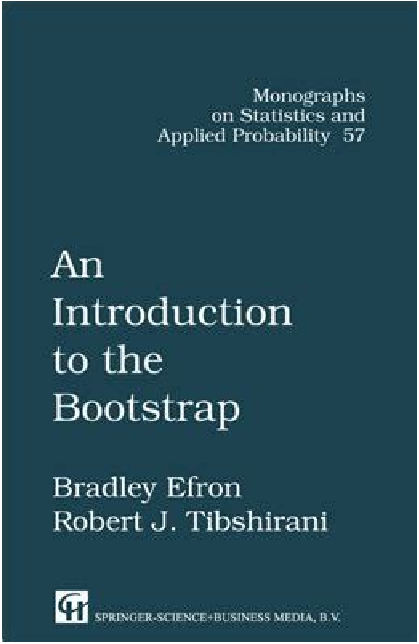
Inferencia Estadística

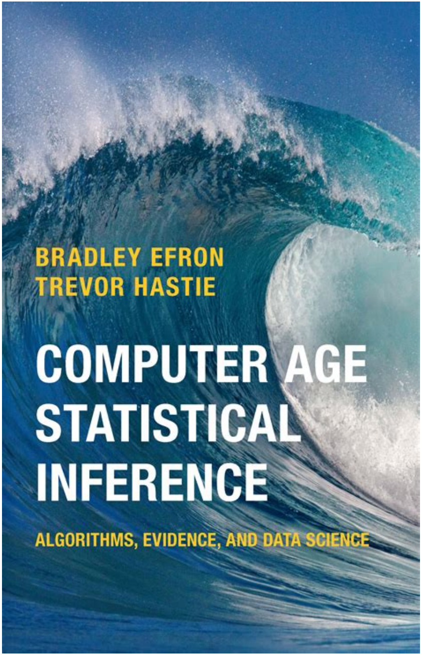
Métodos de Inferencia Basados en Remuestreos

Gabriel Martos Venturini
gmartos@utdt.edu

UTDT

UNIVERSIDAD
TORCUATO DI TELLA


Monographs
on Statistics and
Applied Probability 57
An
Introduction
to the
Bootstrap
Bradley Efron
Robert J. Tibshirani
SPRINGER-SCIENCE+BUSINESS MEDIA, B.V.


BRADLEY EFRON
TREVOR HASTIE
COMPUTER AGE
STATISTICAL
INFERENCE
ALGORITHMS, EVIDENCE, AND DATA SCIENCE

Agenda

Motivación (Jackknife y Bootstrap)

Con $\underline{x} \equiv \{x_1, \ldots, x_n\} \overset{iid}{\sim} F$ y $\widehat{\theta}_n(\underline{x})$, pretendo estimar $\text{se}(\widehat{\theta}_n) \equiv \sqrt{\text{Var}(\widehat{\theta}_n)}$.

- $\widehat{\theta}_n(\underline{x})$ es virtualmente cualquier función de los datos.
- En general resulta complejo (o imposible) obtener una expresión cerrada para (la estimación de) el error estandard de un estimador.
- ¿Y para qué quiero estimar $\text{se}(\widehat{\theta}_n)$?
  - Cuantificar la incertidumbre respecto del parámetro estimado.
- Si pudiera samplear muchas veces de F, podríamos aproximar $\text{se}(\widehat{\theta}_n)$ a partir de la distribución empírica de $\widehat{\theta}_n$ (imposible en la práctica!).
  - Jackknife y Bootstrap proponen re–muestrar (de maneras diferentes) de la única muestra que disponemos ($\underline{X} = \underline{x}$) con el fin de estimar $\text{se}(\widehat{\theta}_n)$.
  - Jackknife tiene algunas ventajas computacionales.
  - Bootstrap tiene algunas propiedades estadísticas.

Agenda

- Disponemos de datos $\underline{x} = \{x_1, \ldots, x_n\}$ y un estadístico $\widehat{\theta}_n(\underline{x})$.
- Consideremos $\underline{x}_{-i} \equiv \{x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_n\}$, para $i = 1, \ldots, n$.
- Sea $\widehat{\theta}_{-i} \equiv \widehat{\theta}_{n-1}(\underline{x}_{-i})$, la estimación Jackknife de $\mathrm{Var}(\widehat{\theta}_n)$ es:

$$\mathrm{v}_{\mathrm{jack}}(\widehat{\theta}_n) = \frac{n-1}{n} \sum_{i=1}^{n} (\widehat{\theta}_{-i} - \overline{\theta}_n)^2 \text{ y } \mathrm{se}_{\mathrm{jack}} = \sqrt{\mathrm{v}_{\mathrm{jack}}},$$

donde $\overline{\theta}_n = \sum_{i=1}^{n} \widehat{\theta}_{-i}/n$.

- La estimación Jackknife del sesgo es:

$$\mathrm{b}_{\mathrm{jack}}(\widehat{\theta}_n) = (n-1)(\overline{\theta}_n - \widehat{\theta}_n).$$

- Por lo que la estimación Jackknife del parámetro se escribe como:

$$\widehat{\theta}_{\mathrm{jack}} = \widehat{\theta}_n - \mathrm{b}_{\mathrm{jack}}(\widehat{\theta}_n) = n\widehat{\theta}_n - (n-1)\overline{\theta}_n.$$

- Ejemplo en R.

Algunas consideraciones importantes

- El método sirve para aproximar la varianza de cualquier estadístico de interés sin necesidad de asumir un modelo particular para los datos.
- ¿Porqué estimar $\mathrm{Var}(\widehat{\theta}_n)$ con $\mathrm{v}_{\mathrm{jack}}(\widehat{\theta}_n)$?
  - Porque quizá no se estimar $\mathrm{Var}(\widehat{\theta}_n)$ (ej: Estimador de correlación).
- Cuando $\widehat{\theta}_n$ no es una función suave de los datos (i.e. $\widehat{\theta}_n =$ mediana muestral), $\mathrm{v}_{\mathrm{jack}}(\widehat{\theta}_n)$ es un estimador inconsistente de $\mathrm{Var}(\widehat{\theta}_n)$.
- La estimación Jackknife de $\mathrm{Var}(\widehat{\theta}_n)$ es más rápida de computar en comparación con la estimación Bootstrap (eficiencia computacional).
- ¿Porqué utilizar $\widehat{\theta}_{\mathrm{jack}}$ en lugar de $\widehat{\theta}_n$?
  - Recordemos que $\mathrm{MSE}(\widehat{\theta}_n, \theta) = \mathrm{bias}^2(\widehat{\theta}_n) + \mathrm{var}(\widehat{\theta}_n)$. Resulta que en general $\mathrm{bias}^2(\widehat{\theta}_n)$ converge más lento a 0 que $\mathrm{bias}^2(\widehat{\theta}_{\mathrm{jack}})$.

Agenda

- Cambiamos el esquema de re-muestreo:
  1. Muestreamos $\underline{x}^* = \{x_1^*, \ldots, x_n^*\}$ de $\{x_1, \ldots, x_n\}$ con reemplazamiento ($\{X_1^*, \ldots, X_n^*\} \overset{iid}{\sim} F_n$) y con esta remuestra computamos $\widehat{\theta}_n^* \equiv \widehat{\theta}_n(\underline{x}_n^*)$.
  2. Repetimos $B \gg n$ veces el paso anterior para obtener $\{\widehat{\theta}_{n,1}^*, \ldots, \widehat{\theta}_{n,B}^*\}$.
  3. La estimación Bootstrap de $\mathrm{Var}(\widehat{\theta}_n)$ es:

$$v_{\mathrm{boot}}(\widehat{\theta}_n) = \frac{1}{B}\sum_{b=1}^{B}(\widehat{\theta}_{n,b}^* - \frac{1}{B}\sum_{b=1}^{B}\widehat{\theta}_{n,b}^*)^2 \text{ y } \mathrm{se}_{\mathrm{boot}} = \sqrt{v_{\mathrm{boot}}}.$$

- Como $F_n \approx F$ cuando $n \gg 0$ ocurre que $v_{\mathrm{boot}}(\widehat{\theta}_n) \approx \mathrm{Var}(\widehat{\theta}_n)$.

$$F_n \overset{n\gg 0}{\approx} F \overset{iid\sim F_n}{\Longrightarrow} \{\underline{x}_1^*, \ldots, \underline{x}_B^*\} \overset{B\gg 0}{\Longrightarrow} \mathrm{se}_{\mathrm{boot}} \approx \mathrm{se}(\widehat{\theta}_n).$$

- Idealmente $B \gg 0$ ($B > 200$ para estimar $\mathrm{Var}(\widehat{\theta}_n)$).
- Ejemplo: Estimando el error standard del R^2.

Intervalos de Confianza Bootstrap

- Empíricos: Con $\{\widehat{\theta}^*_b\}_{b=1}^B$ computamos $G_B(t) = \sum_{b=1}^B \mathbb{I}(\widehat{\theta}^*_b \leq t)/B$.

$$\mathrm{IC}_\alpha(\theta) = [G_B^{-1}(\alpha/2), G_B^{-1}(1-\alpha/2)].$$

- Cuando $\widehat{\theta}_n$ tiene una distribución aproximadamente normal podemos considerar el intervalo: $\widehat{\theta}_n \pm z_{\alpha/2}\mathrm{se}_{\mathrm{boot}}$.
- Existen versiones más sofisticadas en torno a estos métodos (ver por ejemplo t–Bootstrap en § 12 de Efron y Tibshirani y en § 11 de CASI).
- Caso de estudio en `R`.

Agenda

Marco general

- Habitualmente utilizado para comparar poblaciones (no normales).
- Test no paramétrico: $\{X_1, \ldots, X_n\} \overset{iid}{\sim} F_X$ e $\{Y_1, \ldots, Y_m\} \overset{iid}{\sim} F_Y$:

$$H_0 : F_X = F_Y \quad \text{vs.} \quad H_1 : F_X \neq F_Y.$$

- Sea $S \equiv S(X_1, \ldots, X_n, Y_1, \ldots, Y_m)$ un estadístico de contraste con el que generalmente rechazamos H_0 cuando S toma valores grandes:
  - Ej: $S_1 = |\overline{X}_n - \overline{Y}_m|$, $S_2 = (S_X^2 - S_Y^2)^2$, $S_3 = \sup_t |\widehat{F}_X(t) - \widehat{F}_Y(t)|$
  - ¿Cómo proceder cuando no podemos aproximar la distribución de S?
- Bajo H_0 ocurre que: $\{X_1, \ldots, X_n, Y_1, \ldots, Y_m\} \overset{iid}{\sim} F = F_X = F_Y$.
- Llamemos $s^{(\text{obs})} \equiv s(\underline{x}, \underline{y})$ al estadístico observado y consideremos las $N = (n+m)!$ permutaciones de los datos para computar $s^1, \ldots, s^N$.

$$\text{p-val} = \sum_{i=1}^{N} \mathbb{I}(s^i > s^{(\text{obs})})/N.$$

Algunos detalles

- Cuando n y m son grandes aproximamos el p-valor utilizando el método de Monte Carlo (después de elegir un estadístico S):
    1. Con los datos de las muestras computamos $s^{(\text{obs})}$.
    2. Para una secuencia de $K \ll N$ permutaciones aleatorias de los datos (muestreo sin reemplazamiento) calculamos $s^1, \ldots, s^K$ y luego:

$$\text{p-val} \approx \sum_{i=1}^{K} \mathbb{I}(s^i > s^{(\text{obs})})/K.$$

- Este tipo de test es particularmente útil cuando quiero comparar poblaciones y los tamaños muestrales son relativamente pequeños.
- Caso de estudio en `R` (próxima slide).

Caso de estudio

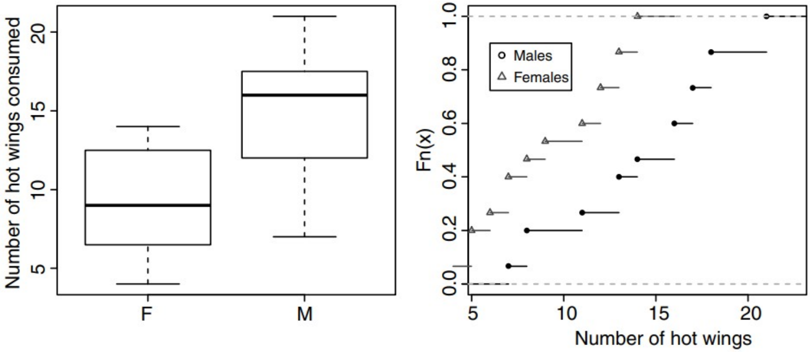


Figure: Diferencia en los patrones de consumo de 'alitas de pollo' entre hombres y mujeres (caso tomado de *Mathematical Statistics with resampling in R*).

- Tamaño muestral reducido: $n = m = 15$.