Extra OHs: Thurs: 2 - 3pm
Monday: 2 - 3pm
Tues: 3 - 4 pm

Defn: Random Sample

If $X_1, X_2, \dots, X_N$ are mutually independent RVs all with marginal dists f, then we say that the Xs are a random sample from f.

sample size

marginal dist.

Alt: $X_n \overset{iid}{\sim} f$.

Notation:

random vector / mv. RV.

$$\underset{\sim}{X} = (X_1, X_2, \dots, X_N)$$

$$\underset{\sim}{x} = (x_1, \dots, x_N) \in \mathbb{R}^N$$

Joint dist. of a RS (rand. sample)

$$f(\underset{\sim}{x}) = f(x_1, x_2, x_3, \dots, x_N)$$

$$= f(x_1) f(x_2) \cdots f(x_N) \quad \text{[by independence]}$$

$$= \prod_{n=1}^{N} f(x_n)$$

Ex. Assume $X_n \overset{iid}{\sim} Exp(\lambda)$

$Exp(\lambda)$

$$f(x) = \lambda e^{-\lambda x} \text{ for } x > 0$$

more explicit

$$f(x) = \begin{cases} \lambda e^{-\lambda x}, & x > 0 \\ 0, & x \le 0 \end{cases}$$

indicator function

$$\mathbb{1}(\text{statement}) = \begin{cases} 1 & \text{statement true} \\ 0 & \text{statement false} \end{cases}$$

more compact

$$f(x) = \lambda e^{-\lambda x} \mathbb{1}(x > 0)$$

What is the joint dist of my RS?

$$f(\underset{\sim}{x}) = \prod_{n=1}^{N} f(x_n)$$

$$= \prod_{n=1}^{N} \lambda e^{-\lambda x_n} \mathbb{1}(x_n > 0)$$

$$= \lambda^N e^{-\lambda \sum_n x_n} \prod \mathbb{1}(x_n > 0)$$

$$\boxed{\begin{aligned} e^a e^b &= e^{a+b} \\ \prod_n e^{a_n} &= e^{\sum_n a_n} \end{aligned}}$$

$$\mathbb{1}(A)\mathbb{1}(B) = \mathbb{1}(A \text{ and } B)$$

$$= \lambda^N e^{-\lambda \sum_n x_n} \underbrace{\prod_n \mathbb{1}(x_n > 0)}$$

$$= \lambda^N e^{-\lambda \sum_n x_n} \mathbb{1}(\text{all } x_n > 0)$$

$= \mathbb{1}(A \text{ and } B)$

$\prod_n \mathbb{1}(A_n) = \mathbb{1}(\text{all } A_n)$

Defn:

Given a RS $X_n \overset{iid}{\sim} f$ and a function

$$T : \mathbb{R}^N \to \mathbb{R}^d$$ ← typically $d \ll N$ e.g. $d = 1$

then $T(\underset{\sim}{X})$ is a statistic.

---

ex.

① Arithmetic Mean: $(d=1)$

$$T(\underset{\sim}{X}) = \frac{1}{N} \sum_{n=1}^{N} X_n = \overline{X}_N$$

② Sample Variance:

$$S_{N-1}^2 = \frac{1}{N-1} \sum_{n=1}^{N} (X_n - \overline{X}_N)^2$$

③ Sample SD:

$$S_{N-1} = \sqrt{S_{N-1}^2}$$

④ Minimum: $X_{(1)} = \min\{X_1, \dots, X_N\}$

⑤ Maximum: $X_{(N)} = \max\{X_1, \dots, X_N\}$

⑥ Range: $X_{(N)} - X_{(1)}$

⑦ Order Statistic: $X_{(r)} = r^{th}$ smallest value among $X_1, \dots, X_N$

Defn: Sampling distribution

The sampling dist. of a Stat. T is just the dist of T.

Ex. If $X_n \overset{iid}{\sim} f$ (cts), what is the dist of $X_{(1)}$?

I want the PDF of the minimum $X_{(1)}$.

$$\mathbb{P}(X_{(1)} \geq t) = \mathbb{P}(X_1 \geq t, X_2 \geq t, \dots, X_N \geq t)$$

$$= \mathbb{P}(X_1 \geq t)\mathbb{P}(X_2 \geq t) \cdots \mathbb{P}(X_N \geq t)$$

[by independence]

$$= \prod_{n=1}^{N} \mathbb{P}(X_n \geq t)$$

$$= \mathbb{P}(X_n \geq t)^N$$

[by identical dist]

F = CDF of X_n

$$= \underline{(1 - F(t))^N}$$

So

$$F_{X_{(1)}}(t) = \mathbb{P}(X_{(1)} \leq t)$$

$$= 1 - \mathbb{P}(X_{(1)} \geq t)$$

$$= 1 - (1 - F(t))^N$$

$$\boxed{f_{X_{(1)}}(t) = \frac{d}{dt} F_{X_{(1)}} = N(1 - F(t))^{N-1} f(t)}$$

Can play a similar game for $X_{(N)}$:

look at $P(X_{(N)} \leq t)$

and get

$$f \qquad \qquad ^{N-1}$$

that you get

$$\boxed{f_{X_{(N)}}(t) = N\,F(t)^{N-1} f(t)}$$

can generalize to $X_{(r)}$

$$\boxed{f_{X_{(r)}}(t) = \frac{N!}{(r-1)!\,(N-r)!} F(t)^{r-1} (1-F(t))^{N-r} f(t)}$$

Ex. $X_n \overset{iid}{\sim} \mathrm{Exp}(\lambda)$

what is the dist of $X_{(1)}$?

$F(x) = 1 - e^{-\lambda x}$

CDF of $\mathrm{Exp}(\lambda)$

$$f_{X_{(1)}}(t) = N(1-F(t))^{N-1} f(t)$$

$$= N(1-(1-e^{-\lambda x}))^{N-1} \lambda e^{-\lambda x}$$

$$= N\lambda (e^{-\lambda x})^{N-1} e^{-\lambda x}$$

$$f_{X_{(1)}}(t) = (N\lambda) e^{-(N\lambda)x}$$

$$\boxed{X_{(1)} \sim \mathrm{Exp}(N\lambda)}$$