Profesor: Héctor Bahamonde, PhD.
e:hector.bahamonde@uoh.cl
w:www.hectorbahamonde.com
Curso: MLE.
TA: Gonzalo Barría.

Probabilidad y Likelihood

Usualmente, nosotros estimamos la relación entre x e y vía un modelo lineal OLS (Equation 1).

$$y_i = \alpha + \beta_1 x_1 + \epsilon_i \tag{1}$$

En este setup, siempre hablamos de "probabilidad". Es claro que Equation 1 nos muestra que los valores de y_i dependen de $\boldsymbol{x}$ (que es una matrix que contiene α y todos los x's), más el residuo ϵ. Cada parámetro tiene un statement de probabilidad (recuerden el *p-value*).

Es decir, la probabilidad de y depende de $\boldsymbol{x}$, lo que significa que en este paradigma, **y depende de un modelo** (M). Recuerda que cada vez que pones una x a la mano derecha de la igualdad, estás haciendo un modelo (estás especificando que la realidad depende de ciertas variables independientes). Más formalmente, esta noción se demuestra en Equation 2:

$$Pr(y|M) = Pr(\text{datos}|\text{modelo}) \tag{2}$$

donde en Equation 2 M es asumido "fijo" o "conocido" ("fixed") y los datos "random". *Lo son?* Recuerda que incluso existe la noción del *true model* (un modelo verdadero que existe "por ahí", pero que es inaccesible para nosotros). King (1998, p. 16) explica que cuando "M is treated as given is a problem because uncertainty in inference lies with the model, not the data."

- **M: no puede ser tratado como conocido**. Los modelos se construyen. Nadie viene a darnos el modelo correcto.
- **datos**: no pueden ser *random*. La naturaleza ya ha producido los datos. Es más, y por esta misma razón, los datos debieran ser considerados como "dados" o "conocidos". No?

Qué diferencias existen entre los conceptos de *probabilidad* y *likelihood*?

Probabilidad: Incertidumbre Absoluta

Las probabilidades, todas, "viven" entre el 0 y 1.

$$\begin{aligned} Y &\sim f(y|\theta, \alpha) \\ \boldsymbol{\theta} &= g(\boldsymbol{X}, \boldsymbol{\beta}) \end{aligned} \tag{3}$$

donde la primera linea see lee "y se distribuye como una función de y que depende de $\boldsymbol{\theta}$ (que es $\hat{y}$, o el valor esperado de y_i, y α". El valor α se puede entender como parametro que está presente sólo en algunos modelos GLMs, o está dado por otros factores. Por ejemplo, σ^2 (la varianza) está dada, no se estima (pero se asume que es **constante**). La segunda línea es una implicación de la primera, y se lee asi: "donde $\boldsymbol{\theta}$ es igual a una funcion g, donde metemos los datos observados $\boldsymbol{X}$ y calculamos los parametros $\boldsymbol{\beta}$".

Lo unico que estamos haciendo, es hacer la transición a la notación de MLE. En este caso, estamos viendo dónde "viven" los parámetros y los datos observados y predichos, pero en lenguaje MLE. Si te fijas, $\boldsymbol{\theta}$ (la caja que contiene $\boldsymbol{X}$ y $\boldsymbol{\beta}$) es tomado como "true parameter". Si el modelo es tomado como "verdadero", qué implicancias tiene el estimado β? Piensa en los intervalos de confianza.

Likelihood: Incertidumbre Relativa

En lenguaje de likelihood (que se expresa con el carácter griego $\boldsymbol{\theta}$) se refiere a un parámetro θ_i de los infinitos parámetros $\boldsymbol{\theta}$ entre todos los modelos posibles ($\tilde{\boldsymbol{\theta}}$). Recuerda, en Equation 3 $\boldsymbol{\beta}$ se refiere a los "true values". Eso quiere decir que OLS y MLE tienen llegadas distintas: en MLE existen tantos $\boldsymbol{\theta}$ como hay datos. Esto implica que la incertidumbre en MLE está anclada a los datos (no al modelo). En otras palabras, en likelihood la incertidumbre es *relativa* porque el likelihood θ no depende del modelo ($\tilde{\boldsymbol{\theta}}$) sino que de los datos. Es por esto que en likelihood uno piensa en términos de qué modelo ($\tilde{\theta}$) **maximiza** el likelihood de haber creado los datos y. **A diferencia del lenguaje de probabilidad, los datos no son tomados como "dados".**

$$\begin{aligned} L(\tilde{\theta}|y) &= k(y)Pr(y|\tilde{\theta}) \\ &\propto Pr(y|\tilde{\theta}) \end{aligned} \tag{4}$$

donde $k(y)$ es una función constante positiva desconocida que depende de y (y cambia cada vez que y cambia). Si te fijas, en probabilidad (Equation 2) la probabilidad es $Pr(y|M)$, pero en likelihood (Equation 4) la probabilidad de $Pr(y|\tilde{\theta})$ es "proporcional" (i.e. $\propto$) a $L(\tilde{\theta}|y)$. Fíjate cómo se invierten los términos.

Esto implica que los likelihoods de dos modelos pueden ser comparados, pero sólo si bienen de los mismos datos (recuerda que $k(y)$ depende de y). Al contrario, en probabilidad, la probabilidad de que algo pase, *siempre* va entre 0 y 1 (independiente de los datos o los modelos usados). En cambio, los likelihoods pueden tomar cualquier número y no son necesariamente comparables (son como los errores estándard en un modelo lineal—sólo toman significancia dentro de ese modelo en particular).

References

King, Gary (1998). *Unifying Political Methodology: The Likelihood Theory of Statistical Inference*. Ann Arbor, MI: University of Michigan Press, pp. 1–274.