Recap:
1. Order statistic: its marginal distribution, joint dist. of two order statistics.
2. Convergence concepts for 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 \( \varepsilon > 0 \), 
\[
\lim_{n \to \infty} P(|X_n - X| > \varepsilon) = 0
\]

Ex: \( X_n \sim N(0, \frac{1}{n}) \), \( \forall \ n \geq 1 \)
\[
P(|X_n| > \varepsilon) \leq \frac{E[|X_n|]}{\varepsilon} = \frac{1}{n\varepsilon} \to 0 \quad \text{as} \quad n \to \infty
\]
that \( X = 0 \) w.p. 1.

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

EX: \( U_1, \ldots, U_n, \ldots \sim U(0,1) \)
- Define \( X_n = n(1-U(n)) \) where \( U(n) = \max\{U_i\} \), \( 1 \leq i \leq n \)
- \( U(n) \) where does the sequence \( X_1, \ldots \) converges to?
\[
\Pr(X_n \leq x) = \Pr\left(n(1 - U(n)) \leq x \right)
= \Pr\left(1 - U(n) \leq \frac{x}{n} \right) = \Pr\left(U(n) \geq 1 - \frac{x}{n} \right)
= 1 - \Pr(U(n) < 1 - \frac{x}{n})
= 1 - \Pr\left(\max_{1 \leq i \leq n} U_i < 1 - \frac{x}{n} \right) = 1 - \Pr\left(U_1 < 1 - \frac{x}{n}, \ldots, U_n < 1 - \frac{x}{n}\right)
= 1 - \Pr(U_1 < 1 - \frac{x}{n}) \Pr(U_2 < 1 - \frac{x}{n}) \cdots \Pr(U_n < 1 - \frac{x}{n})
= 1 - \left(1 - \frac{x}{n}\right)^n
\]

\[
\lim_{n \to \infty} \Pr(X_n \leq x) = 1 - \lim_{n \to \infty} \left(1 - \frac{x}{n}\right)^n = 1 - e^{-x}
\]

\[
\lim_{n \to \infty} F_{X_n}(x) = 1 - e^{-x} = \Pr(X \leq x) = F_X(x) \text{ where } X \sim \text{Exp}(1)
\]

Thus by the definition, \(X_1, \ldots, X_n, \ldots\) converges in distribution to \(X\) where \(X \sim \text{Exp}(1)\).

**An important fact:**

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

**An important fact:**

- For any sequence \(X_1, \ldots, X_n, \ldots\) 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 \( X_1, X_2, \ldots, X_n, \ldots \) 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 \( X \sim \text{Ber}(\frac{1}{2}) \).

where \( X \sim \text{Ber}(\frac{1}{2}) \), \( 1 - X \sim \text{Ber}(\frac{1}{2}) \)

thus \( X_n \xrightarrow{d} 1 - X \).

\[
P(|X_n - (1 - X)| > \varepsilon) = P(|X - (1 - X)| > \varepsilon) = P(|2X - 1| > \varepsilon)
\]

let, \( \varepsilon = \frac{1}{2} \)

\[
P(|2X - 1| > \frac{1}{2}) = 1 \implies \text{that } \lim_{n \to \infty} P(|X_n - (1 - X)| > \theta_{\frac{1}{2}}) \neq 0
\]

\( \implies X_n \xrightarrow{d} X \)

Another interesting fact:

In the definition of convergence in distribution,

\[
x_n \xrightarrow{d} \lim_{n \to \infty} F_{X_n}(x) = F_x(x) \quad \text{for all } x \text{ where}
\]

\( F_{X}(x) \) in cont.

let, \( X_n = \frac{1}{n} \quad \forall \ n \geq 1 \)

\( \implies P(X_n = \Theta_{\frac{1}{n}}) = 1 \)

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

\( \lim_{n \to \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)}
\]
\( \text{Intuitively, } \{X_n\} \text{ sequence should converge in dist. to a r.v. } X \text{ which is } P(X=0)=1 \)

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

Note that \( \lim_{n \to \infty} F_{X_n}(x) \neq F_X(x) \) 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
\[ X_n + Y_n \xrightarrow{d} X + Y \text{ in general.} \]

\( X_n \sim \mathcal{N}(0) \) \( \implies \) \( Y_n = -X_n \)
\[ X_n + Y_n = 0 \]
\( X_n \sim \mathcal{N}(0,1) \) \( \implies \) \( X_n \sim \mathcal{N}(0,1) \)
\[ X_n \sim \mathcal{N}(0,1) \]
\[ Y_n = -X_n \sim \mathcal{N}(0,1) \]
\[ X + Y \sim \mathcal{N}(0,2) \]
\[ X_n + Y_n \xrightarrow{d} X + Y \]

**Slutzky's Theorem:**
\[ \text{If } X_n \xrightarrow{d} X \text{ and } Y_n \xrightarrow{P} a, \text{ where } a \text{ is a const} \]
then (a) \( Y_nX_n \xrightarrow{d} aX \)
(b) \( X_n + X_n \xrightarrow{d} X + a \)
Most important application of these two types of convergence:

weak law of large numbers

Let $X_1, \ldots, X_n, \ldots$ 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 $\varepsilon > 0$,

$$\lim_{n \to \infty} P( |\bar{X}_n - \mu| > \varepsilon) = 0 \quad \text{i.e.} \quad \bar{X}_n \xrightarrow{P} \mu.$$

central limit theorem:

Let $X_1, \ldots, X_n, \ldots$ be a sequence of iid random variables whose moment generating function exists in a $\theta$ 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 $\sqrt{n} \left( \bar{X}_n - \mu \right)$. Then for any $-\infty < x < \infty$,

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

$$\frac{\sqrt{n} \left( \bar{X}_n - \mu \right)}{\sigma} \xrightarrow{d} N(0,1)$$
Delta theorem:

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

Then

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

Significance: \( X_1, \ldots, X_n \) are i.i.d. and

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

\[
\sqrt{n} \left( \overline{X}_n - \mu \right) \xrightarrow{d} N(0, 1) \quad \text{(by Central Limit theorem (CLT))}
\]

\[
\Rightarrow \sqrt{n} \left( \overline{X}_n - \mu \right) \xrightarrow{d} N(0, \sigma^2)
\]

Question: \( \sqrt{n} \left( \frac{1}{n} \sum_{i=1}^{n} X_i^2 - \mu^2 \right) \xrightarrow{d} ? \)

\[ g(x) = x^2 \] let's assume \( \mu \neq 0 \) \( \Rightarrow g'(\mu) = 2\mu \neq 0 \)

\[
\sqrt{n} \left( \overline{X}_n^2 - \mu^2 \right) \xrightarrow{d} N(0, \sigma^2 (2\mu)^2) = N(0, 4\mu^2 \sigma^2)
\]

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

When \( g'(\theta) \neq 0 \) but \( g''(\theta) \neq 0 \) then

\[
\sqrt{n} \left[ g(X_n) - g(\theta) \right] \xrightarrow{d} \frac{\sigma^2}{2} g''(\theta) X_1
\]
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

\[ X_1, \ldots, X_n \overset{iid}{\sim} f_0(x) \quad \overline{X}_n \overset{iid}{\sim} N(M, 1) \]

Goal: When we reduce data with a sample \( X_1, \ldots, X_n \), we are trying to understand what type of data reduction technique will not lose any information on \( \theta \).

Sufficiency:

Let \( \theta \), \( \mathbf{X} = (X_1, \ldots, X_n) \) and \( \overline{X} \sim F_\theta(x) \). Then \( T(X) \) is known to be a sufficient statistic for \( \theta \) if the conditional distribution \( X | T(X) \) is independent of \( \theta \). Intuitively it means that \( T(X) \) contains the same information on \( \theta \) that \( X \) contains. There is no "additional information" required to make inference on \( \theta \).
Example: \( x_1, x_2, x_3 \overset{iid}{\sim} \text{Ber}(p) \)

\[ P(x) = p^x (1-p)^{1-x}, \ x = 0, 1 \]

Claim: \( T(X) = X_1 + X_2 + X_3 \) is a sufficient statistic for \( p \).

Proof: \( P(x_1 = x_1, x_2 = x_2, x_3 = x_3 \mid T(X) = t) = 0 \) if \( t \neq x_1 + x_2 + x_3 \)

\( \text{if} \ t = x_1 + x_2 + x_3 \)

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

\[ \frac{P(x_1 = x_1, x_2 = x_2, x_3 = x_3)}{P(x_1 + x_2 + x_3 = t)} \]

\[ = P(x_1 = x_1) P(x_2 = x_2) P(x_3 = x_3) \]

\[ \frac{1}{P(x_1 + x_2 + x_3 = t)} \]

\[ = \frac{p^{x_1} (1-p)^{1-x_1} p^{x_2} (1-p)^{1-x_2} p^{x_3} (1-p)^{1-x_3}}{p^{x_1 + x_2 + x_3} (1-p)^{3-x_1-x_2-x_3}} \]

\[ \text{as} \ \sum_{i=1}^{3} x_i \sim \text{Bin}(3, p) \]

\[ \frac{(3)}{t} \frac{p^{t} (1-p)^{3-t}}{p^{x_1 + x_2 + x_3} (1-p)^{3-x_1-x_2-x_3}} \]

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

Hence, by the definition, \( T(X) = X_1 + X_2 + X_3 \) is a sufficient statistic.
$X_1, X_2, X_3 \sim \text{Bern}(p)$

<table>
<thead>
<tr>
<th>Cases</th>
<th>prob.</th>
</tr>
</thead>
<tbody>
<tr>
<td>000</td>
<td>$(1-p)^3$</td>
</tr>
<tr>
<td>001</td>
<td>$p(1-p)^2$</td>
</tr>
<tr>
<td>010</td>
<td>$p(1-p)^2$</td>
</tr>
<tr>
<td>100</td>
<td>$p^2(1-p)$</td>
</tr>
<tr>
<td>011</td>
<td>$p^3(1-p)$</td>
</tr>
<tr>
<td>101</td>
<td>$p^3(1-p)$</td>
</tr>
<tr>
<td>110</td>
<td>$p^3(1-p)$</td>
</tr>
<tr>
<td>111</td>
<td>$p^3$</td>
</tr>
</tbody>
</table>

$\sum x_i = 0$

$\sum x_i = 1$

$\sum x_i = 2$

$\sum x_i = 3$

To write the likelihood you do not need to know the entire data. If you know $\sum_{i=1}^3 x_i$, 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 sufficient statistics from the distribution?

**Factorization theorem:**

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

3
Examples: \( x_1, \ldots, x_n \) iid \( \text{Bin}(p) \)

\[
p(x) = \frac{n!}{i!} p^x (1-p)^{n-x} = p^{\sum_{i=1}^n x_i} (1-p)^{n-\sum_{i=1}^n x_i}
\]

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

2. Suppose \( x_1, \ldots, x_n \) iid \( \text{Pois} (\lambda) \)

\[
p(\lambda) = \frac{\lambda^{\sum_{i=1}^n x_i}}{\sum_{i=1}^n x_i!} = e^{-\lambda} \frac{\lambda^{\sum_{i=1}^n x_i}}{\sum_{i=1}^n x_i!}
\]

\[
g(\sum_{i=1}^n x_i, \lambda) = e^{-\lambda} \frac{\lambda^{\sum_{i=1}^n x_i}}{\sum_{i=1}^n x_i!}, \quad R(x) = \frac{1}{\sum_{i=1}^n x_i!}
\]

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

3. Suppose \( x_1, \ldots, x_n \) iid \( \text{N}(\mu, \nu) \), \( \nu \) in known.

\[
p(\mu) = \frac{1}{\sqrt{2\pi\nu}} \exp\left\{-\frac{(X_1-\mu)^2}{2\nu}\right\}
\]

\[
= \left(\frac{1}{\sqrt{2\pi\nu}}\right)^n \exp\left\{-\sum_{i=1}^n \frac{(X_i-\mu)^2}{2\nu}\right\}
\]

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

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

\[ R(x) = \frac{1}{\left( \sqrt{2\pi\nu} \right)^n} \exp \left\{ -\frac{\sum_{i=1}^{n} x_i^2}{2\nu} \right\} \]

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

4. \( x_1, \ldots, x_n \sim \mathcal{N}(\mu, \nu) \) both \( \mu, \nu \) unknown.

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

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

\[ R(x) = 1 \]

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

5. \( x_1, \ldots, x_n \sim \mathcal{U}(\theta, \theta+1) \)

\[ f_{\theta}(x) = 1 \text{ if } \theta < x_1 < \theta + 1, \ldots, \theta < x_n < \theta + 1 \]

\[ = I \left( \theta < x_1 < \theta + 1, \ldots, \theta < x_n < \theta + 1 \right) \]

\[ I(\theta) = 1 \text{ if } \theta \text{ happens} \]

\[ = 0 \quad \text{otherwise} \]

\[ = I \left( \theta < x_{(1)} \land x_{(n)} < \theta + 1 \right) \]

\[ = I \left( x_{(n)} - 1 < \theta < x_{(1)} \right) \]
By factorization theorem $T(x) = (X_{(1)}, X_{(n)})$

Here $\theta$ is one-dimensional, but sufficient statistic in two dimensional.

\( x_1, \ldots, x_n \overset{iid}{\sim} U(0, \theta) \)

\[
\tilde{f}_0(x) = 1 \quad \text{if} \quad 0 < x_1 < \theta, \ldots, 0 < x_n < \theta \\
= I(0 < x_1 < \theta, \ldots, 0 < x_n < \theta) \\
= I(x_{(1)} > 0, x_{(n)} < \theta) = I(x_{(1)} > 0) I(x_{(n)} < \theta)
\]

Thus $T(x) = X_{(n)}$ is sufficient statistic.

\( x_1, \ldots, x_n \overset{iid}{\sim} U(\theta_1, \theta_2) \)

\[
\tilde{f}_{\theta_1, \theta_2}(x) = 1 \quad \text{if} \quad \theta_1 < x_1 < \theta_2, \ldots, \theta_1 < x_n < \theta_2 \\
= I(\theta_1 < x_1 < \theta_2, \ldots, \theta_1 < x_n < \theta_2) \\
= I(x_{(1)} > \theta_1, x_{(n)} < \theta_2) = I(x_{(1)} > \theta_1) I(x_{(n)} < \theta_2)
\]

$T(x) = (X_{(1)}, X_{(n)})$ in jointly sufficient for $(\theta_1, \theta_2)$. 