What to do today (Nov 2, 2020)?

1. Introduction

2. Probability and Distribution (Chp 1-3)

3. Essential Topics in Mathematical Statistics

3.1 Elementary Statistical Inferences (Chp 4)

- *3.1.1 Sampling and Statistics*
- *3.1.2 Confidence Interval*
- *3.1.3 Order Statistics*
- *3.1.4 Hypothesis Testing*
- **3.1.5 Statistical Simulation and Bootstrap**

3.2 Consistency and Limiting Distributions (Chp 5)

- **3.2.1 Convergence in Probability**
- **3.2.2 Convergence in Distribution**

3.3 Maximum Likelihood Methods (Chp 6)

3.1.5 Statistical Simulation and Bootstrap: Preparation for Bootstrap

Consider rv $X \sim F(\cdot)$: iid observations $X_1, \ldots, X_n$

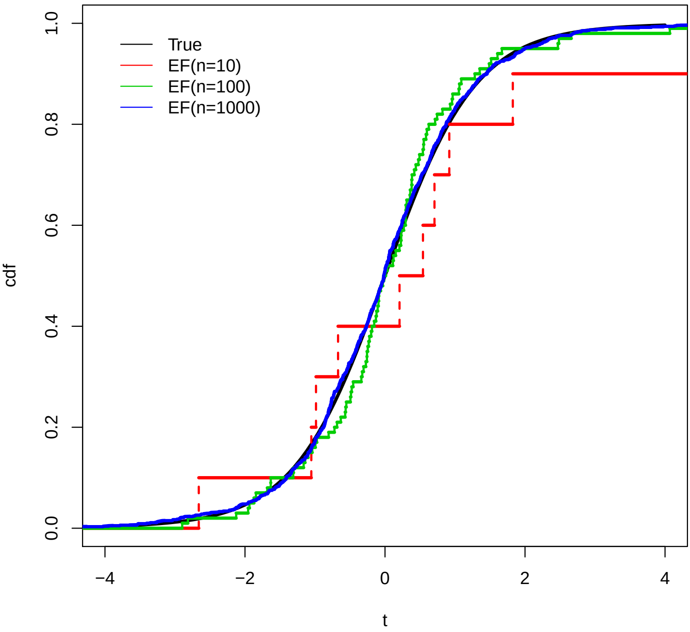
Definition. The following is the **empirical function** with the random sample:

$$F_n(x) = \frac{1}{n}\sum_{i=1}^{n} \mathrm{I}(X_i \le x) \quad \text{for} \quad -\infty < x < \infty.$$

For all $x \in (-\infty, \infty)$,

- $E\{F_n(x)\} = F(x)$ and $Var\{F_n(x)\} = F(x)\big[1 - F(x)\big]/n$.
- more ... ...

$F_n(\cdot)$ is a very good estimator of $F(\cdot)$.



$T \sim t(6)$.

3.1.5 Statistical Simulation and Bootstrap: Bootstrap

Consider rv $X \sim F(\cdot)$: iid observations $X_1, \ldots, X_n$

- When to use a point estmator $\widehat{\theta}(X_1, \ldots, X_n)$ of a population parameter θ, how to estimate its variance $Var(\widehat{\theta})$?

Thinking ... ...

- If we could have a random sample $\widehat{\theta}_b$ for $b = 1, \ldots, B$ from the same population as $\widehat{\theta}$, we can estimate the variance with

$$s^2_{\widehat{\theta}} = \sum_{b=1}^{B} (\widehat{\theta}_b - \bar{\widehat{\theta}})^2 / (B - 1)$$

  with $\bar{\widehat{\theta}} = \sum_{b=1}^{B} \widehat{\theta}_b / B$.
- That can be achieved if there are $X_{1b}, \ldots, X_{nb}$ iid from $F(\cdot)$. However, $F(\cdot)$ is unknown. How to overcome it?

3.1.5 Statistical Simulation and Bootstrap: Bootstrap

Consider rv $X \sim F(\cdot)$: iid observations $X_1, \ldots, X_n$

- When to use a point estmator $\widehat{\theta}(X_1, \ldots, X_n)$ of a population parameter θ, how to estimate its variance $Var(\widehat{\theta})$?

Bootstrap variance estimation:

- Step 1. Generate $X_{1b}^*, \ldots, X_{nb}^*$ iid from the empirical function $F_n(\cdot)$.
  (*Resample with size n from* $X_1, \ldots, X_n$ *with replacement.*)
- Step 2. Calculate $\widehat{\theta}(X_{1b}^*, \ldots, X_{nb}^*)$, denoted by $\widehat{\theta}_b^*$.
- Repeat Steps 1. and 2. B times and obtain $\{\widehat{\theta}_b^* : b = 1, \ldots, B\}$.
- With $\bar{\widehat{\theta}}^* = \sum_{b=1}^{B} \widehat{\theta}_b^* / B$, calculate $s_{\widehat{\theta}^*}^2 = \sum_{b=1}^{B} (\widehat{\theta}_b^* - \bar{\widehat{\theta}}^*)^2 / (B-1)$.
- Use $s_{\widehat{\theta}^*}^2$ to estimate $Var(\widehat{\theta})$.

Resampling methods: Jackknife (J.W. Tukey, 1958); Bootstrap (Bradley Efron, 1979)

3.1.5 Statistical Simulation and Bootstrap: Bootstrap

Consider rv $X \sim F(\cdot)$: iid observations $X_1, \ldots, X_n$

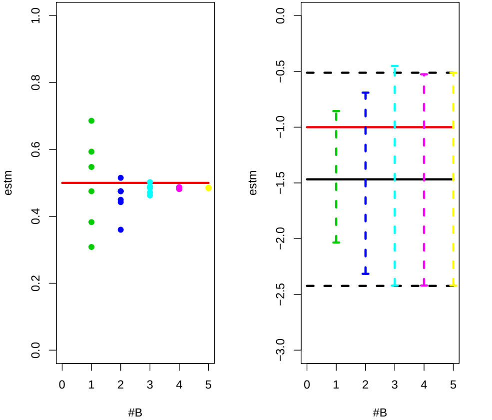
- How to obtain an interval estimator of a population parameter θ based on a point estmator $\widehat{\theta}(X_1, \ldots, X_n)$?

Bootstrap confidence interval:

- Step 1. Generate $X_{1b}^*, \ldots, X_{nb}^*$ iid from the empirical function $F_n(\cdot)$.
  (*Resample with size n from* $X_1, \ldots, X_n$ *with replacement.*)
- Step 2. Calculate $\widehat{\theta}(X_{1b}^*, \ldots, X_{nb}^*)$, denoted by $\widehat{\theta}_b^*$.
- Repeat Steps 1. and 2. B times and obtain $\left\{\widehat{\theta}_b^* : b = 1, \ldots, B\right\}$.
- Sort the sequence as $\widehat{\theta}_{(1)}^* \le \ldots \le \widehat{\theta}_{(B)}^*$, and obtain bootstrap percentiles: $\widehat{\theta}_{((\alpha/2)100)}^*$ and $\widehat{\theta}_{((1-\alpha/2)100)}^*$.
- Use $(\widehat{\theta}_{((\alpha/2)100)}^*, \widehat{\theta}_{((1-\alpha/2)100)}^*)$ as a $(1-\alpha)100\%$ CI for θ.

Bootstrap example

Consider $X \sim F(\cdot)$ with $\mu = E(X)$ and iid obs $X_1, \ldots, X_n$.



(a) Bootstrap Estm for SD of Xbar

(b) Bootstrap CI Estm for Population Mean

$X \sim N(-1, 5^2)$ with $n = 100$ and $B = 10^k$ for $k = 1, \ldots, 5$.

3.2.1 Convergence in Probability

Definition. We say a sequence of random variables (rvs) $\{Y_n : n = 1, 2, \ldots\}$ **converges in probability** to rv Y if, for any $(\forall)$ $\epsilon > 0$,

$$\lim_{n\to\infty} P\left(\left|Y_n - Y\right| \geq \epsilon\right) = 0.$$

Denote it by "$Y_n \to Y$ in probability" as $n \to \infty$, or "$Y_n \xrightarrow{P} Y$ as $n \to \infty$".

- A special case is that $Y_n \xrightarrow{P} c$, a constant.

Theorem. (Weak Law of Large Numbers (WLLN)). Let $\{X_n\}$ be a sequence of iid rvs with the common mean μ and variance $\sigma^2 < \infty$. Let $Y_n = \bar{X}_n = \sum_{i=1}^n X_n / n$. Then the sequence $\{Y_n\}$ converges in probability to μ. That is, $\bar{X}_n \xrightarrow{P} \mu$ as $n \to \infty$.

Definition. Let $X_1, \dots, X_n$ be a sample from $F(x;\theta), \theta \in \Omega$. A statistic of the sample, denoted by T_n, is a **consistent** estimator of θ if $T_n \xrightarrow{P} \theta$ as $n \to \infty$.

Comments:

- By WLLN, $\bar{X}_n \xrightarrow{P} \mu = E(X)$ as $n \to \infty$ if the observations on X are iid. That is, $\bar{X}_n$ is a consistent estimator of μ.
- The sequence of sample proportions converges in probability to the population proportion. (*View it as the theoretical support to the frequentist definition for probability.*)

Theorem. Suppose $X_n \xrightarrow{P} X$ and $Y_n \xrightarrow{P} Y$ as $n \to \infty$. The following results hold:

- $X_n + Y_n \xrightarrow{P} X + Y$, and $X_n Y_n \xrightarrow{P} XY$.
- $g(X_n) \xrightarrow{P} g(X)$ for any continuous function $g(\cdot)$.

eg, the sample variance $\left[\sum_{i=1}^n X_i^2 - n\bar{X}^2\right]/(n-1)$ is a consistent estimator of $\sigma^2 = Var(X)$ if $X_1, \dots, X_n$ is a random sample from the population.

3.2.2 Convergence in Distribution

Definition. Consider rv $X \sim F_X(\cdot)$. We call a sequence of rvs $\{X_n\}$ **converges in distribution** to X if

$$\lim_{n\to\infty} F_{X_n}(x) = \lim_{n\to\infty} P(X_n \leq x) = F(x)$$

for all $x \in \mathcal{C}(F_X)$, the set of all continuous points of $F_X(.)$. Denote it by $X_n \xrightarrow{D} X$. In other words, the **limiting distribution** or the **asymptotic distribution** of $\{X_n\}$ is $F_X(\cdot)$.

Recall the Central Limit Theorem (CLT):
Theorem. (CLT) If $X_1, \ldots, X_n$ are iid with mean μ and variance σ^2,

$$Y_n = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left(\frac{X_i - \mu}{\sigma} \right) \xrightarrow{D} N(0,1) n \to \infty.$$

- That is, $(\bar{X} - \mu)/\sqrt{\sigma^2/n} \xrightarrow{D}\rightarrow N(0,1)$ as $n \to \infty$.
- That is, $\sum_{i=1}^{n} X_i \sim N(n\mu, n\sigma^2)$ approximately as $n >> 1$.

3.2.2 Convergence in Distribution

Theorem. Consider a sequence of rvs $\{X_n\}$.

- If $X_n \xrightarrow{P} X$, then $X_n \xrightarrow{D} X$.
- If $X_n \xrightarrow{D} c$ with c a constant, then $X_n \xrightarrow{P} c$.

Theorem. Consider sequences of rvs $\{X_n\}$ with $X_n \xrightarrow{D} X$.

- If $g(\cdot)$ is a continuous function on the support of X, then $g(X_n) \xrightarrow{D} g(X)$.
- If rvs $A_n \xrightarrow{P} a$ and rvs $B_n \xrightarrow{P} b$ with both a and b constant, then

$$A_n + B_n X_n \xrightarrow{D} a + bX.$$

(Slutsky's Theorem)

Example 3.6 Suppose $X_1, \ldots, X_n$ is a random sample from the uniform distn $U(0, \theta)$.

- $Y_n = \max(X_1, \ldots, X_n)$ can be a "good estimator" of θ.
- Y_n's distn:

$$F_{Y_n}(y) = \begin{cases} 1, & y > \theta \\ \left(y/\theta\right)^n, & 0 < y \le \theta \\ 0, & t \le 0; \end{cases} \quad f_{Y_n}(y) = \begin{cases} ny^{n-1}/\theta^n, & 0 < y \le \theta \\ 0, & \textit{elsewhere}. \end{cases}$$

- $E(Y_n) = n\theta/(n+1)$, a biased estimator of θ.
- $F_{Y_n}(y) \to 1$ or 0 for $y \ge \theta$ or $y < \theta$, respectively. Thus $Y_n \xrightarrow{D} \theta$.
- Since θ is a constant, $Y_n \xrightarrow{P} \theta$: Y_n is consistent.
- Furhter, let $W_n = n(\theta - Y_n)$. The distn of W_n

$$P(W_n \le t) = P(Y_n \ge \theta - t/n) = 1 - \left(1 - \frac{t/\theta}{n}\right)^n$$

  converges to $1 - \exp(-t/\theta)$. That is $W_n \xrightarrow{D} W$, which follows the exponential distn $NE(\theta)$.

What will we study next class?

1. *Introduction*
2. *Probability and Distribution (Chp 1-3)*
3. **Essential Topics in Mathematical Statistics (Chp 4-6)**
    - *3.1 Elementary Statistical Inferences (Chp 4)*
    - *3.2 Consistency and Limiting Distributions (Chp 5)*
        - *3.2.1 Convergence in Probability*
        - *3.2.2 Convergence in Distribution*
    - **3.3 Maximum Likelihood Methods (Chp 6)**
4. *Further Topics, Selected from Chp 7-11*