STA260 Notes

Wed, Jul 2 2025 STA260 Lec I (Mathematical or Theoretical Statistics)

Office hours: Can ask any questions

Tips: Come to lectures, come to tutorial, etc

Thank god: don't expect to see too many surprises.

Hogg McKean textbook: read the second textbook maybe?

First term test: July 18

Need to remember STA256H5

Resources: Mathpix

The beginning: In a typical statistical problem, we have a random variable (r.v.) whose underlying distribution is known, but its parameters are unknown.

- We call θ a parameter of dist
- The goal is to estimate the unknown θ based on
- available data or observations
- A statistic (singular) is a function of the observations which summarizes (reduces) the data

-> Statistics is the study of statistic

Ex: $Exp(\lambda) \to \lambda$ is unknown

$X \sim Exp(\lambda)$ $\quad X \sim Poisson(\lambda)$ $\quad \to \quad X \sim Exp(\theta)$ $\quad X \sim Poisson(\theta)$

$X \sim N(\mu, \sigma^2)$ $\quad X \sim N(\theta_1, \theta_2)$

Ex: lifetime $82, 55, 65, \ldots 10^6$ obs.

$$\bar{x} = \frac{82+55+\ldots}{10^6} = 80.5$$

-> Sufficiency of statistics

-> entails all information of sample

- Let us assume that

$$X_1, X_2, \ldots, X_n \sim f(x; \theta)$$

(n: sample size; θ: parameterized)

The sequence of r.v.'s $X_1, \ldots, X_n$ is called a sample where each r.v. is called an observation

Once the sample is realized, we denote numerical values taken by each observation using lower-case letters

$x_1, x_2, \ldots, x_n$

realized observations once the sample is drawn

$X_1 \sim Exp(\theta)$

$X_1 = 85 \to$ now is not random: is predetermined

A statistic is always a function of sample

$$T = g(X_1, X_2, \ldots, X_n)$$

$$\bar{X} = \frac{X_1 + X_2 + \ldots + X_n}{n}$$

$$X_{(n)} = \max\{X_1, X_2, \ldots, X_n\}$$

Once the sample is drawn, the t is called the realization of statistic T, where

$$t = g(x_1, x_2, \ldots, x_n)$$

$$\bar{x} = \frac{x_1 + x_2 + \ldots + x_n}{n} = 80.5$$

-> called an estimate

Estimator vs. Estimate:

While we call T an estimator of Θ, we call its realization t an estimate of Θ

Estimator $\bar{X}$

estimate $\bar{x}$

★ A sample is called random if

$$X_1, X_2, \ldots, X_n \overset{iid}{\sim} f(x;\theta)$$

Break over:

After you get Statistics, get Sampling Dist is the underlying dist. of Statistics

$\bar{X} = \frac{\sum X_i}{n}$ sample mean → new distribution

$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$ Sample variance estimates σ^2 → new distribution

$E[\bar{X}] = \mu$ $\quad E[S^2] = \sigma^2$

unbiased u $u+10$

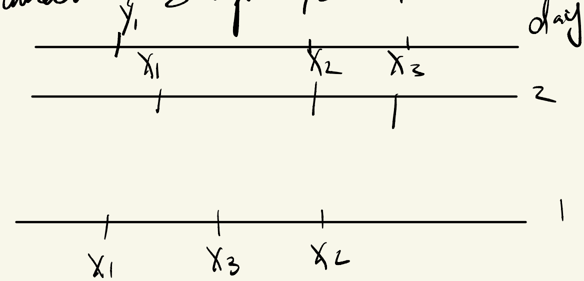
$\bar{X} \xrightarrow{P} \mu$ $\quad S^2 \xrightarrow{P} \sigma^2$

Order Statistics

Let $X_1, X_2, \ldots, X_n$ be a random Sample from $f(x;\theta)$

Then $Y_1, Y_2, \ldots, Y_n$ represents the order statistics associated to the random Sample

Y_1 → minimum $\quad Y_2 \quad Y_3$ → maximum



$X_1 :=$ price AAPL

$X_2 :=$ " " IBM

$X_3 :=$ " " TESLA

$X_{(1)}, X_{(2)}, \ldots, X_{(n)}$

minimum $\quad$ maximum

Find the Sampling dist. of maximum order statistic

Interested in ordered statistics to find maximum profit.

$\max\{2,3\} \le 5 \Rightarrow \begin{cases} 2 \le 5 \\ 3 \le 5 \end{cases}$

$Y_n = \max\{X_1, X_2, \dots, X_n\}$

$F_{Y_n}(y) = P[Y_n \le y] = P[\max(X_1, \dots, X_n) \le y]$

$= P[X_1 \le y, X_2 \le y, \dots, X_n \le y]$

$\overset{\text{indep}}{=} P[X_1 \le y]\, P[X_2 \le y] \cdots P[X_n \le y]$

$= F_X(y) F_X(y) \dots F_X(y)$

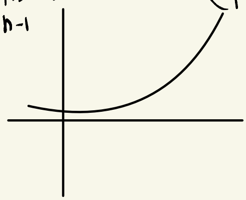
$= [F_X(y)]^n$

$F_{Y_n}(y) = [F_X(y)]^n$

$f_{Y_n}(y) = \frac{d}{dy} F_{Y_n}(y) = n f_X(y) [F_X(y)]^{n-1}$

$X_1, X_2, \dots, X_n \overset{iid}{\sim} \text{Unif}(0,1)$ — $f_X(x) = 1$, $F_X(x) = x$

$f_{Y_n}(y) = n \cdot y^{n-1}$

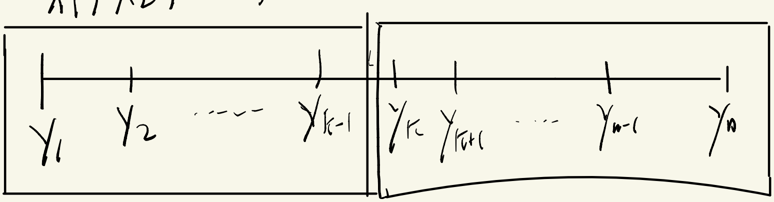


Sampling dist. of maximum

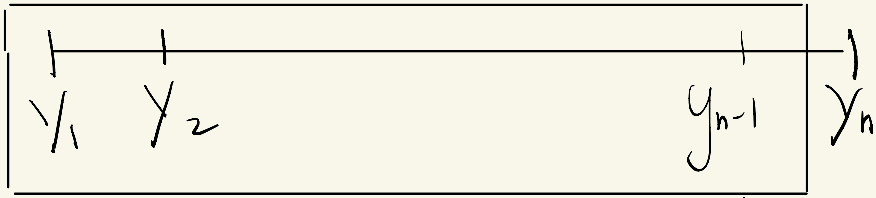
Heuristic Approach

$Y_n - Y_1$ range statistic

$X_1, X_2, \dots, X_n \overset{iid}{\sim} f_X(x)$



$f_{Y_k}(y_k) = \frac{n!}{(k-1)!(n-k)!1!} [F_X(y_k)]^{k-1} [1 - F_X(y_k)]^{n-k} f_X(y_k)$



$$f_{y_n}(y_n) = \frac{n!}{(n-1)!\,1!}\left[F_X(y_n)\right]^{n-1} f_X(y_n)$$

$$= n\left[F_X(y_n)\right]^{n-1} f_X(y_n)$$

Let $Y_1 < Y_2 < Y_3 < Y_4$ be ordered statistic of random sample $f_X(x) = 2x \quad 0<x<1$

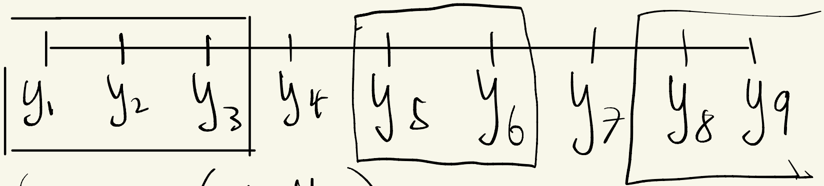
$$F_X(x) = x^2$$

$P\left[Y_3 > \frac{1}{2}\right]$

$$f_{Y_3}(y_3) = \left[F_X(y_3)\right]^2 \frac{4!}{2!\,1!\,1!}\left[1 - F_X(y_3)\right]^1 f_X(y_3)$$

$$P\left[Y_3 \geq \frac{1}{2}\right] = \int_{1/2}^{\infty} f_{Y_3}(y_3)\,dy_3$$

$$= \int_{1/2} 24\left(y_3^5 - y_3^7\right) dy_3 = \frac{243}{256}$$

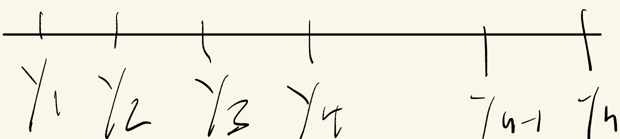


$$f_{Y_4,Y_7}(y_4,y_7) = \frac{9!}{3!\,2!\,2!}\left[F_X(y_4)\right]^3\left[1-F_X(y_7)\right]^2\left[F_X(y_7)-F_X(y_4)\right]^2 f_X(y_4)\, f_X(y_7)$$

☆ Review MGF yourself

For order statistics, can always do the heuristic approach.

Practice: distr. of $X_{(1)}$



$$f_{X_{(1)}}(x_1) = \frac{n!}{1!\,(n-1)!} = (n)\left(f_X(x_1)\right)\left(1-F(x_1)\right)^{n-1}$$

Actual STA256 Review time

1. $Y_1, \dots, Y_n$ be independent random variables with the exponential dist. with mean λ, where $\lambda > 0$.

(a) For $y \geq 0$,

cdf of Y_1 is $F_{Y_1}(y) = P_Y(Y \leq y)$

$$= \int^{y} f(t)\,dt$$

$Y_1 \sim Exp(\lambda)$

$f(y) = \frac{1}{\lambda} e^{-\frac{y}{\lambda}}$

$$= \int_0^{y} \frac{1}{\lambda} e^{-\frac{t}{\lambda}}\,dt$$

$$= -e^{-\frac{t}{\lambda}} \Big|_0^{y} = -e^{-\frac{y}{\lambda}} + 1, \quad y \geq 0.$$

$\therefore F_{Y_1}(y) = \begin{cases} 0 & \text{if } y < 0 \\ 1 - e^{-\frac{y}{\lambda}} & \text{if } y \geq 0 \end{cases}$

Order Statistics Topic 1:

Ex. $n=4$ Data: 5, 9, 3, 7 : $X_{(1)} \leq X_{(2)} \leq X_{(3)} \leq X_{(4)}$

order: 3, 5, 7, 9

min $x_{(1)}$

max $x_{(4)}$

- new data: 10, 15, 0, 3

0, 3, 10, 15 → max

min

$F_{X_i}(y) = \int_{-\infty}^{y} f(t)\,dt$

Let $X_1, \ldots, X_n \sim f(x)$

The order of Statistics are defined as:

$X_{(1)} \le X_{(2)} \le \ldots \le X_{(n)}$

Let $X_1, \ldots, X_n \sim f(x)$ be samples. all possible values for r.v.

Distribution of the minimum $F_{X_1}(x) = P(X_1 \le x)$ ≡ Prob. that min value will be less than x

$X_1, X_2, \ldots, X_n \sim f(x)$ its own distr.

Derivation: $X_{(n)}$: let $Y = X_{(n)} = \max(X_1, \ldots, X_n)$

Use the cdf approach

The cdf of Y is $F_Y(y) = P(Y \le y)$

$$= P(\max(X_1, \ldots, X_n) \le y)$$

$$= P(X_1 \le y, X_2 \le y, \ldots, X_n \le y)$$

$$\overset{\text{indep}}{=} P(X_1 \le y)\,P(X_2 \le y) \ldots P(X_n \le y)$$

$$= F_{X_1}(y) \times F_{X_2}(y) \times \ldots \times F_{X_n}(y) = F(y)^n$$

Now,

$$F_{X_{(n)}}(y) = [F(y)]^n$$

$$f_{X_{(n)}}(y) = n[F(y)]^{n-1} \times f(y)$$ "Chain rule"

eq. doesn't matter

$X_{(1)}$: Let $Y = \min(X_1, \ldots, X_n)$

cdf of $Y = P(Y \le y)$

$$= P(\min(X_1, \ldots, X_n) \le y)$$

$$= 1 - P(\min(X_1, \ldots, X_n) \ge y)$$

$$= 1 - P(X_1 \ge y, X_2 \ge y, \ldots, X_n \ge y)$$

$$\overset{\text{indep}}{=} 1 - P(X_1 \ge y) \times P(X_2 \ge y) \times \ldots \times P(X_n \ge y)$$

$$= 1 - (1 - P(X_1 \le y))(1 - P(X_2 \le y))(1 - P(X_n \le y))$$

$$= 1 - (1 - P(X_1 \le y))^n$$

$$= 1 - (1 - F(y))^n$$

$$F_{X(1)}(y) = 1-(1-F_x(y))^n$$

Now, $f_{X(1)}(y) = 0 - n(1-F_x(y))^{n-1} \times (0 - f_x(y))$

$$= n(1-F_x(y))^{n-1} \times f_x(y)$$

Example: $X_1, \ldots, X_n$ be iid with pdf $f_x(x) = 1$ for $0 \le x \le 1$ $\Rightarrow X \sim U(0,1)$

Let $Y = \max(X_1, \ldots, X_n)$.

① $f_y(y) = n(F_x(y))^{n-1} \times f_x(y)$

$F_x(x) = \int_0^x (1)\,dt = t\Big|_0^1 = 1$

• $F_x(y) = P(X \le y) = \int_{-\infty}^{y} f(t)\,dt = \int_0^y 1\,dt = y$

$\therefore f_y(y) = n * (y)^{n-1} * 1 = ny^{n-1}, \quad 0 \le y \le 1$

$$E[Y] = \int_0^1 y f_y(y)\,dy = \int_0^1 y\, n y^{n-1}\,dy = n\int_0^1 y^n\,dy = n \frac{y^{n+1}}{n+1}\Big|_0^1 = \frac{n}{n+1}$$

Ex: $n=4$, $X_i \overset{iid}{\sim} U(0,1)$

$$E(X_{(n)}) = \frac{4}{4+1} = \frac{4}{5}.$$

cdf: $F_X(x) = \int_{-\infty}^{x} f(t)\,dt$ defined on the domain of r.v. X

Order statistics review:

$X_1, X_2, \ldots, X_{15}$ be Sample of Size 15 from $X_i \sim U(0,1)$
↓ experiments

r.v. ≡ experiments

Let $X_1, X_2, \ldots, X_n$ be a random Sample of Size n from Some distribution.

Denote order statistics by

$$X_{(1)} = \min(X_1, X_2, \ldots, X_n)$$

$$\vdots$$

$$X_{(n)} = \max(X_1, X_2, \ldots, X_n)$$

For order statistics, usually easier to begin with the cdf.

cdf for min:

$$F_{X_{(1)}}(x) = P(X_{(1)} \le x) = P(\min(X_1, X_2, \ldots X_n) \le x)$$

$$= 1 - P(x \le \min(X_1, X_2, \ldots X_n))$$

$$= 1 - P(x \le X_1, x \le X_2, \ldots, x \le X_n)$$

$$= 1 - P(x \le X_1)P(x \le X_2) \cdots P(x \le X_n)$$

$$= 1 - (P(x \le X))^n$$

$$= 1 - (1 - P(X \le x))^n$$

$$= 1 - [1 - F(x)]^n$$

$X_i \sim U(0,1) \equiv I_{(0,1)}(x)$ with cdf

$$F(x) = x \quad \text{for} \quad 0 \le x \le 1$$

$$F_{X(n)}(x) = P(X_{(n)} \le x) = P(X_1 \le x, X_2 \le x, \dots, X_n \le x)$$
$$= P(X \le x)^n = [F(x)]^n$$
(by independence)

Beta/exponential part of the family of gamma distributions.

Joint Distribution of min, max.

$$F_{X(1), X(n)}(x,y) = P(X_{(1)} \le x, X_{(n)} \le y).$$

$$P(X_{(n)} \le y) = P(X_{(1)} \le x, X_{(n)} \le y) + P(x \le X_{(1)}, X_{(n)} \le y)$$

Case ①: $x \ge y$, then $P(X_{(1)} \le x, X_{(n)} \le y) = P(X_{(n)} \le y)$

Case ②: $x < y$,

$$P(X_{(1)} \ge x, X_{(n)} \le y) = P(x < X_1 \le y, x < X_2 \le y, \dots x < X_n \le y)$$
$$\overset{iid}{=} [P(x \le X_1 \le y)]^n$$
$$= [F(y) - F(x)]^n$$

So $F_{X(1), X(n)}(x,y)$
$$= [F(y)]^n - [F(y) - F(x)]^n$$

$f_{X(1), X(n)}(x,y) \longrightarrow$ review later

holds for $x < y$ and for x and y both in the support of the original dist.

A heuristic: min and max are continuous and the joint pdf does not represent probability. Heuristic approach

$$f_{X_{(1)}, X_{(n)}}(x,y) = P(X_{(1)} = x, X_{(n)} = y)$$

$$= f(x)\, f(y) \left[F[y] - F[x]\right]^{n-2}$$

After accounting for diff. cases, the "probability" of getting a minimum of x and a maximum of y is

$$n(n-1) f(x) \left[F(y) - F(x)\right]^{n-2} f(y)$$

5. Joint Distribution for all the other order statistics

Find $f_{X_{(1)}, X_{(2)}, \dots, X_{(n)}}(x_1, x_2, \dots, x_n)$

$$f_{X_{(1)}, X_{(2)}, X_{(3)}}(x_1, x_2, x_3) = P(X_{(1)} = x_1, X_{(2)} = x_2, X_{(3)} = x_3)$$

Fix $x_1 < x_2 < x_3$.

So $f_{X_{(1)}, X_{(2)}, X_{(3)}}(x_1, x_2, x_3) = 3!\, f(x_1) f(x_2) f(x_3)$

Joint CDF:

$$P(X_{(1)} \leq x_1, X_{(2)} \leq x_2, \dots, X_{(n)} \leq x_n),$$

$$P(y_1 \leq X_{(1)} \leq x_1, y_2 \leq X_{(2)} \leq x_2, \dots, y_n < X_{(n)} < x_n)$$

for $y_1 < \dots$

$$= n!\, P(y_1 < X_1 \leq x_1, \dots, y_n < X_n < x_n)$$

$$= \prod_{i=1}^{n} P(y_i < X_i < x_i) = \prod_{i=1}^{n} \left[F(x_i) - F(y_i)\right]$$

6 Distribution of $X_{(i)}$

$$f_{X_{(2)}, \ldots, X_{(n)}}(x_2, \ldots, x_n) = \int_{-\infty}^{x_2} f_{X_{(1)}, X_{(2)}, \ldots, X_{(n)}}(x_1, x_2, \ldots, x_n)\,dx_1$$

$$= \int_{-\infty}^{x_2} n!\, f(x_1) f(x_2) \ldots f(x_n)\,dx_1$$

$$= n!\, f(x_2) \ldots f(x_n) F(x_2)$$

Continue until reach $X_{(i)}$:

$$f_{X_{(i)}, \ldots, X_{(n)}}(x_i, \ldots, x_n) = \frac{n!}{(i-1)!} f(x_i) \ldots f(x_n) \left[F(x_i)\right]^{i-1}$$

$$f_{X_{(i)}} = \frac{n!}{(n-i)!\,(i-1)!} \left[F(x_i)\right]^{i-1} f(x_i) \left[1 - F(x_i)\right]^{n-i}$$