Profesor: Héctor Bahamonde, PhD.

e:hector.bahamonde@uoh.cl

w:www.hectorbahamonde.com

Curso: MLE.

TA: Gonzalo Barría.

I. Likelihood e Inferencia

Recuerda que Equation 1 es el "axiom likelihood". Lo que dice que likelihood es "proporcional" a la probabilidad (King 1998, p. 59). Y donde también la constante $k(y)$ asegura que el likelihood es relativo al modelo, y no un absoluto como ocurre en el paradigma de la probabilidad.

$$\begin{aligned} L(\tilde{\theta}|y) &= k(y)Pr(y|\tilde{\theta}) \\ &\propto Pr(y|\tilde{\theta}) \end{aligned} \tag{1}$$

Ahora, tratemos de estimar nuestro parámetro β pero usando MLE. Es decir, en vez de estimar $\beta = (x^\top x)^{-1}x^\top y$, lo estimaremos usando la Equation 1 de la siguiente manera:

$$\begin{aligned} L(\tilde{\beta}|y) &= k(y)Pr(y|\tilde{\beta}) \\ &= k(y)\prod_{i=1}^{n} f(y_i|\tilde{\beta}) \\ &\propto \prod_{i=1}^{n} f(y_i|\tilde{\beta}) \\ &= \prod_{i=1}^{n}(2\pi)^{-\frac{1}{2}} \exp\left[\frac{-(y_i-\beta)^2}{2}\right] \end{aligned} \tag{2}$$

Importantemente, Equation 2 nos entrega el likelihood relativo de que **el modelo (capturado por β) produzca los datos** y_i. Sobre esto, hay que rescatar los siguientes puntos:

1. MLE es inverso a la probabilidad: en MLE el modelo es construido, y los datos son "dados".

2. Para hacer obtener el likelihood (que es simplemente un número), trabajaremos con el conocido "log-likelihood function", que es básicamente sacar el log de Equation 2 (tal como lo demuestra

Equation 3).

3. Usamos *product operators* ($\prod$) porque necesitamos multiplicar el likelihood de cada i particular (cada fila en la columna y). Esto es una consecuencia de que por ejemplo i_1 es independiente a i_2 y asi para todos los i_n.

Veamos ahora el "log-likelihood function":

$$\begin{aligned} L(\tilde{\beta}|y) &= k(y)Pr(y|\tilde{\beta}) \\ ln\ L(\tilde{\beta}|y) &= ln\left\{k(y)Pr(y|\tilde{\beta})\right\} \\ ln\ L(\tilde{\beta}|y) &= -\frac{1}{2}\sum_{i=1}^{n}(y_i - \tilde{\beta})^2 \end{aligned} \tag{3}$$

Debido a que el logaritmo de un producto ($k(y)Pr(y|\tilde{\beta})$), y usando el *Fisher-Neyman Factorization Lemma*, la tercera línea es la forma simplificada del "log-likelihood function". Nota que $ln\ L(\tilde{\beta}|y)$ depende de los errores cuadrados, es decir, la distancia entre lo predicho y lo observado, i.e. $(y_i - \tilde{\beta})^2 = \epsilon_i^2$.