Profesor: Héctor Bahamonde, PhD.

e:hector.bahamonde@uoh.cl

w:www.hectorbahamonde.com

Curso: MLE.

TA: Gonzalo Barría.

I. Propiedades de los Estimadores MLE

1. Invariance to reparameterization: estimaciones se pueden transformar y no alteran su valor substantivo. Por ej., tomar un log o elevar al cuadrado.

2. Invariance to sampling plans: debido a que los estimadores dependen de los datos usados (son particulares a ellos, y no son generales), Es por esto que variando el tamaño de la muestra no varía la calidad de los estimadores.

3. $\hat{\theta}$ no está sesgado. Es decir, en la medida que se repitan los experimentos, $\hat{\theta}$ no cambia.

4. $\hat{\theta}$ es consistente. Dentro de todo el "parameter space", existe un "spike" (una punta) que representa el verdadero parametro $\hat{\theta}$.

5. El "parameter space" de $\hat{\theta}$ esta normalmente distribuido.

II. Precisión de los Estimadores MLE

Pensar en la presición, es pensar en cuán bien (o mal) nuestro estimador $\hat{\theta}$ (i.e. "nuestro β") maximice el likelihood the que $E(y) = \pi$. Para esto, debemos recordar que la función del LL es relativa (depende de los datos). No es como antes en OLS, donde podíamos comparar *p-values*, r^2 y otros *entre modelos con distintos datasets*. Aquí no.

Likelihodd Ratio Test (LLR)

$$\begin{aligned} \Re &= 2\ln(\frac{L_{R_\star}}{L_\star}) \\ &= 2(\ln L_\star - L_{R_\star}) \end{aligned} \tag{1}$$

donde $L_\star$ es el likelihood del modelo "entero" y $L_{R_\star}$ es el likelihood del modelo "restringido". "Entero" significa que tiene todos los efectos estimados, mientras que el "restringido" tiene los efectos *restringidos* a 0.

Volvamos a usar el lenguaje OLS para ver qué significa "entero" y "restringido". Supongamos que tenemos dos modelos compitiendo; **ambos se construyeron usando la misma base de datos** (entonces, los podemos comparar).

$$y_i = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \epsilon_i \tag{2}$$

$$\begin{aligned} y_i &= \beta_0 + 0x_1 + \beta_2 x_2 + \epsilon_i \\ y_i &= \beta_0 + 0 + \beta_2 x_2 + \epsilon_i \end{aligned} \tag{3}$$

Aquí lo que hemos hecho es que el modelo en Equation 2 está "entero". Sin embargo, el modelo en la Equation 3 está "restringido"; hemos "restringido" el valor de $\beta_1 = 0$. Nota que $0 \times x_1 = 0$. La entonces pregunta que nos ayuda a resolver el LLR es *qué modelo es mejor*? *El restringido o el entero?* Comparemos sus respectivos LLRs.

Si $L_\star = L_{R_\star}$ entonces $\Re = 1$. Pero si el likelihood del modelo "entero" $L_\star$ es mejor, entonces $\Re \geq 1$. Por ejemplo, si $L_\star = -72.86$ y $L_{R_\star} = -81.27$, entonces,

$$\begin{aligned} \Re &= 2\ln(\frac{L_{R_\star}}{L_\star}) \\ &= 2(\ln -72.86 - -81.27) \\ &= 16.8 \end{aligned} \tag{4}$$

Aqui podemos concluir que el modelo "entero" es mejor, y que debemos quedarnos con este.

Wald Test El LRT funciona cuando mejor cuando tenemos un número limitado de hipótesis. En el ejemplo anterior contrastamos la hipotesis donde Equation 2 era mejor que Equation 3 (y concluimos que Equation 2 era mejor). *Qué pasa cuando tienes muchas hipótesis y resulta complicado restringir varios parametros a cero?* Para eso usamos el *Wald test*. La ventaja de este test es que "only the unrestricted model need be estimated; from that model, one can perform tests against any number of alternatives" (King 1998, p. 87).

El Wald test es un tipo de t-test. De nuestro conocimiento de OLS sabemos que una vez obtenido el t-test, es posible calcular el p-value.

Recordemos primero el t-test:

$$t = \frac{\hat{\mu}_{\text{s}} - \tilde{\mu}_{\text{p}}}{\frac{\hat{\sigma}}{\sqrt{n}}} \tag{5}$$

donde $\hat{\mu}_{\text{s}}$ es el *sample mean* (lo que observamos), $\tilde{\mu}_{\text{p}}$ es el *population mean* (lo que no observamos), y $\hat{\sigma}$ es la desviación estándar estimada. Ahora veamos el Wald test,

$$\mathbb{W} = \frac{\hat{\beta} - \tilde{\beta}}{\hat{\sigma}} \tag{6}$$

donde, de manera análoga al t-test en Equation 5, $\hat{\beta}$ es el valor promedio estimado (lo que observamos), y a diferencia del t-test, $\tilde{\beta}$ es un posible valor de β, mientras que $\hat{\sigma}$ es el error estándar de $\hat{\beta}$.

Veamos el siguiente ejemplo, donde de nuevo trataremos de testear la diferencia entre nuestro modelo "entero" (Equation 2) y nuestro modelo "restringido" (Equation 3). Supongamos que nuestro modelo OLS en la Equation 2 $\hat{\beta}_1 = -2.30$ (con error estandár estimado de $\hat{\sigma} = 0.5$), y que nosotros queremos testear si $\hat{\beta}_1 = 0$, o más precisamente, que uno de los posibles valores de $\hat{\beta}_1$ es $\tilde{\beta}_1 = 0$. Para eso, restringimos nuestro modelo (como en Equation 3).

$$\begin{aligned}
\mathbb{W} &= \frac{\hat{\beta}_1 - \tilde{\beta}_1}{\hat{\sigma}} \\
\mathbb{W} &= \frac{-2.30 - 0}{0.5} \\
\mathbb{W} &= -4.60
\end{aligned} \tag{7}$$

Dado que $\mathbb{W} = -4.60$, podemos sugerir que el modelo restringido (Equation 3) esta -4.60 **desviaciones estándar lejos** del modelo entero (Equation 2)—es decir, ambos son substancialmente diferentes. Si realmente $\hat{\beta}_1 = 0$, entonces $\tilde{\beta}_1 = 0$, lo que hubiera implicado que $\mathbb{W} = 0$, pero en este caso es -4.60. Esto significa que debemos rechazar la idea de que $\hat{\beta}_1 = 0$ (rechazando así el modelo restringido, Equation 3), y quedándose con el modelo entero (Equation 2) donde $\hat{\beta}_1 = -2.30$.

- La ventaja del Wald test es que podemos testear si múltiples parámetros son iguales a cero (u otro valor).

- Debiera ser claro en este momento que ultimamente, lo que busca este test, es sugerir si un parámetro aporta al modelo (mejorando su likelihood) o no.

References

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