Defn: Random Sample

RVs

sample size

If $X_1, X_2, X_3, \ldots, X_N$ are mutually independent

all w/ marginal dist f marginal PMF/PDF

then we say these Xs are a random sample from f.

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

indep. and ident. dist.

Notation:

$\underset{\sim}{X} = (X_1, \ldots, X_N)$ random. vect. or multivariate RV

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

Joint dist. of a RS (rand. sample)

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

$= f(x_1) f(x_2) \cdots f(x_N)$ (by independence)

joint dist.

N

$= f(x_1)f(x_2)\cdots f(x_N)$

joint dist.

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

$\underline{\text{Ex}_1}$ If $X_n \overset{iid}{\sim} \text{Exp}(\lambda)$ what's the joint?

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

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

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

$$= \lambda^N e^{-\lambda \sum_n x_n}$$

$$f(x) = \begin{cases} \lambda e^{-\lambda x} & \text{for } x > 0 \\ 0 & \text{for } x \leq 0 \end{cases}$$

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

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

$f(x) = \text{~~~} \; \mathbb{1}(x \in \text{support})$

$$e^a e^b = e^{a+b}$$

$\mathbb{1}(\text{all } x_n > 0)$

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

Defn: Statistic

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

and a function

sample size

$$T : \mathbb{R}^{d} \to \mathbb{R}^{q}$$ typically $d \ll N$ (typically $d=1$)

typically

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

<u>Ex.</u> <u>Arithmetic mean:</u> $(d=1)$

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

<u>Sample Variance:</u>

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

<u>Sample SD:</u>

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

<u>Minimum:</u> $X_{(1)} = \min_n \{X_1, \dots, X_N\}$

<u>Max.:</u> $X_{(N)} = \max_n X_n$

<u>Range:</u> $X_{(N)} - X_{(1)}$

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

Defn: Sampling Distribution

The sampling dist. of a stat. $T = T(\underset{\sim}{X})$ is just the dist. of T.

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

Let $X_n \overset{iid}{\sim} f$ (where f cts)

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

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

$$= \prod_{n=1}^{N} P(X_n \geq t) \quad [\text{by independence}]$$

$$= P(X_n \geq t)^N$$

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

CDF of X_n

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

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

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

Play Similar game for $X_{(N)}$:

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

and get

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

General formula for order stats:

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

Famous result from Intro Stat.

$$X_n \overset{iid}{\sim} N(\mu, \sigma^2) \text{ then } \bar{X}_N \sim N(\mu, \sigma^2/N)$$

put on hold

Facts: Sums of RVs

Let $g: \mathbb{R} \to \mathbb{R}$ and $X_n \overset{iid}{\sim} f$

① $\mathbb{E}\left[\sum_{n=1}^{N} g(X_n)\right] = N\,\mathbb{E}[g(X_n)]$ any of them

pf.

$$\mathbb{E}\left[\sum_n g(X_n)\right] = \sum_n \mathbb{E}[g(X_n)]$$ all the same

$$= N\,\mathbb{E}[g(X_1)]$$

② $\mathrm{Var}\left(\sum_n g(X_n)\right) = N\,\mathrm{Var}(g(X_n))$

pf. basically same as above (NEEDS INDEPENDENCE)

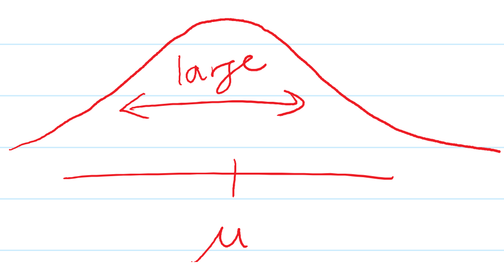
Theorem: If $X_n \overset{iid}{\sim} f$ and

$$\mathbb{E}X_n = \mu \quad \text{and} \quad \mathrm{Var}\,X_n = \sigma^2$$

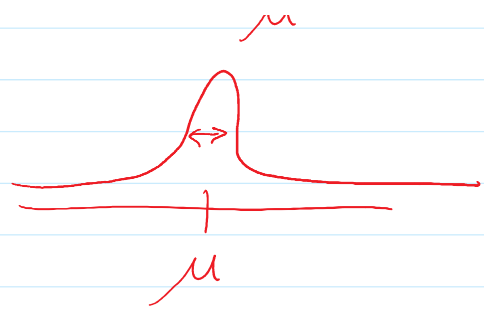
then ① $\mathbb{E}[\bar{X}_N] = \mu$

② $\mathrm{Var}(\bar{X}_N) = \sigma^2/N$

Small N



large N



(3) $\mathbb{E}[S_{N-1}^2] = \sigma^2$

Pf.

(1)

$$\mathbb{E}[\overline{X}_N] = \mathbb{E}\left[\frac{1}{N}\sum_{n=1}^{N} X_n\right] = \frac{1}{N}\mathbb{E}\left[\sum_n X_n\right] = \frac{1}{N} N\mu$$

$$= \mu$$

(2) $$\mathrm{Var}(\overline{X}_N) = \mathrm{Var}\left(\frac{1}{N}\sum_n X_n\right)$$

$$= \frac{1}{N^2}\mathrm{Var}\left(\sum_n X_n\right)$$

$$= \frac{1}{N^2} N\sigma^2 = \sigma^2/N.$$

(3) $$\mathbb{E}[S_{N-1}^2]$$

$$= \mathbb{E}\left[\frac{1}{N-1}\sum_n (X_n - \overline{X})^2\right]$$

$$= \frac{1}{N-1}\mathbb{E}\left[\qquad\qquad\right]$$

$$\sum_n (X_n - \overline{X})^2 = \sum_n X_n^2 - N\overline{X}_N^2$$

recall: $\mathrm{Var}(X) = \mathbb{E}[X^2] - (\mathbb{E}X)^2$

$$\downarrow$$

$$\mathbb{E}[X^2] = \mathrm{Var}(X) + \mathbb{E}[X]^2$$

$$= \frac{1}{N-1} E\Big[$$

$$= \frac{1}{N-1} E\left[\sum_n X_n^2 - N \bar{X}_N^2\right]$$

$$= \frac{1}{N-1}\left(\sum_n E[X_n^2] - N\, E[\bar{X}_N^2]\right)$$

$$\sigma^2 + \mu^2 \qquad\qquad \sigma^2/N + \mu^2$$

$$\overset{\text{algebra}}{=} \cdots = \sigma^2.$$