

## Recap:

- ① Order statistic: its marginal distribution, joint dist. of two order statistics.
- ② Convergence concepts for ~~a~~ a sequence of random variables.

### Convergence in Probability:

A sequence of random variables  $\{X_n\}_{n \geq 1}$  is said to converge in prob. to a r.v.  $X$  if for every  $\epsilon > 0$   $\lim_{n \rightarrow \infty} P(|X_n - X| > \epsilon) = 0$

Ex:  $X_n \sim N(0, \frac{1}{n}) \quad \forall n \geq 1$

$$P(|X_n| > \epsilon) \leq \frac{E[X_n^2]}{\epsilon^2} = \frac{1}{n\epsilon^2} \rightarrow 0 \quad \text{as } n \rightarrow \infty$$

that  $X = 0$  w.p. 1.

### Convergence in distribution

A sequence of random variables  $X_1, \dots, X_n, \dots$  is said to converge to a random variable  $X$  if  $\lim_{n \rightarrow \infty} F_{X_n}(x) = F_X(x)$ , at all points  $x$  where  $F_X(x)$  is cont. Here  $F_{X_n}(x) = P(X_n \leq x) = \text{C.D.F of } X_n$  and  $F_X(x) = P(X \leq x)$  i.e. C.D.F of  $X$ .

Ex:  $X_1, \dots, X_n, \dots \stackrel{iid}{\sim} U(0, 1)$

Ex:  $U_1, \dots, U_n, \dots \stackrel{iid}{\sim} U(0, 1)$

Define,  $X_n = n(1 - U_{(n)})$  where  $U_{(n)} = \max_{1 \leq i \leq n} \{U_i\}$

Qn: where does the sequence  $X_1, \dots$  converges to?

$$\begin{aligned}
 P(X_n \leq x) &= P(n(1 - U_{(n)}) \leq x) \\
 &= P\left(1 - U_{(n)} \leq \frac{x}{n}\right) = P\left(U_{(n)} \geq 1 - \frac{x}{n}\right) \\
 &= 1 - P\left(U_{(n)} < 1 - \frac{x}{n}\right) \\
 &= 1 - P\left(\max_{1 \leq i \leq n} U_i < 1 - \frac{x}{n}\right) = 1 - P\left(U_1 < 1 - \frac{x}{n}, \dots, U_n < 1 - \frac{x}{n}\right) \\
 &= 1 - P\left(U_1 < 1 - \frac{x}{n}\right) P\left(U_2 < 1 - \frac{x}{n}\right) \cdots P\left(U_n < 1 - \frac{x}{n}\right) \\
 &= 1 - \left(1 - \frac{x}{n}\right)^n
 \end{aligned}$$

$$\lim_{n \rightarrow \infty} P(X_n \leq x) = 1 - \lim_{n \rightarrow \infty} \left(1 - \frac{x}{n}\right)^n = 1 - e^{-x}$$

$\lim_{n \rightarrow \infty} F_{X_n}(x) = 1 - e^{-x} = P(X \leq x) = F_X(x)$  where  $X \sim \text{Exp}(1)$

thus by  $\circ$  the definition,  $X_1, \dots, X_n, \dots$  converges  
in distribution to  $X$  where  $X \sim \text{Exp}(1)$ .

An important fact:

Note that when  $X_1, \dots, X_n, \dots$  converges in distribution to  $X$ , we denote it mathematically by  $X_n \xrightarrow{d} X$ .

An important fact:

For any sequence  $X_1, \dots, X_n, \dots$  of random variables  $X_n \xrightarrow{P} X \Rightarrow X_n \xrightarrow{d} X$ .

But  $X_n \xrightarrow{d} X$  does not necessarily imply that  $X_n \xrightarrow{P} X$ .

Suppose  $x_1, x_2, \dots, x_n, \dots$  is a sequence of random variables defined by

$$P(x_n = 0) = P(x_n = 1) = \frac{1}{2}$$

then trivially  $x_n \xrightarrow{d} x$  where  $x \sim \text{Ber}\left(\frac{1}{2}\right)$ .

where  $x \sim \text{Ber}\left(\frac{1}{2}\right)$ ,  $1-x \sim \text{Ber}\left(\frac{1}{2}\right)$

thus  $x_n \xrightarrow{d} 1-x$ .

$$P(|x_n - (1-x)| > \varepsilon) = P(|x - (1-x)| > \varepsilon) = P(|2x - 1| > \varepsilon)$$

$$\text{let, } \varepsilon = \frac{1}{2}$$

$$P(|2x - 1| > \frac{1}{2}) = 1 \Rightarrow \text{that } \lim_{n \rightarrow \infty} P(|x_n - (1-x)| > \frac{1}{2}) \neq 0$$

$$\Rightarrow x_n \not\xrightarrow{P} x.$$

Another interesting fact.

In the definition of convergence in distribution,  
~~if~~  $\lim_{n \rightarrow \infty} F_{X_n}(x) = F_X(x)$  for all  $x$  where

$F_X(x)$  is cont.

$$\text{let, } x_n = \frac{1}{n} \quad \forall n \geq 1$$

$$\Rightarrow P\left(x_n = \frac{1}{n}\right) = 1$$

$$F_{X_n}(x) = \begin{cases} 0 & \text{if } x < \frac{1}{n} \\ 1 & \text{if } x \geq \frac{1}{n} \end{cases}$$

$$\lim_{n \rightarrow \infty} F_{X_n}(x) = \begin{cases} 0 & \text{if } x \leq 0 \\ 1 & \text{if } x > 0 \end{cases} \quad (\text{by the definition of limits of sets})$$

③

④ intuitively  ~~$\{X_n\}_{n \geq 1}$~~  sequence should converge in dist. to a r.v.  $X$  which is  $P(X=0)=1$ .

$$F_X(x) = \begin{cases} 0 & \text{if } x < 0 \\ 1 & \text{if } x \geq 0 \end{cases}$$

Note that  $\lim_{n \rightarrow \infty} F_{X_n}(x) \neq F_X(x)$  at  $x=0$

and also note that  $x=0$  is the discontinuity point of  $F_X$ .

thus the convergence in distribution definition does not include all those  $x$ 's which are discontinuity points of  $F_X$ .

Fact:  $x_n \xrightarrow{d} x$  and  $y_n \xrightarrow{d} y$  then

$x_n + y_n \not\xrightarrow{d} x+y$  in general.

$$\cancel{x_n \sim N(0,1)} \quad x_n \stackrel{\text{iid}}{\sim} N(0,1) \quad y_n = -x_n$$

$$x_n + y_n = 0 \quad \cancel{x_n \stackrel{\text{iid}}{\sim} N(0,1)} \Rightarrow x \sim N(0,1)$$

$$\text{and } y_n = -x_n \sim N(0,1) \Rightarrow y \sim N(0,1)$$

$$x+y \sim N(0,2) \Rightarrow x_n + y_n \not\xrightarrow{d} x+y.$$

Slutsky Theorem:

If  $x_n \xrightarrow{d} x$  and  $y_n \xrightarrow{P} a$ , where  $a$  is a constant

then (a)  $y_n x_n \xrightarrow{d} ax$

(b)  $x_n + y_n \xrightarrow{d} x+a$

Most important application of these two types of convergence!

### weak law of large numbers

Let  $x_1, \dots, x_n, \dots$  be iid random variables with  $E[x_i] = \mu$  and  $\text{Var}(x_i) = \sigma^2 < \infty$ . Define  $\bar{x}_n = \frac{1}{n} \sum_{i=1}^n x_i$ . Then, for every  $\epsilon > 0$ ,  $\lim_{n \rightarrow \infty} P(|\bar{x}_n - \mu| > \epsilon) = 0$  i.e.  $\bar{x}_n \xrightarrow{P} \mu$ .

### central limit theorem

Let  $x_1, \dots, x_n, \dots$  be a sequence of iid random variables whose moment generating function exists in a neighborhood of 0. Let  $E[x_i] = \mu$  and  $\text{Var}(x_i) = \sigma^2 > 0$ . Define  $\bar{x}_n = \frac{1}{n} \sum_{i=1}^n x_i$ . Let  $G_n(x)$  be the C.D.F.

of  $\sqrt{n}(\bar{x}_n - \mu)$ . Then for any  $-\infty < x < \infty$ ,

$$\lim_{n \rightarrow \infty} G_n(x) = \int_{-\infty}^x \frac{1}{\sqrt{2\pi}} e^{-y^2/2} dy$$

$$\sqrt{n}(\bar{x}_n - \mu) \xrightarrow{d} N(0, 1)$$

## Delta theorem :

Let  $y_n$  be a sequence of random variables such that  $\sqrt{n}(y_n - \theta) \xrightarrow{d} N(0, \sigma^2)$ . For a given function  $g$  and a specific value of  $\theta$ , suppose  $g'(\theta)$  exists and  $g'(\theta) \neq 0$  ( $\frac{d}{dx} g(x)|_{x=\theta}$ ).

Then

$$\sqrt{n}(g(y_n) - g(\theta)) \xrightarrow{d} N(0, \sigma^2 [g'(\theta)]^2)$$

Significance:  $x_1, \dots, x_n$  are i.i.d. and

$$E[x_i] = \mu, \quad \text{Var}(x_i) = \sigma^2 > 0$$

$$\sqrt{n}(\bar{x}_n - \mu) \xrightarrow{d} N(0, 1) \quad (\text{by Central limit theorem (CLT)})$$

$$\Rightarrow \sqrt{n}(\bar{x}_n - \mu) \xrightarrow{d} N(0, \sigma^2)$$

Question:  $\sqrt{n}(\bar{x}_n^2 - \mu^2) \xrightarrow{d} ?$

$$g(x) = x^2 \quad \text{lets assume } \mu \neq 0 \Rightarrow g'(\mu) = 2\mu \neq 0$$

$$\sqrt{n}(\bar{x}_n^2 - \mu^2) \xrightarrow{d} N(0, \sigma^2(2\mu)^2) = N(0, 4\mu^2 \sigma^2)$$

What happens when  $g'(\theta) = 0$

when  $g'(\theta) = 0$  but  $g''(\theta) \neq 0$  then

$$n[g(y_n) - g(\theta)] \xrightarrow{d} \frac{\sigma^2}{2} g''(\theta) \tilde{x}_1$$

## Recap:

- ① Order statistic
- ② Convergence properties: convergence in dist. and convergence in prob.
- ③ Properties on results regarding these two modes of convergence.
- ④ Slutsky's theorem, Weak law of large number, CLT, Delta theorem.

## Statistical Inference tool

$$x_1, \dots, x_n \stackrel{\text{iid}}{\sim} f_\theta(x) \quad x_1, \dots, x_n \stackrel{\text{iid}}{\sim} N(\mu, 1)$$

Goal: When we reduce ~~the~~ data with  $n$  sample  $x_1, \dots, x_n$ , we ~~want~~ are trying to understand what type of data reduction technique will not lose any information on  $\theta$ .

## Sufficiency:

Let  $\underline{x} = (x_1, \dots, x_n)$  and  $\underline{x} \sim F_\theta(\underline{x})$ . Then  $T(\underline{x})$  is known to be a sufficient statistic for  $\theta$  if the conditional distribution  $\underline{x} | T(\underline{x})$  is independent of  $\theta$ . Intuitively it means that  $T(\underline{x})$  contains the same information on  $\theta$  that  $\underline{x}$  contains. There is no "additional information" required to make inference on  $\theta$ .

Example:  $x_1, x_2, x_3 \stackrel{iid}{\sim} \text{Ber}(p)$

$$f_p(x) = p^x (1-p)^{1-x}, x=0, 1$$

claim:  $T(\underline{x}) = x_1 + x_2 + x_3$  is a sufficient statistic for  $p$ .

Pf:  $P(x_1 = x_1, x_2 = x_2, x_3 = x_3 \mid T(\underline{x}) = t) = 0$  if  $t \neq x_1 + x_2 + x_3$

if  $t = x_1 + x_2 + x_3$

$$= \frac{P(x_1 = x_1, \cancel{x_2 = x_2}, x_3 = x_3, \sum_{i=1}^3 x_i = t)}{P(\sum_{i=1}^3 x_i = t)}$$

$$= \frac{P(x_1 = x_1, x_2 = x_2, x_3 = x_3)}{P(\sum_{i=1}^3 x_i = t)}$$

$$= \frac{P(x_1 = x_1) P(x_2 = x_2) P(x_3 = x_3)}{P(\sum_{i=1}^3 x_i = t)}$$

$$= \frac{p^{x_1} (1-p)^{1-x_1} p^{x_2} (1-p)^{1-x_2} p^{x_3} (1-p)^{1-x_3}}{\binom{3}{t} p^t (1-p)^{3-t}} \quad \left[ \text{as } \sum_{i=1}^3 x_i \sim \text{Bin}(3, p) \right]$$

$$= \frac{p^{x_1+x_2+x_3} (1-p)^{3-t}}{\binom{3}{t} p^t (1-p)^{3-t}}$$

$$= \frac{p^t (1-p)^{3-t}}{\binom{3}{t} p^t (1-p)^{3-t}} = \frac{1}{\binom{3}{t}}$$

Hence by the definition  $T(\underline{x}) = x_1 + x_2 + x_3$  is a sufficient statistic

$x_1, x_2, x_3 \stackrel{iid}{\sim} \text{Ber}(p)$

cases	prob.	
0 0 0	$(1-p)^3$	$\sum x_i = 0$
0 0 1	$p(1-p)^2$	$\sum x_i = 1$
0 1 0	$p^2(1-p)$	
1 0 0	$p^2(1-p)$	
0 1 1	$p^2(1-p)$	$\sum x_i = 2$
1 0 1	$p^2(1-p)$	
1 1 0	$p^2(1-p)$	
1 1 1	$p^3$	$\sum x_i = 3$

To write the likelihood you do not need to know the entire data. If you knew  $\sum_{i=1}^3 x_i$  then you will be able to write the likelihood. Thus  $\sum_{i=1}^3 x_i$  contains all information about  $p$ .

Question: How to find ~~the~~ sufficient statistics from the distribution?

Factorization theorem:

Let  $\underline{x}$  have a joint p.d.f. (or p.m.f.)  $f_\theta(\underline{x})$ , where  $\theta$  is the unknown parameter. A statistic  $T(\underline{x})$  is sufficient for  $\theta$  if and only if  $f_\theta(\underline{x}) = g(T(\underline{x}), \theta) h(\underline{x})$ , where  $h(\underline{x})$  is a function of  $\underline{x}$  not dependent on  $\theta$ .

examples: ①  $x_1, \dots, x_n \stackrel{iid}{\sim} \text{Ber}(p)$

$$f_p(\underline{x}) = \prod_{i=1}^n p^{x_i} (1-p)^{1-x_i} = p^{\sum_{i=1}^n x_i} (1-p)^{n - \sum_{i=1}^n x_i}$$

$$= \left( \frac{p}{1-p} \right)^{\sum_{i=1}^n x_i} (1-p)^n$$

②  $g\left(\sum_{i=1}^n x_i, p\right) = \left( \frac{p}{1-p} \right)^{\sum_{i=1}^n x_i} (1-p)^n, R(\underline{x}) = 1$

thus by factorization theorem,  $T(\underline{x}) = \sum_{i=1}^n x_i$

③ Suppose  $x_1, \dots, x_n \stackrel{iid}{\sim} \text{Pois}(\lambda)$

$$f_\lambda(\underline{x}) = \prod_{i=1}^n \frac{e^{-\lambda} \lambda^{x_i}}{x_i!} = \frac{e^{-n\lambda} \lambda^{\sum_{i=1}^n x_i}}{\prod_{i=1}^n x_i!}$$

$$g\left(\sum_{i=1}^n x_i, \lambda\right) = e^{-n\lambda} \lambda^{\sum_{i=1}^n x_i}, R(\underline{x}) = \frac{1}{\prod_{i=1}^n x_i!}$$

by factorization theorem  $T(\underline{x}) = \sum_{i=1}^n x_i$ .

④ Suppose  $x_1, \dots, x_n \stackrel{iid}{\sim} N(\mu, \sigma^2)$ ,  $\sigma^2$  is known.

$$f_\mu(\underline{x}) = \prod_{i=1}^n \left\{ \frac{1}{\sqrt{2\pi\sigma^2}} \exp \left\{ -\frac{(x_i - \mu)^2}{2\sigma^2} \right\} \right\}$$

$$= \left( \frac{1}{\sqrt{2\pi\sigma^2}} \right)^n \exp \left\{ -\sum_{i=1}^n \frac{(x_i - \mu)^2}{2\sigma^2} \right\}$$

$$= \left( \frac{1}{\sqrt{2\pi\sigma^2}} \right)^n \exp \left\{ -\frac{1}{2\sigma^2} \left[ \sum_{i=1}^n x_i^2 - 2\mu \sum_{i=1}^n x_i + n\mu^2 \right] \right\}$$

$$= \left( \frac{1}{\sqrt{2\pi\sigma^2}} \right)^n \exp \left\{ -\frac{\sum_{i=1}^n x_i^2}{2\sigma^2} \right\} \exp \left\{ -\frac{2\mu}{2\sigma^2} \sum_{i=1}^n x_i \right\} \exp \left\{ -\frac{n\mu^2}{2\sigma^2} \right\}$$

④

$$g\left(\sum_{i=1}^n x_i, \mu\right) = \exp\left\{ \frac{\mu}{\sigma} \sum_{i=1}^n x_i \right\} \exp\left\{ -\frac{n\mu^2}{2\sigma^2} \right\}$$

$$h(\underline{x}) = \left( \frac{1}{\sqrt{2\pi\sigma^2}} \right)^n \exp\left\{ -\frac{\sum_{i=1}^n x_i^2}{2\sigma^2} \right\}$$

By factorization theorem  $T(\underline{x}) = \sum_{i=1}^n x_i$

(4)  $x_1, \dots, x_n \stackrel{iid}{\sim} N(\mu, \sigma^2)$  both  $\mu, \sigma^2$  unknown.

$$f_{\mu, \sigma^2}(\underline{x}) = \left( \frac{1}{\sqrt{2\pi\sigma^2}} \right)^n \exp\left\{ -\frac{\sum_{i=1}^n x_i^2}{2\sigma^2} \right\} \exp\left\{ -\frac{n\mu^2}{2\sigma^2} \right\}$$

$$g\left(\sum_{i=1}^n x_i^2, \sum_{i=1}^n x_i, \mu, \sigma^2\right) = \left( \frac{1}{\sqrt{2\pi\sigma^2}} \right)^n \exp\left\{ -\frac{\sum_{i=1}^n x_i^2}{2\sigma^2} \right\} \exp\left\{ \frac{\mu}{\sigma^2} \sum_{i=1}^n x_i \right\}$$

$$\exp\left\{ -\frac{n\mu^2}{2\sigma^2} \right\}$$

$$h(\underline{x}) = 1$$

$$T(\underline{x}) = \left( \sum_{i=1}^n x_i, \sum_{i=1}^n x_i \right)$$

(5)  $x_1, \dots, x_n \stackrel{iid}{\sim} U(\theta, \theta+1)$

$$f_\theta(\underline{x}) = 1 \quad \text{if } \theta < x_1 < \theta+1, \dots, \theta < x_n < \theta+1$$

$$= I(\theta < x_1 < \theta+1, \dots, \theta < x_n < \theta+1)$$

$$I(\text{re}) = 1 \quad \text{if } \text{re} \text{ happens}$$

$$= 0 \quad \text{o.w.}$$

$$= I(x_{(1)}, x_{(n)} < \theta+1)$$

$$= I(x_{(n)} - 1 < \theta < x_{(1)})$$

(5)

By factorization theorem  $T(\underline{x}) = (x_{(1)}, x_{(n)})$   
 Here  $\theta$  is one-dimensional, but sufficient statistic is two dimensional.

$$\textcircled{6} \quad x_1, \dots, x_n \stackrel{\text{iid}}{\sim} U(0, \theta)$$

$$\begin{aligned} f_\theta(\underline{x}) &= 1 \quad \text{if } 0 < x_1 < \theta, \dots, \textcircled{6} \quad 0 < x_n < \theta \\ &= I(0 < x_1 < \theta, \dots, 0 < x_n < \theta) \\ &= I(x_{(1)} > 0, x_{(n)} < \theta) = I(x_{(1)} > 0) I(x_{(n)} < \theta) \end{aligned}$$

thus  $\textcircled{6} T(\underline{x}) = x_{(n)}$  is sufficient statistic.

$$\textcircled{7} \quad x_1, \dots, x_n \stackrel{\text{iid}}{\sim} U(\theta_1, \theta_2)$$

$$\begin{aligned} f_{\theta_1, \theta_2}(\underline{x}) &= 1 \quad \text{if } \theta_1 < x_1 < \theta_2, \dots, \theta_1 < x_n < \theta_2 \\ &= I(\theta_1 < x_1 < \theta_2, \dots, \theta_1 < x_n < \theta_2) \\ &= I(x_{(1)} > \theta_1, x_{(n)} < \theta_2) = I(x_{(1)} > \theta_1) I(x_{(n)} < \theta_2) \end{aligned}$$

$T(\underline{x}) = (x_{(1)}, x_{(n)})$  is jointly sufficient for  $(\theta_1, \theta_2)$ .