Teoría Econométrica I - EAE- 250-A

Técnicas de Remuestreo - Bootstrap

Tatiana Rosá

Instituto de Economía - Pontificia Universidad Católica de Chile

Noviembre 2021

Introducción

- En esta sección veremos algunas técnicas de remuestreo que nos proveen un método alternativo de inferencia teoría asintótica
  - Ni inferencia exacta basada en supuestos distribucionales
  - Ni inferencia asintótica (implica muestras grandes)
  - Bootstrap no paramétrico (Efron, 1979) y algunas variantes paramétricas
- Vamos a seguir las notas de Hansen ((https://www.ssc.wisc.edu/ bhansen/econometrics/Econometrics.pdf))

Bootstrap

- Sea F la distribución conjunta de las observaciones (y_i, x_i)
- Sea un estadístico de interés

$$T_n = T_n((y_1, x_1), ..., (y_n, x_n))$$

- Por ejemplo puede ser un estimador θ o un test-t $\left(\frac{\hat{\theta}-\theta}{SE(\theta)}\right)$
- El estadístico depende de una u otra manera de F
- La cdf exacta de T_n cuando los datos son "sampleados" de la distribución F es:

$$G_n(u, F) = Pr(T_n \leq u|F)$$

- En general $G_n(u, F)$ depende de F, lo que implica que G cambia cuando cambia F
- Idealmente nos gustaría poder hacer inferencia basada en $G_n(u, F)$. Esto es generalmente imposible, puesto que F es desconocido.

Bootstrap

- Las aproximaciones asintóticas antes vista se basan en aproximar $G_n(u, F)$ con

$$G_n(u, F) = \lim_{n \to \infty} G_n(u, F)$$

- Cuando $G(u, F) = G(u)$ no depende de la F usamos la distribución $G(u)$ para hacer inferencia.
- Efron (1979) propone el **bootstrap** que hace una aproximación distinta. La cdf desconocida, F, es reemplazada por un estimador consistente F_n. Si reemplazamos F_n por F, obtenemos

$$G_n^*(u) = G_n(u, F_n)$$

Bootstrap

- Llamemos a $G_n^*(u)$ la "distribución bootstrap". La inferencia basada en $G_n^*(u)$ es inferencia alternativa a la teoría asintótica.
- Sea (y_i^*, x_i^*) una variable aleatoria con distribución F_n. Una muestra aleatoria con dicha distribución se llama **"Bootstrap Data"**.
- El estadístico $T_n^* = T_n((y_1^*, x_1^*), ..., (y_n^*, x_n^*), F_n)$ construido con esta "Bootstrap Data" es una variable aleatoria con distribución $G_n^*(u)$.
- Llamemos a T_n^* un estadístico de Bootstrap

Función de distribución empírica

- Recuerde que

$$\begin{aligned} F(y,x) &= Pr(y_i \leq y, x_i \leq x) \\ &= \mathbb{E}(1_{\{y_i \leq y\}}, 1_{\{x_i \leq x\}}) \end{aligned}$$

  Donde $1_{\cdot}$ es una función indicadora
- El método de los momentos nos dice que igualemos a los cocientes muestrales, así:

$$F_n(y,x) = \frac{1}{n}\sum_{i=1}^{n} 1_{\{y_i \leq y\}} 1_{\{x_i \leq x\}}$$

  $F_n(y,x)$ es la **función de distribución Empírica (EDF)**
- F_n es un estimador no paramétrico de F. Note que F puede ser discreta o continua, pero F_n es una step function.

Función de distribución empírica

- La gracia es que la EDF es un estimador consistente de la CDF
- Note que
  $\forall(y, x)$; $1_{\{y_i \leq y\}} \cdot 1_{\{x_i \leq x\}}$ es una variable aleatoria i.i.d, cuya esperanza es $F(y, x)$, luego por LLN, tenemos que:

$$F_n(y, x) \xrightarrow{p} F(y, x)$$

- Es más, la naturaleza Bernoulli de $1_{\{y_i \leq y\}} \cdot 1_{\{x_i \leq x\}}$ nos dice que:

$$Var(1_{\{y_i \leq y\}} \cdot 1_{\{x_i \leq x\}}) = F(y, x)(1 - F(y, x))$$

  $\Rightarrow$ por CLT L-L,

$$\sqrt{n}(F_n(y, x) - F(y, x)) \xrightarrow{d} \mathcal{N}(0, F(y, x)(1 - F(y, x)))$$

- **Importante:** la EDF es una distribución de probabilidad discreta y válida, que pone igual probabilidad $\frac{1}{n}$ para cada par (y_i, x_i), $i = 1, ..., n$

Bootstrap no paramétrico

- El **bootstrap no paramétrico** se obtiene cuando la distribución $G_n^*(u) = G_n(u, F_n)$ se define usando la EDF: $F_n(y, x) = \frac{1}{n}\sum_{i=1}^{n} 1_{\{y_i \leq y\}} \cdot 1_{\{x_i \leq x\}}$, como un estimador de F
- Dada la EDF, F_n, es en esencia una distribución multinomial (con n puntos de soporte). Note que un trial de una multinomial es lo mismo que un trial de una binomial o bernoulli. Luego, para un par (y_i^*, x_i^*) teníamos una bernoulli, para n pares tenemos una multinomial
- En principio la distribución $G_n^*(u)$ se podría calcular por métodos directos (método de la transformada, es decir, una transformación conocida de variables aleatorias con distribución conocida) lo cual es en general implausible
- Una alternativa es generar muestras a partir de la EDF (que es un estimador consistente de F) y obtener así la distribución $G_n^*(u)$. Sin embargo, dado que hay $(2n-1)!/(n!(n-1)!)$ muestras posibles $\{(y_1^*, x_1^*), ..., (y_n^*, x_n^*)\}$, ese cálculo es no factible

Bootstrap no paramétrico

- Luego, en lugar de calcular $(2n-1)!/(n!(n-1)!)$ realizaciones de la EDF lo haremos para un número grande pero acotado. Así, el método puede parecer conceptualmente complejo pero es sencillo de implementar:

1. Genere B muestras de bootstrap de tamaño n (esto equivale a muestrear la EDF que pone la misma masa de probabilidad a cada par, $1/n$. Dado eso, esto equivale a hacer una muestra aleatoria simple con reemplazo para generar cada una de las B muestras).
2. Calcule el estadístico de bootstrap para cada una de las B muestras.

$$T_n^* = T_n((y_1^*, x_1^*), ..., (y_n^*, x_n^*), F_n)$$

   Andrews y Budinsky (2000) sugieren como calcular B, pero típicamente $B = 1000$ es suficiente.
3. Obtenga el sesgo, varianza e intervalos de confianza para su estadístico T_n a partir de los B valores de su estadístico de bootstrap T_n^*.

Sesgo

- **Sesgo:** Suponga que tenemos un estimador $\hat{\theta}$, el sesgo de $\hat{\theta}$ lo expresamos:

$$\tau_n = \mathbb{E}(\hat{\theta} - \theta_0) \quad \text{(sesgo)}$$

- Sea $T_n(\theta) = \hat{\theta} - \theta$, luego $\tau_n = \mathbb{E}(T_n(\theta_0))$.
- La contraparte del bootstrap es:

$$\hat{\theta}^* = \hat{\theta}((y_1^*, x_1^*), ..., (y_n^*, x_n^*)) \quad \text{y}$$
$$T_n^* = \hat{\theta}^* - \hat{\theta}$$

- El estimador del sesgo via bootstrap es $\tau_n^* = \mathbb{E}(T_n^*)$. Así,

$$\hat{\tau}_n^* = \frac{1}{B}\sum_{b=1}^{B} T_{nb}^*$$

$$\hat{\tau}_n^* = \frac{1}{B}\sum_{b=1}^{B}(\hat{\theta}_b^* - \hat{\theta}) = \bar{\hat{\theta}}^* - \hat{\theta}$$

Biased corrected

- Luego si $\hat{\theta}$ es sesgado podemos querer calcular un estimador **biased-corrected**. Idealmente

$$\tilde{\theta} = \hat{\theta} - \tau_n$$

- Pero τ_n es desconocido. Luego, el estimador biased-corrected mediante bootstrap es:

$$\begin{aligned}\tilde{\theta}^* &= \hat{\theta} - \hat{\tau}_n^* \\ &= \hat{\theta} - (\bar{\hat{\theta}}^* - \hat{\theta}) \\ &= 2\hat{\theta} - \bar{\hat{\theta}}^*\end{aligned}$$

Varianza

- **Varianza:** Sea $T_n = \hat{\theta}$, la varianza de $\hat{\theta}$ es:

$$V_n = \mathbb{E}(T_n - \mathbb{E}(T_n))^2$$

- Sea $T_n^* = \hat{\theta}^*$, este tiene varianza:

$$V_n^* = \mathbb{E}(T_n^* - \mathbb{E}(T_n^*))^2$$

- La estimación mediante bootstrap es:

$$\hat{V}_n^* = \frac{1}{B}\sum_{b=1}^{B}(\hat{\theta}_b^* - \bar{\hat{\theta}}^*)^2$$

  $\therefore$ Un error estándar bootstrap para $\hat{\theta}$ es:

$$SE^*(\hat{\theta}) = \sqrt{\hat{V}_n^*}$$

- Si bien este error estándar puede ser calculado no siempre es muy útil. Generalmente nos interesa construir intervalo de confianza mediante bootstrap (no tenemos que asumir normalidad)

IC con bootstrap - método del percentil

- Para una distribución $G_n(u, F)$ sea $q_n(\alpha, F)$ un cuantil, la función que satisface

$$G_n(q_n(\alpha, F), F) = \alpha$$

- Si tenemos contrapartes de bootstrap $q_n^*(\alpha, F) = q_n^*(\alpha)$, a un $(1-\alpha)\%$, el **intervalo de confianza de EFRON** para $T_n = \hat{\theta}$ es

$$C_1 = [q_n^*(\alpha/2), q_n^*(1-\alpha/2)]$$

- Este se conoce como el **método del percentil** y es muy popular en el trabajo empírico a pesar que no está muy bien motivado, pero tiene la ventaja de ser invariante a transformaciones monotónicas.

Método del percentil - pasos

- Para hacer el intervalo de confianza, el método recomendado por Efron (1979) es el siguiente:
  1. Estime $q_n(\alpha, F)$ mediante bootstrap, obteniendo
  $$q_n^*(\alpha, F_n) = q_n^*(\alpha)$$
  2. Si $T_n = \hat{\theta}$, se debe construir el intervalo de confianza en base a la empírica
  $$C_1 = [q_n^*(\alpha/2); q_n^*(1-\alpha/2)]$$
- Este es el **método del percentil**, con un intervalo de confianza al $(1-\alpha)\%$. Este método es tan simple e intuitivo que se puede entender simplemente como “cortar las colas” de la distribución.

Intuición

Motivación alternativa

- A pesar de lo intuitivo del método, no tiene una motivación estadística potente, ni un punto de conexión con la teoría tradicional de intervalos de confianza
- Una manera más parecida a la que conocemos:
- Sea $T_n = \hat{\theta} - \theta$ y $q_n(\alpha)$ el cuantil α de T_n (luego es el cuantil anterior pero desplazado hacia la izquierda en θ). Estime $q_n^*(\alpha)$ por bootstrap y construya el intervalo de confianza de la siguiente manera:

$$C_1 = \left[\hat{\theta} + q_n^*(\alpha/2); \hat{\theta} + q_n^*(1-\alpha/2)\right]$$

- $T_n^* = \hat{\theta^*} - \hat{\theta}$, lo que explica por qué se suma $\hat{\theta}$, para compensar lo que había sido desplazado
- Esto sigue siendo el método del percentil, pero con una motivación distinta

Motivación alternativa

- C_1 es la contraparte bootstrap del intervalo teórico

$$C_1^0 = \left[\hat{\theta} + q_n(\alpha/2); \hat{\theta} + q_n(1-\alpha/2)\right]$$

- La probabilidad de cobertura será

$$\begin{aligned}
\Pr(\theta_0 \in C_1^0) &= \Pr[\hat{\theta} + q_n(\alpha/2) \leq \theta_0 \leq \hat{\theta} + q_n(1-\alpha/2)] \\
&= \Pr[-q_n(1-\alpha/2) \leq \hat{\theta} - \theta_0 \leq -q_n(\alpha/2)] \\
&= G_n(-q_n(\alpha/2), F_0) - G_n(-q_n(1-\alpha/2), F_0)
\end{aligned}$$

- ¿Es esto igual a $1-\alpha$? Por lo general, no lo será. Sólo lo será cuando G_n es simétrica.

Motivación alternativa

- Si G_n es simétrica entonces

$$G_n(-u, F) = 1 - G_n(u, F) \quad \forall u$$

- Si esto se cumple, entonces

$$\Pr(\theta_0 \in C_1^0) = (1 - G_n(q_n(\alpha/2), F_0)) - (1 - G_n(q_n(1 - \alpha/2)))$$

Teníamos que el punto de soporte era $q_n(\alpha, F)$ para un cuantil α, o sea

$$G_n(q_n(\alpha, F), F) = \alpha$$

- Aplicando eso, tenemos

$$\begin{aligned} \Pr(\theta_0 \in C_1^0) &= (1 - \alpha/2) - (1 - (1 - \alpha/2)) \\ &= 1 - \alpha/2 - \alpha/2 \\ &= 1 - \alpha \end{aligned}$$

Sólo funciona con G_n simétrica

- Este es el primer método que sugiere Efron en 1979. Sólo se puede utilizar si la distribución es simétrica y el estimador es insesgado

Método de Hall

- Sea $T_n(\theta) = \hat{\theta} - \theta$. Sea $q_n(\alpha)$ el cuantil α de T_n
- La probabilidad de que $q_n(\alpha/2) \leq T_n(\theta_0) \leq q_n(1-\alpha/2)$ es:

$$\Pr[q_n(\alpha/2) \leq T_n(\theta_0) \leq q_n(1-\alpha/2)] = 1-\alpha$$

- Si reemplazamos $T_n(\theta_0)$ por $T_n(\theta_0) = \hat{\theta} - \theta_0$

$$\Pr[q_n(\alpha/2) \leq \hat{\theta} - \theta_0 \leq q_n(1-\alpha/2)] = 1-\alpha$$

$$\Pr[\hat{\theta} - q_n(1-\alpha/2) \leq \theta_0 \leq \hat{\theta} - q_n(\alpha/2)] = 1-\alpha$$

- ¿Es esto factible? No, no lo es porque no conocemos la distribución F.
- Sólo podemos establecer el intervalo teórico que proviene de la expresión anterior como

$$C_2^0 = [\hat{\theta} - q_n(1-\alpha/2); \hat{\theta} - q_n(\alpha/2)]$$

Método de Hall

- Es tal como lo anterior, un intervalo no factible porque no conocemos F
- Sin embargo, podemos estimar F mediante bootstrap.

$$C_2^* = [\hat{\theta} - q_n^*(1-\alpha/2); \hat{\theta} - q_n^*(\alpha/2)] \quad \text{Y } C_1 \neq C_2 \text{ generalmente}$$

  C_1 será igual a C_2 si $G_n^*(u)$ es simétrica con respecto a $\hat{\theta}$.
- En éste método, por construcción podemos establecer que
  $\Pr[\hat{\theta} - q_n(1-\alpha/2) \leq \theta_0 \leq \hat{\theta} - q_n(\alpha/2)] = 1-\alpha$
- En el método de Efron, no se podía decir ni siquiera teóricamente que este resultado era alcanzable. En la práctica, ambos métodos se manejan por igual.

Método de Hall - Implementación

- Computacionalmente hablando:
  1. Estime $T_n^* = \hat{\theta}^* - \hat{\theta}$ (B replicaciones, ej. $B = 1.000$) donde $\hat{\theta}^*$ es una realización del estimador para una muestra de bootstrap. Luego tendremos B de ellos.
  2. Ordénelos de menor a mayor: $\{T_{n1}^*, T_{n2}^*, ..., T_{nB}^*\}$
  3. Obtenga $q^*(\alpha/2)$ y $q^*(1-\alpha/2)$ cortando las colas. Ejemplo: si $\alpha = 5\%$ y $B = 1.000$
  $$q^*(0,025) = T_{n,25}^*$$
  $$q^*(0,975) = T_{n,975}^*$$
  4. Construya C_2 con esos valores.

Percentile-t Equal-tailed Interval

- Suponga que queremos testear la siguiente hipótesis a una cola a un $\alpha\%$

$$H_0 : \theta = \theta_0$$

$$H_1 : \theta < \theta_0$$

- Como se puede observar, esto es un test a una cola con una zona de rechazo implícita en la cola izquierda
- Construímos el estadístico t

$$T_n(\theta) = \frac{\hat{\theta} - \theta}{se(\hat{\theta})}$$

- Un test basado en dicho estadístico rechaza cuando $T_n(\theta) < c$, con c valor crítico
- ¿Cómo se obtiene c? Necesitamos obtener un cuantil que acumule $\alpha\%$. Entonces, elegimos c tal que

$$\Pr(T_n(\theta_0 < c) = \alpha$$

$$\Rightarrow c = q_n(\alpha, F) = q_n(\alpha)$$

Percentile-t Equal-tailed Interval

- Pero F es desconocido. Podemos estimarlo mediante bootstrap. Luego, se rechaza si $T_n^*(\theta) < q_n^*(\alpha)$.
- Esto nos da pie para invertir el test y formar un intervalo de confianza. Tomar en cuenta que

$$T_n^*(\theta) = \frac{\hat{\theta}^* - \hat{\theta}}{se(\hat{\theta}^*)}$$

- Construyamos un intervalo de confianza

$$\Pr[q_n(\alpha/2) \leq T_n(\theta_0) \leq q_n(1-\alpha/2)] = 1-\alpha$$

- Este resultado se da por construcción, sin asumir simetría. Esto es similar al caso anterior, salvo que el $T_n(\theta_0)$ es distinto

$$\Pr\left[q_n(\alpha/2) \leq \frac{\hat{\theta} - \theta_0}{se(\hat{\theta})} \leq q_n(1-\alpha/2)\right] = 1-\alpha$$

$$\Pr[\hat{\theta} - se(\hat{\theta})q_n(1-\alpha/2) \leq \theta_0 \leq \hat{\theta} - se(\hat{\theta})q_n(\alpha/2)] = 1-\alpha$$

Percentile-t Equal-tailed Interval

- El intervalo teórico (no factible) queda

$$C_3^0 = [\hat{\theta} - (\hat{\theta})q_n(1-\alpha/2); \hat{\theta} - se(\hat{\theta})q_n(\alpha/2)]$$

- Pero sabemos que podemos tener una estimación desde la empírica. Luego, la contraparte bootstrap de esto último es

$$C_3 = [\hat{\theta} - se(\hat{\theta})q_n^*(1-\alpha/2); \hat{\theta} - se(\hat{\theta})q_n^*(\alpha/2)]$$

- Nótese que estamos utilizando los errores estándar ordinarios (no bootstrap).
- El nombre **"Equal-tailed"** se le da porque la probabilidad de estar a la izquierda de $\hat{\theta} - se(\hat{\theta})q_n^*(1-\alpha/2)$ es aproximadamente igual a la probabilidad de estar a la derecha de $\hat{\theta} - se(\hat{\theta})q_n^*(\alpha/2)$.

Symmetric Percentile-t Interval

- Sea $H_0 : \theta = \theta_0$, $H_1 : \theta \neq \theta_0$ un test a $\alpha\%$
- Armamos el estadístico t:

$$T_n(\theta) = \frac{\hat{\theta} - \theta}{se(\hat{\theta})}$$

- se rechaza si $|T_n(\theta)| > c$. ¿Cómo obtenemos el valor de c?
- Lo podemos obtener de resolver la siguiente ecuación:

$$\Pr(|T_n(\theta)| > c) = \alpha \tag{1}$$

- Pero no conocemos (aún) la distribución de $|T_n(\theta)|$ El módulo nos pone un problema adicional: la distribución de $T_n(\theta)$ es $G_n(\theta)$

Symmetric Percentile-t Interval

- Este problema no es tan complicado de resolver. Note que

$$\begin{aligned} \Pr(|T_n(\theta)| < c) &= 1-\alpha \\ \Pr(-c < T_n(\theta_0) < c) &= 1-\alpha \\ G_n(c) - G_n(-c) &= 1-\alpha \\ \overline{G}_n(c) &= 1-\alpha \end{aligned}$$

  donde $\overline{G}_n(c) = G_n(c) - G_n(-c)$ es la distribución de una distribución simétrica.
- Ahora podemos obtener $c = q_n(\alpha)$ donde $q_n(\alpha)$ es el cuantil $1-\alpha$ de la distribución $\overline{G}_n$. (No es exactamente la mejor notación, pero es la que usa Bruce Hansen)
- Luego, la estimación mediante bootstrap de $q_n^*(\alpha)$ se obtiene de ordenar de menor a mayor las estimaciones bootstrap del test-t, $|T_n^*| = |\hat{\theta}^* - \hat{\theta}|/se(\hat{\theta}^*)$ tomando el cuantil $1-\alpha$.

Symmetric Percentile-t Interval

- Note que tomamos dicho cuantil y no $1 - \alpha/2$ porque al tener un valor absoluto la distribución está evaluada en los reales positivos y dicho cuantil acumula las dos zonas de rechazo "si la distribución fuese simétrica" en una sola zona, en este caso.

$$C_4 = [\hat{\theta} - se(\hat{\theta})q_n^*(\alpha); \hat{\theta} - se(\hat{\theta})q_n^*(\alpha)]$$

  y claramente el test-t bootstrap rechaza si $|T_n(\theta_0)| > q_n^*(\alpha)$. Por último note que este intervalo está diseñado para funcionar "bien" y su probabilidad de cobertura es:

$$\begin{aligned} Pr(\theta_0 \in C_4) &= Pr[\hat{\theta} - se(\hat{\theta})q_n^*(\alpha) < \theta_0 < \hat{\theta} - se(\hat{\theta})q_n^*(\alpha)] \\ &= Pr[|T_n(\theta_0)| < q_n^*(\alpha)] \\ &\simeq Pr[|T_n(\theta_0)| < q_n(\alpha)] \\ &= 1 - \alpha \end{aligned}$$

- Si no tenemos certeza de la simetría de la función de distribución, este método es preferible sobre los anteriores.

Bootstrap del modelo de regresión lineal

- El modelo de regresión lineal en su forma más críptica se puede expresar de la siguiente manera

$$y_i = x_i'\beta + e_i, \qquad \mathbb{E}(e_i|x_i) = 0$$

- Si quisiéramos hacer inferencia sobre β mediante bootstrap usaríamos el método de bootstrap no-paramétrico, que es lo que hemos visto hasta ahora
- Este método remuestrea pares (y_i^*, x_i^*) de la EDF e implícitamente impone $\mathbb{E}(e_i^* x_i^*) = 0$
- Esto no garantiza $\mathbb{E}(e_i^*|x_i^*) = 0$. Luego, la distribución de bootstrap no impone los supuestos del modelo de regresión lineal y, por lo tanto, es un estimador ineficiente de la verdadera distribución cuando los supuestos del modelo de regresión se cumplen
- Un manera de lograr el supuesto de media condicional es imponer independencia de los errores y los regresores remuestreados pero es un supuesto más fuerte que lo necesario

Remuestreo de los errores - Residual and parametric bootstrap

- Primero note que la lógica del bootstrap aquí es obtener e_i^* y x_i^* de la EDF, lo cual nos obliga a tener una estimación de los errores para así generar $y_i^* = x_i^* \hat{\beta} + e_i^*$. Note que también necesitamos un estimador de β que en este caso será el de MCO.
  Existen varias formas de imponer independencia pero todas requieren remuestrear de distribuciones del error y de los regreseres independientes.
  Para generar los errores:
  - No paramétricamente: obtenga los errores bootstrap e_i^* remuestreando de los errores obtenidos mediante MCO $\hat{e}_i$ (Residual Bootsrap)
  - Paramétricamente: genere errores bootstrap de una distribución paramétrica, por ejemplo $\mathcal{N}(0, \hat{\sigma}^2)$ (Parametric Bootstrap)

Remuestreo de los regresores

- Para los regresores:
  - No parámetricamente: obtenga x_i^* remuestreando de la EDF o, en palabras sencillas, remuestreando de $\{x_1, x_2, ..., x_n\}$..
  - Paramétricamente: genere regresores bootstrap de una distribución paramétrica, por ejemplo $\mathcal{N}(\bar{x}, Var(x))$.
  - Fije: $x_i^* = x_i$, lo que equivale a tratar a los regresores como fijos en muestras repetidas. Todo el análisis será condicional en x, lo que es válido en estadística.
- Los métodos anteriores generarán errores independientes de los regresores y funcionan bajo el supuesto de homocedasticidad
- Existe un método relativamente nuevo que logra un supuesto más débil que $\mathbb{E}(e_i^* | x_i^*) = 0$ y además permite heterocedasticidad.

Wild bootstrap

- Es un método particular que construye la distribución condicional de e_i^* tal que

$$\begin{aligned} \mathbb{E}(e_i^*|x_i) &= 0 \\ \mathbb{E}(e_i^{*2}|x_i) &= \hat{e}_i^2 \\ \mathbb{E}(e_i^{*3}|x_i) &= \hat{e}_i^3 \end{aligned}$$

- Esta preservará las características más importantes de los datos
- Se puede obtener con una distribución con sólo dos puntos de masa:

$$P\left(e_i^* = \frac{1+\sqrt{5}}{2}\hat{e}_i\right) = \frac{\sqrt{5}-1}{2\sqrt{5}}$$
$$P\left(e_i^* = \frac{1-\sqrt{5}}{2}\hat{e}_i\right) = \frac{\sqrt{5}+1}{2\sqrt{5}}$$

- nos dice que cada error bootstrap será el error de MCO pero ajustado por $(1-\sqrt{5})/2$ con probabilidad $p = (\sqrt{5}-1)/2\sqrt{5}$ o $(1+\sqrt{5})/2$ con probabilidad $1-p = \sqrt{5}+1)/2\sqrt{5}$

Implementación

- Note que esta es una distribución bernoulli.
- Para implementar eso Ud. puede generar números aletorios de una distribución uniforme $[0, 1]$ y debe "invertirlos" para generar esta distribución. Esta inversión es particular porque es no lineal pero muy sencilla.
- Así, para $u < ((\sqrt{5} - 1)/2\sqrt{5})$ multiplica $\hat{e}_i$ por $(1 + \sqrt{5})/2$ y para $u \geq ((\sqrt{5} - 1)/2\sqrt{5})$ multiplica $\hat{e}_i$ por $(1 - \sqrt{5})/2$