi}{1-\pi}\right)\right\}}{1+\exp\left\{\log\left(\frac{\pi}{1-\pi}\right)\right\}} \\ =& \frac{\frac{\pi}{1-\pi}}{1+\frac{\pi}{1-\pi}} \\ =& \frac{\frac{\pi}{1-\pi}}{\frac{1}{1-\pi}} \\ =& \pi, \end{aligned}$$

luego $h = g^{-1}$.
Por lo tanto

$$\pi = h(g(\pi)) = h(x'\beta) = \frac{\exp\{x'\beta\}}{1+\exp\{x'\beta\}},$$

que es lo que queríamos ver.

$\square$

Problema 5

(Solución)

(a) Notamos que si

$$\mu_i = \mathbb{E}[Y_i] = (\beta_0 + \beta_1 x_i)^2$$

entonces la función liga cumple que g

$$g^{-1}(\mu_i) = (\beta_0 + \beta_1 x_i)^2,$$

por lo que $g(t) = \sqrt{t}$, luego esta es candidata a ser función liga por se monótona y diferenciable.
Sin embargo, pese a que g podría ser candidata a función liga y la distribución Pareto pertenece a la familia exponencial esta factorización como hemos visto no es canónica, es decir, no se cumple que $T(y) = y$, por lo que en este caso no cumple para tener las expresiones en términos de $b(\theta)$ para $E[Y]$ y $V[Y]$, de lo que concluimos que este modelo no es apto para considerarlo un modelo lineal generalizado.

(b)

Este modelo no puede ser lineal generalizado. Sabemos

$$\mu_i = \beta_0 + \log(\beta_1 + \beta_2 z_i),$$

luego

$$\begin{aligned}\mu_i =& \log(e^{\beta_0}) + \log(\beta_1 + \beta_2 x_i)\\ =& \log(e^{\beta_0}\beta_1 + e^{\beta_0}\beta_2 x_i) \qquad \forall i \in \{1, 2, \dots, n\}.\end{aligned}$$

Por lo que si $g : \mathbb{R} \to \mathbb{R}$ es tal que $g(t) = e^t$ entonces

$$g(\mu_i)] = e^{\beta_0}\beta_1 + e^{\beta_0}\beta_2 x_i,$$

además la función $g : \mathbb{R} \to \mathbb{R}$ es monótona y diferenciable, pero no es lineal en los parámetros, por lo que g no puede ser función liga en consecuencia el modelo no puede ser lineal generalizado.

$\square$