What to do today (Oct 21, 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)

3.3 Maximum Likelihood Methods (Chp 6)

Example 3.2 From a sample of 1250 BC voters, 420 of them indicate that they support the NDP. Obtain an approximate 95% CI for the proportion of BC voters who support the NDP. [(.310, .362)]

- Population: r.v. X=1 or 0 to indicate a vote for NDP in BC. $X \sim (1, p)$.
- Random sample: iid r.v.s $X_1, X_2, \ldots, X_{1250}$ (votes from BC). with $\bar{X}_{obs} = \bar{x} = 420/1250$ ($\hat{p}$).
- 95% Confidence Interval of $\theta = \mu = E(X) = p$:

$$\hat{\theta}_L = \bar{x} - 1.96 \frac{s}{\sqrt{1250}}, \quad \hat{\theta}_U = \bar{x} + 1.96 \frac{s}{\sqrt{1250}}$$

$s^2 = \frac{n}{n-1} \hat{p}(1-\hat{p}) \approx \hat{p}(1-\hat{p}) = 0.223 \Longrightarrow$ an approximate 95% CI: $(.310, .362)$.

3.1.2 Confidence Interval: to estimate other population parameter

Consider the following topics:

- How to estimate $\sigma^2 = Var(X)$ in the population with disn $N(\mu, \sigma^2)$?

- How to estimate $\mu_X - \mu_Y$ for the two populations, say, X, Y?

- How about to estimate θ when $X \sim F(x;\theta)$ in general?

3.1.3 Order Statistics: Definition

Consider a collection of rvs $X_1, \ldots, X_n$.

Definition. The **order statistics** of the rvs are $Y_1 = \min(X_1, \ldots, X_n) \leq \ldots \leq Y_n = \max(X_1, \ldots, X_n)$, often denoted by $X_{(1)}, \ldots, X_{(n)}$.

Definition. If $X_1 = Y_{R_1} = X_{(R_1)}, \ldots, X_n = Y_{R_n} = X_{(R_n)}$, $R_1, \ldots, R_n$ are the ranks of $X_1, \ldots, X_n$ in the collection.

eg, if X_1, X_2, X_3's obsns are $x_1 = 2.3, x_2 = 5.9, x_3 = 1.1$,

- the obsn of the order statistics are $x_{(1)} = y_1 = 1.1, x_{(3)} = y_2 = 2.3, x_{(3)} = y_3 = 5.9$
- the obsn of the rank statistics are $r_1 = 2, r_2 = 3, r_3 = 1$.

3.1.3 Order Statistics: Property

Consider a random sample $X_1, \ldots, X_n$ from a population with cdf $F(x)$.

- What're the distn of $Y_1 = X_{(1)} = \min(X_1, \ldots, X_n)$ and $Y_n = X_{(n)} = \max(X_1, \ldots, X_n)$?

$$P(Y_n \leq y) = P(X_1 \leq y, \ldots, X_n \leq y) = F(y)^n$$

$$P(Y_1 \leq y) = 1 - P(X_1 > y, \ldots, X_n > y) = 1 - [1 - F(y)]^n$$

- When the population distn is continuous with pdf $f(x)$:

Theorem. The joint pdf of its order statistics $Y_1, \ldots, Y_n$ is

$$g(y_1, \ldots, y_n) = \begin{cases} n! \prod_{i=1}^n f(y_i) & 0 < y_1 \leq \ldots \leq y_n \\ 0 & \textit{elsewhere}. \end{cases}$$

For example, if $f(x) = 1$ for $0 < x < 1$, the pdf of the *range* $Y_n - Y_1$ with $n = 3$ is $h(z) = 6z(1 - z)$ for $0 < z < 1$.

3.1.3 Order Statistics: Related

Consider a random sample $X_1, \ldots, X_n$ from a population with cdf $F(x)$.

Definition. The **population** p **quantile** of $F(x)$ is $\xi_p = F^{-1}(p)$ for $0 < p < 1$.

Definition. The **sample** p **quantile** of the random sample is $X_{(k)}$ if $k = \lfloor p(n+1) \rfloor$, the largest integer that not larger than $p(n+1)$.

- If $p = 0.5$, $\xi_{0.5}$ is the population median, and the sample median is often $X_{(.5(n+1))}$ if n's odd and $\{X_{(.5n+1)} + X_{(.5n)}\}/2$ if n's even.
- $\xi_{0.25}$ and $\xi_{0.75}$ are the population 1st and 3rd *quartiles*, respectively; the sample quartiles $Q_0 = X_{(1)}, Q_1 = X_{(.25(n+1))}, Q_2 = X_{(.5(n+1))}, Q_3 = X_{(.75(n+1))}, Q_4 = X_{(n)}$ are often used to do the *five-number summary*.
- The population/sample p-quantile is often referred to as the $100p$th percentile of the population/sample.

3.1.4 Hypothesis Testing: Basic Setup

- **Population.** Suppose r.v. $X \sim F(\cdot;\theta)$, where θ is an unknown *parameter*.
- **Data (observations).** Suppose $X_1, \cdots, X_n$ are iid and arise from $F(\cdot;\theta)$.
- **Goal.** To test on $H_0 : \theta \in \Omega_0$ vs $H_1 : \theta \in \Omega_1$ using the random sample. $(\Omega_0 \bigcap \Omega_1 = \emptyset)$

How to achieve the goal?

$\Longrightarrow$ testing procedures for making an inference on **the null hypothesis** H_0 vs **the alternative hypothesis** H_1?

3.1.4 Hypothesis Testing: Basic Concepts

Let $\mathcal{D} = \{\textit{all possible realizations of } (X_1, \ldots, X_n)\}$.

Definition. Set $\mathcal{C} \subseteq \mathcal{D}$ is called the **rejection region** for a hypothesis test if the test's decision rule is as follows:

Reject H_0 (Accept H_1) if $(x_1, \ldots, x_n) \in \mathcal{C}$;

Acdept H_0 (Reject H_1) if $(x_1, \ldots, x_n) \notin \mathcal{C}$.

- The **Type I error** of a test occurs if H_0 is rejected when H_0 is true; the **Type II error** of a test occurs if H_0 is accepted when H_1 is true.
- We say the rejection region $\mathcal{C}$ is of **size** (or significance level) α if $\alpha = \max_{\theta \in \Omega_0} P_\theta\{(X_1, \ldots, X_n) \in \mathcal{C}\}$.
  The **power function** of the test is
  $\textit{power}(\theta) = P_\theta\{(X_1, \ldots, X_n) \in \mathcal{C}\}$ for $\theta \in \Omega_1$.

3.1.4 Hypothesis Testing: Examples

Example 3.3 To test for a binomial proportion of success at size α: $X \sim B(1, \theta)$ with $H_0 : \theta = \theta_0$ vs $H_1 : \theta < \theta_0$, provided a random sample $X_1, \ldots, X_n$

Example 3.4 To test at $\alpha = .05$ on whether a six-face die is even by rolling it 60 times indptly with the outcomes

face	1	2	3	4	5	6
count	13	19	11	8	5	4

What will we study next week?

1. *Introduction*

2. *Probability and Distribution (Chp 1-3)*

3. **Important Topics in Mathematical Statistics (Chp 4-6)**

   - **3.1 Elementary Statistical Inferences**
     - *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*
   - *3.3 Maximum Likelihood Methods*

4. *Further Topics, Selected from Chp 7-11*