

Tema 7. Intervalos de confianza y contrastes de hipótesis basados en remuestreos

basado en

- B. Efron, R. Tibshirani (1993). An Introduction to the bootstrap.
- O. Kirchkamp (2019). Resampling methods.

Curso 2023/2024

Introducción

- ▶ Los errores estándar se usan a menudo para calcular intervalos de confianza *aproximados* para los parámetros que se estudian.
- ▶ Dado un estimador $\hat{\theta}$ y un error estándar \hat{se} , el intervalo de confianza al 95 % habitual es

$$\hat{\theta} \pm 1,96 \cdot \hat{se}$$

donde $z_{0,025} = 1,96$ procede de una distribución normal estándar.

- ▶ En el campo de los intervalos de confianza existen diferentes técnicas bootstrap y es un área de desarrollo teórico en evolución constante.

Introducción

- ▶ Supongamos que los datos $\mathbf{x} = (x_1, x_2, \dots, x_n)$ proceden de una distribución desconocida F .
- ▶ Si el tamaño muestral n es grande la distribución de $\hat{\theta}$ converge a una normal de media θ y varianza \widehat{se}^2 , es decir

$$\frac{\hat{\theta} - \theta}{\widehat{se}} \sim N(0, 1)$$

- ▶ De este modo, se obtiene un intervalo de confianza estándar con probabilidad de recubrimiento igual a $1 - \alpha$

$$\left[\hat{\theta} - z_{1-\frac{\alpha}{2}} \widehat{se}; \hat{\theta} - z_{\frac{\alpha}{2}} \widehat{se} \right]$$

donde z_α es el percentil 100α -ésimo de una distribución normal.

Introducción

- ▶ La propiedad de recubrimiento implica que aproximadamente en el $(1 - \alpha) \cdot 100\%$ de las ocasiones este intervalo contiene el verdadero valor de θ .
- ▶ El bootstrap se puede usar para mejorar los intervalos de confianza; de hecho, cuando n es muy grande los intervalos bootstrap y los aproximados convergen a los mismos valores.
- ▶ Se puede calcular el estimador *plug-in* ($\hat{\theta} = t(\hat{F})$) del parámetro de interés $\theta = t(F)$ y, a su vez, un error estándar \hat{s}_e basado en un método bootstrap o en un método *jackknife*.

Ejemplo

- ▶ Se podría usar un enfoque *exacto* para los intervalos de confianza:
 - (I) Funciona incluso con muestras pequeñas.
 - (II) Pero requiere el conocimiento de la *verdadera* distribución.
 - (III) Para aplicarse a veces hay que hacer transformaciones de los datos.
- ▶ En el caso del ejemplo de los aparatos de aire acondicionado, si el número de fallos sigue una distribución de **Poisson**, entonces los tiempos entre fallos siguen una **distribución exponencial**.

Ejemplo

- ▶ Supongamos que la función de densidad es

$$f(x; \lambda) = \begin{cases} \lambda e^{-\lambda x} & x \geq 0 \\ 0 & x < 0 \end{cases}$$

- ▶ Para estimar λ se puede usar el estimador de máxima verosimilitud (*MLE*):

$$\hat{\lambda} = \frac{n}{\sum_i x_i} = \frac{1}{\bar{x}}$$

Ejemplo

- ▶ Al asumir una distribución exponencial, se puede calcular el intervalo de confianza exacto para el parámetro (o para la media). Ver:

`en.wikipedia.org/wiki/Exponential_distribution`

$$IC_{\frac{1}{\lambda}} = \left[\frac{\bar{x} \cdot 2n}{\chi_{2n, 1 - \frac{\alpha}{2}}^2}; \frac{\bar{x} \cdot 2n}{\chi_{2n, \frac{\alpha}{2}}^2} \right]$$

donde $\chi_{n, \alpha}^2$ es el percentil 100α -ésimo de una distribución χ_n^2 .

- ▶ De este modo se obtiene un intervalo igual a

[1] 65.89765 209.17415

Aproximación a la normal

- ▶ La aproximación a la normal se puede hacer cuando la muestra es grande.
- ▶ Así, la media muestral se distribuye asintóticamente como una normal aplicando el *TCL* (Teorema Central del Límite).
- ▶ Aproximación normal basada en estimadores paramétricos de σ :

```
confint(lm(hours ~ 1, data=aircondit))
```

```
                2.5 %    97.5 %  
(Intercept) 21.52561 194.6411
```


Intervalos t de Student

- ▶ Una manera de mejorar los intervalos asintóticos, cuando las muestras son pequeñas, es usar intervalos t de Student.
- ▶ En 1908 Gosset (the *Student*) calculó la distribución de:

$$t = \frac{\hat{\theta} - \theta}{\widehat{se}} \sim t_{n-1}$$

- ▶ Mediante esta aproximación el intervalo queda como

$$\left[\hat{\theta} - t_{n-1, 1-\alpha/2} \cdot \widehat{se} ; \hat{\theta} - t_{n-1, \alpha/2} \cdot \widehat{se} \right]$$

donde $t_{n-1, \alpha}$ es el percentil α de una distribución t de Student.

- ▶ Pero no tiene en cuenta el ajuste necesario de los intervalos de confianza cuando aparece *asimetría*.
- ▶ Se puede mejorar esta aproximación mediante los **intervalos bootstrap t** .

Intervalos bootstrap t

- ▶ Usando el bootstrap se pueden obtener intervalos ajustados sin necesidad de asumir normalidad en los datos.
- ▶ Los intervalos que se construyen reciben el nombre de aproximación *bootstrap-t*.
- ▶ En este método se calcula una tabla de cuantiles de t directamente a partir de los datos que se tienen.
- ▶ Esta tabla se usa entonces para construir intervalos de confianza de la misma manera que en el caso de los intervalos clásicos.
- ▶ La tabla de valores remuestreados se construye generando B muestras bootstrap y luego calculando la versión bootstrap del estadístico t para cada uno de ellos.

Intervalos bootstrap t

- ▶ El método consiste en generar B muestras bootstrap $\mathbf{x}^{*1}, \mathbf{x}^{*2}, \dots, \mathbf{x}^{*B}$. A continuación, para cada una de ellas se calcula

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

- ▶ donde $\hat{\theta}^*(b) = s(\mathbf{x}^{*b})$ es el valor de $\hat{\theta}$ para la muestra bootstrap \mathbf{x}^{*b} y $\hat{se}^*(b)$ es el valor estimado del error estándar de $\hat{\theta}^*$ para la muestra bootstrap \mathbf{x}^{*b} .
- ▶ El α -ésimo percentil de $Z^*(b)$ se estima mediante el valor \hat{t}_α tal que

$$\frac{\#\{Z^*(b) \leq \hat{t}_\alpha\}}{B} = \alpha$$

Intervalos bootstrap t

- ▶ Por ejemplo, si $B = 1000$ el estimador del punto 5 % es el 50-esimo valor mayor de los $Z^*(b)$ ordenados y el estimador del punto 95 % es el 950-esimo valor mayor de los $Z^*(b)$ ordenados.

- ▶ De este modo, el estimador bootstrap-t que queda es

$$\left[\hat{\theta} - \hat{t}_{1-\alpha/2} \cdot \hat{se}; \hat{\theta} - \hat{t}_{\alpha/2} \cdot \hat{se} \right]$$

- ▶ Es necesario considerar un número de réplicas B mayor que en el caso del cálculo de los errores estándar bootstrap.
- ▶ Por otro lado, los percentiles bootstrap-t pueden diferir bastante de los percentiles t de Student habituales.

Ejemplo

- ▶ Se puede observar que se necesitan extraer muchas muestras bootstrap para obtener unos resultados ajustados.
- ▶ Si comparamos ahora los cuantiles de la distribución original t de Student con los cuantiles bootstrap se obtienen resultados muy diferentes.

```
alfa = 0.05  
(extremos = quantile(bT, c(1-alfa/2, alfa/2)))
```

```
      2.5%      97.5%  
-4.504548  1.572487
```

```
(t.standard = qt(c(alfa/2, 1-alfa/2), N-1))
```

```
[1] -2.200985  2.200985
```

Características de los intervalos bootstrap-t

- ▶ Los valores de la t de Student o de la normal son simétricos alrededor del cero. De este modo se obtienen intervalos simétricos alrededor del estimador del parámetro $\hat{\theta}$.
- ▶ Por el contrario, los percentiles bootstrap-t pueden ser *asimétricos* con respecto al estimador, lo que lleva a obtener intervalos que son más largos o cortos a la derecha o la izquierda.
- ▶ Esta asimetría representa una mejora en el recubrimiento que tienen estos intervalos.
- ▶ El procedimiento bootstrap-t es una buena opción particularmente en estadísticos de localización como la media muestral, la mediana o los percentiles muestrales.

Desventajas de los intervalos bootstrap-t

- ▶ Pero hay problemas computacionales y de interpretación en los intervalos bootstrap-t.
- ▶ En el denominador del estadístico $Z^*(b)$ se necesita conocer $\widehat{se}^*(b)$, es decir, la desviación estándar de $\hat{\theta}^*$ para cada muestra bootstrap \mathbf{x}^{*b} .
- ▶ En el caso de la media, existe una expresión explícita para calcularlo, pero para muchos estadísticos no se tiene una expresión formal para ello.
- ▶ La única opción sería calcular un estimador bootstrap del error estándar para **cada** muestra bootstrap lo cual no es muy operativo.

Desventajas de los intervalos bootstrap-t

- ▶ Otro problema que se presenta en los intervalos bootstrap-t es que se puede comportar de manera errática con muestras pequeñas.
- ▶ También los intervalos bootstrap-t **no** son invariantes cuando se aplican transformaciones al parámetro original que se estima.
- ▶ A veces, interesa transformar los datos para encontrar mejores estimadores de θ . Por ejemplo, cambiando la escala de las observaciones.
- ▶ Pero a menudo no se sabe cuál es la mejor transformación para aplicar a los datos.

Intervalos *básicos*

- ▶ Otra opción para intervalos bootstrap, son los *intervalos básicos*
- ▶ Si se asume que $\hat{\theta} - \theta$ se puede aproximar o tiene la misma distribución que $\hat{\theta}^* - \hat{\theta}$, entonces, el intervalo básico se define como

$$IC_{bas} = \left[2\hat{\theta} - \hat{\theta}_{1-\frac{\alpha}{2}}^*; 2\hat{\theta} - \hat{\theta}_{\frac{\alpha}{2}}^* \right]$$

- ▶ Propiedades de los intervalos básicos:
 - (I) No son invariantes mediante transformaciones.
 - (II) Corrigen el sesgo siempre que $\hat{\theta} - \theta$ tenga la misma distribución que $\hat{\theta}^* - \hat{\theta}$.

Intervalos tipo percentil

- ▶ Se pueden considerar otro tipo de intervalos basados en los *percentiles*.
- ▶ Supongamos que se obtiene una muestra \mathbf{x}^* a partir de $\hat{P} \rightarrow \mathbf{x}^*$ y se calculan luego los estadísticos $\hat{\theta}^* = s(\mathbf{x}^*)$.
- ▶ Si se denomina \hat{G} la función de distribución acumulada de $\hat{\theta}^*$, entonces, el *intervalo percentil* de nivel $1 - \alpha$ se define como los percentiles $\frac{\alpha}{2}$ y $1 - \frac{\alpha}{2}$ de \hat{G} :

$$\left[\hat{\theta}_{\%low}; \hat{\theta}_{\%up} \right] = \left[\hat{G}^{-1} \left(\frac{\alpha}{2} \right); \hat{G}^{-1} \left(1 - \frac{\alpha}{2} \right) \right]$$

Intervalos tipo percentil

- ▶ Como por definición $\hat{G}^{-1}(\alpha) = \hat{\theta}_{(\alpha)}^*$ se puede reescribir también el intervalo como

$$\left[\hat{\theta}_{\%low}; \hat{\theta}_{\%up} \right] = \left[\hat{\theta}_{(\frac{\alpha}{2})}^*; \hat{\theta}_{(1-\frac{\alpha}{2})}^* \right]$$

- ▶ La expresión anterior se refiere a una situación ideal donde hay infinitas réplicas bootstrap. En la práctica se toma un número finito B de ellas.
- ▶ Se generan B conjuntos de datos bootstrap independientes $\mathbf{x}^{*1}, \mathbf{x}^{*2}, \dots, \mathbf{x}^{*B}$ y se calculan los respectivos estimadores $\hat{\theta}^*(b) = s(\mathbf{x}^{*b})$, donde $b = 1, \dots, B$.

Intervalos tipo percentil

- ▶ Se denomina como $\hat{\theta}_{B(\alpha)}^*$ el percentil α -ésimo de los valores $\hat{\theta}^*(b)$, es decir, el $B \times \alpha$ -ésimo valor en la lista ordenada de las B réplicas de $\hat{\theta}^*$.
- ▶ Por ejemplo, si $B = 2000$ y $\alpha = 0,05$ entonces $\hat{\theta}_{B(\alpha)}^*$ es el valor 100-ésimo en la lista ordenada de réplicas.
- ▶ Del mismo modo, se razonaría con $\hat{\theta}_{B(1-\alpha)}^*$, el percentil $1 - \alpha$ -ésimo de los valores.

Intervalos tipo percentil

- ▶ De este modo, el intervalo percentil $1 - \alpha$ se define como

$$\left[\hat{\theta}_{\%low}; \hat{\theta}_{\%up} \right] \approx \left[\hat{\theta}_{B(\frac{\alpha}{2})}^*; \hat{\theta}_{B(1-\frac{\alpha}{2})}^* \right]$$

- ▶ Si la distribución de $\hat{\theta}^*$ es aproximadamente normal, entonces los intervalos percentil y normal estándar coinciden.

2.5 %	97.5 %
48.6625	190.7583

Intervalos tipo percentil

- ▶ Para mejorar los intervalos de confianza es necesario realizar a veces transformaciones distintas a la clásica del logaritmo.
- ▶ El método del percentil permite incorporar de manera automática dicho tipo de transformaciones.
- ▶ El siguiente lema formaliza el hecho de que el método del percentil siempre *descubre* cuál es la transformación correcta.

Intervalos tipo percentil

- ▶ **Lema del intervalo del percentil:**

Supongamos que existe una función o transformación m del parámetro tal que convierte en normal a la distribución de $\hat{\theta}$, es decir $\hat{\phi} = m(\hat{\theta})$

$$\hat{\phi} \sim N(\phi, c^2)$$

con una desviación estándar dada c .

- ▶ Entonces el intervalo del percentil basado en $\hat{\theta}$ se obtiene como la transformación inversa

$$\left[m^{-1} \left(\hat{\phi} - z_{1-\alpha} \cdot c \right); m^{-1} \left(\hat{\phi} - z_{\alpha} \cdot c \right) \right]$$

Intervalos tipo percentil

- ▶ Supongamos un ejemplo donde se genera una muestra de tamaño 10 de una distribución normal estándar.
- ▶ El parámetro de interés que se trata de estimar es $\theta = e^\mu$ donde μ es la media poblacional.
- ▶ El verdadero valor de θ es $e^0 = 1$ y el estimador muestral es $\hat{\theta} = e^{\bar{x}}$.
- ▶ Si se generan muestras bootstrap $\hat{\theta}^*$ se obtiene una distribución bastante asimétrica y los intervalos percentil al 95 % resultan ser

$$\left[\hat{\theta}_{\%low}; \hat{\theta}_{\%up} \right] = [0,75; 2,07]$$

- ▶ Si se calculan los intervalos estándar normales basados en un error estándar $\hat{se} = 0,34$, se obtiene $1,25 \pm 1,96 \cdot 0,34 = [0,59; 1,92]$ que son muy diferentes a los anteriores.

Intervalos tipo percentil

- ▶ Sin embargo esta aproximación no es muy buena dado que la distribución del estadístico no es normal.
- ▶ Si se toma la transformación logaritmo sobre los datos y se considera el parámetro $\mu = \log(\theta)$ entonces se obtiene un intervalo estándar normal igual a $[-0,28; 0,73]$
- ▶ La distribución de los datos transformados es más simétrica y el intervalo estándar normal es más ajustado.
- ▶ Si se toma el antilogaritmo de los extremos de este intervalo, se obtiene $[0,76; 2,08]$ que es casi idéntico al obtenido mediante el percentil bootstrap.
- ▶ Es decir, el intervalo percentil encuentra de manera automática la transformación adecuada para los datos.

Intervalo de sesgo-correcto y acelerado BCa

- ▶ Los intervalos bootstrap-t y el percentil pueden conducir a estimaciones del intervalo de confianza algo erráticas, especialmente el intervalo percentil cuando el estimador es sesgado respecto al parámetro cuyo intervalo de confianza queremos aproximar.
- ▶ Se puede considerar una versión mejorada del intervalo percentil llamada BCa , abreviatura que procede de sesgo-correcto (*bias-corrected*) y acelerado (*accelerated*).
- ▶ En la determinación del intervalo BCa utilizaremos dos cantidades, \hat{z}_0 y \hat{a} .
- ▶ La primera, \hat{z}_0 , se introduce para corregir el sesgo del estimador $\hat{\theta}$.

Intervalo de sesgo-correcto y acelerado BCa

- ▶ \hat{z}_0 se define como

$$\hat{z}_0 = \Phi^{-1} \left(\frac{\# \{ \hat{\theta}^*(b) < \hat{\theta} \}}{B} \right)$$

- ▶ Representa el sesgo mediano de la estimación dada por las B réplicas bootstrap del estimador.
- ▶ Es decir, la discrepancia entre la mediana de dicha distribución de frecuencias y $\hat{\theta}$, en unidades normales.
- ▶ Obtenemos el valor $\hat{z}_0 = 0$, cuando exactamente la mitad de las observaciones son menores que $\hat{\theta}$.

Intervalo de sesgo-correcto y acelerado BCa

- ▶ La segunda cantidad, \hat{a} , denominada *aceleración*, corrige para el caso en que el error estándar $se(\hat{\theta})$ no sea constante, y se define en términos de los valores *jackknife*, como

$$\hat{a} = \frac{\sum_{i=1}^n (\hat{\theta}_{(\cdot)} - \hat{\theta}_{(i)})^3}{6 \left[\sum_{i=1}^n (\hat{\theta}_{(\cdot)} - \hat{\theta}_{(i)})^2 \right]^{3/2}}$$

- ▶ donde $\hat{\theta}_{(i)}$ es la i -ésima réplica *jackknife* del estimador tal que

$$\begin{aligned}\hat{\theta}_{(i)} &= s(\mathbf{x}_{(i)}) \\ \hat{\theta}_{(\cdot)} &= \frac{1}{n} \sum_{i=1}^n \hat{\theta}_{(i)}\end{aligned}$$

Intervalo de sesgo-correcto y acelerado BCa

- ▶ El intervalo de sesgo-correcto y acelerado (intervalo BCa) de coeficiente de confianza $1 - \alpha$ se define igual que el intervalo percentil con unos valores particulares α_1 y α_2

$$\left[\hat{\theta}_{\alpha_1}^* ; \hat{\theta}_{\alpha_2}^* \right]$$

- ▶ tal que

$$\alpha_1 = \Phi \left(\hat{z}_0 + \frac{\hat{z}_0 + z_{1-\frac{\alpha}{2}}}{1 - \hat{a}(\hat{z}_0 + z_{1-\frac{\alpha}{2}})} \right)$$
$$\alpha_2 = \Phi \left(\hat{z}_0 + \frac{\hat{z}_0 + z_{\frac{\alpha}{2}}}{1 - \hat{a}(\hat{z}_0 + z_{\frac{\alpha}{2}})} \right)$$

- ▶ siendo $z_{\frac{\alpha}{2}}$ y $z_{1-\frac{\alpha}{2}}$ los cuantiles habituales de la normal estándar y Φ la función de distribución de una normal estándar.

Intervalo de sesgo-correcto y acelerado *BCa*

- ▶ Se puede observar que si $\hat{z}_0 = \hat{a} = 0$ queda

$$\alpha_1 = \Phi\left(z_{1-\frac{\alpha}{2}}\right) = \frac{\alpha}{2}$$
$$\alpha_2 = \Phi\left(z_{\frac{\alpha}{2}}\right) = 1 - \frac{\alpha}{2}$$

de forma que el intervalo *BCa* coincide con el intervalo percentil.

- ▶ El valor de \hat{z}_0 traslada el intervalo a la derecha o la izquierda y \hat{a} hace que el intervalo sea más ancho o más estrecho.
- ▶ En las aplicaciones de este intervalo se recomienda utilizar, al menos $B = 1000$ muestras bootstrap.
- ▶ Al igual que el intervalo percentil, el *BCa* también *adivina* las transformaciones del parámetro.

Contrastes de hipótesis Bootstrap

- ▶ Contraste de hipótesis de dos muestras bootstrap:
 - (I) Se eligen B submuestras independientes \mathbf{x}' e \mathbf{y}' tomadas **CON reemplazamiento** de la distribución conjunta (\mathbf{x}, \mathbf{y}) .
 - (II) Se calcula $\hat{\theta}^*(b)$ para cada muestra.
 - (III) se calcula el nivel de significación alcanzado (ASL) (equivalente al *p-valor*)

$$ASL = \frac{\#\{\hat{\theta}^*(b) > \hat{\theta}\}}{B}$$

Contrastes de hipótesis BCa

- ▶ Los intervalos *BCa* resultan ser bastante eficientes para calcular intervalos de confianza.
- ▶ De este modo, la mejor opción es adaptarlos al caso de los contrastes de hipótesis.
- ▶ Se trata de calcular el valor de *ASL* a partir de intervalos de confianza para una probabilidad de recubrimiento cualquiera α .
- ▶ Se tiene que encontrar un nivel α tal que los extremos superior e inferior del intervalo de confianza coincidan con el α_1 o el α_2 observados.

Contrastes de hipótesis BCa

- ▶ Se tenía que los intervalos *BCa* se definían como

$$BCa = \left[\hat{\theta}_{(\alpha_1)}^*; \hat{\theta}_{(\alpha_2)}^* \right]$$

- ▶ con

$$\alpha_2 = \Phi \left(\hat{z}_0 + \frac{\hat{z}_0 + z_{\frac{\alpha}{2}}}{1 - \hat{a}(\hat{z}_0 + z_{\frac{\alpha}{2}})} \right)$$

- ▶ Si al hacer los remuestreos se obtiene el valor α_2 correspondiente al estadístico observado $\hat{\theta}$, entonces despejando el valor de $z_{\alpha/2}$ se obtiene que

$$ASL_{BCa} = \Phi \left(\frac{\Phi^{-1}(\alpha_2) - \hat{z}_0}{1 - \hat{a}(\Phi^{-1}(\alpha_2) - \hat{z}_0)} - \hat{z}_0 \right)$$

Varianza de los estimadores bootstrap

- ▶ La varianza de los estimadores bootstrap tiene dos causas o fuentes de variación:
 - (I) Una procede del muestreo aleatorio simple de tamaño n de la población cuyo parámetro queremos estimar.
 - (II) La otra procede del remuestreo bootstrap de tamaño B .
- ▶ Se trata de diseñar un método para estimar la varianza de los estimadores bootstrap, que se denomina *jackknife-after-bootstrap*.
- ▶ Se basa en utilizar aproximaciones jackknife en la estimación de la varianza de un estimador bootstrap.

Varianza de los estimadores bootstrap

- ▶ Tomamos el ejemplo de la estimación de $Var(\widehat{se}_B)$, la varianza del estimador bootstrap del error estándar de un estimador $\hat{\theta}$.
- ▶ El método *jackknife-after-bootstrap* consiste en eliminar primero uno de los i valores muestrales, $i = 1, \dots, n$ y luego recalcular \widehat{se}_B con las muestras bootstrap obtenidas, denominando a este valor $\widehat{se}_{B(i)}$.
- ▶ El *jackknife-after-bootstrap* se define como

$$Var_{jack}(\widehat{se}_B) = \frac{n-1}{n} \sum_{i=1}^n (\widehat{se}_{B(i)} - \widehat{se}_{B(\cdot)})^2$$

donde

$$\widehat{se}_{B(\cdot)} = \frac{1}{n} \sum_{i=1}^n \widehat{se}_{B(i)}$$

Varianza de los estimadores bootstrap

- ▶ Este método presenta la dificultad del cálculo de $\widehat{se}_{B(i)}$, ya que éste requiere un conjunto de muestras completamente nuevas para cada i .
- ▶ No obstante, podemos evitar este exceso de cálculo de la siguiente manera.
- ▶ Para cada dato x_i de los valores observados de la muestra, existen muestras bootstrap en las que no aparece x_i , pudiendo utilizarse estas muestras para estimar $\widehat{se}_{B(i)}$
- ▶ Por ejemplo, en este caso, lo estimaremos con la desviación típica de las muestras bootstrap que **no** contengan el dato x_i .

Varianza de los estimadores bootstrap

- ▶ Si C_i es el conjunto de muestras bootstrap que **no** contienen a x_i , el cual contiene $B_i \leq B$ elementos, utilizaremos el $\widehat{se}_{B(i)}$ siguiente

$$\widehat{se}_{B(i)} = \sqrt{\frac{1}{B_i} \sum_{b \in C_i} (s(\mathbf{x}^{*b}) - \bar{s}_i)^2}$$

donde

$$\bar{s}_i = \frac{1}{B_i} \sum_{b \in C_i} s(\mathbf{x}^{*b})$$

- ▶ Este método se puede aplicar, no solo en la estimación de la varianza del estimador bootstrap del error de muestreo, sino en todos los estimadores bootstrap.
- ▶ Esto falla si todas las B muestras bootstrap contienen algún dato x_i .
- ▶ Así, el número de muestras bootstrap que se debe utilizar es bastante elevado.