Recall :-

Sample $X_1, X_2, \ldots, X_n$ from the population
i.i.d X

Empirical Distribution : Discrete distribution with

p.m.f. :- $$f_n(t) = \frac{|\{i \mid X_i = t \;\; 1 \le i \le n\}|}{n}$$

Weak Law of Large numbers :- If $E[X] = \mu$, $Var(X) = \sigma^2$

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

(Tschebyschev) $\forall \varepsilon > 0$ $\quad \mathbb{P}(|\bar{X} - \mu| > \varepsilon) \longrightarrow 0$ as $n \to \infty$.

Sample Proportion :- Interest $p = \mathbb{P}(X \in A)$

$$\hat{p}_n = \frac{|\{i : X_i \in A, 1 \le i \le n\}|}{n}$$

let
$$Z_i = \begin{cases} 1 & \text{if } X_i \in A \\ 0 & \text{otherwise} \end{cases}$$

$\{Z_i\}_{i=1}^{n}$ are i.i.d. Bernoulli (p) ; $E[Z_i] = p$

$$\hat{p}_n = \frac{\sum_{i=1}^{n} Z_i}{n}$$

- $E[\hat{p}_n] = p$

  $Var[\hat{p}_n] = \dfrac{p(1-p)}{n}$

  Consistent estimator
  unbiased

- weak law of large numbers

  $\mathbb{P}(|\hat{p}_n - p| > \varepsilon) \longrightarrow 0$ as $n \to \infty$

- strong law of large numbers

  $\mathbb{P}(\lim_{n\to\infty} \hat{p}_n = p) = 1$

$\mathbb{P}(\lim_{n\to\infty}$ Relative frequency of Event A $=$ True Probability of A $) = 1$

Sample Mean and Sample Variance

$E[X] = \mu$,

$Var[X] = \sigma^2$

- $X_1, X_2, \ldots, X_n$ be an i.i.d. random sample from a population. — X

  Recall the sample mean

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

Recall

$E[\overline{X}] = \mu$

$Var[\overline{X}] = \frac{\sigma^2}{n}$

  and sample variance

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2.$$

$E[S^2] = \sigma^2$

- What are the limiting behaviour and distributional properties ?

Recall-Joint Distribution

Let $n \geq 1$ and $X_1, X_2, \ldots, X_n$ be random variables defined on the same probability space.

$\hookrightarrow$ discrete or Continuous random variables.

- The joint distribution function $F : \mathbb{R}^n \to [0, 1]$ is given by

$$F(x_1, x_2, \ldots, x_n) = P(X_1 \leq x_1, X_2 \leq x_2, \ldots, X_n \leq x_n),$$

for $x_1, x_2, \ldots, x_n \in \mathbb{R}$.

Example :- $X_i \in (0,1) \quad 1 \leq i \leq 3$

$$\mathbb{P}\left(\sqrt{X_1^2 + X_2^2 + X_3^2} \leq \frac{1}{2} \right) = ?$$

Dependent random variables / independent random variables.

Recall-Joint Distribution- Discrete and Independence

Joint p.m.f

Let $n \geq 1$ and $X_1, X_2, \ldots, X_n$ be random variables defined on the same probability space.

- If all the random variables were discrete, from the joint distribution function one can determine

$$P(X_1 = t_1, X_2 = t_2, \ldots, X_n = t_n),$$

and vice versa

p.m.f $\longleftrightarrow$ Joint distribution function

for all $t_i \in \text{Range}(X_i), 1 \leq i \leq n$.

Independence

- A finite collection $(X_1, X_2, X_3, \ldots, X_n)$ are mutually independent discrete random variables then their joint probability mass function is given by

$$P(X_1 = t_1, X_2 = t_2, \ldots, X_n = t_n) = \prod_{i=1}^{n} P(X_i = t_i)$$

for all $t_i \in \text{Range}(X_i), 1 \leq i \leq n$.

Joint p.m.f.
$\equiv$
Product of
marginal p.m.f.

Recall-Joint Distribution -Continuous and Independence

Joint p.d.f

Let $n \geq 1$ and $X_1, X_2, \ldots, X_n$ be random variables defined on the same probability space.

- If all the random variables were continuous, from the joint distribution function one can determine for every event $A \subset \mathbb{R}^n$

Viceversa

$f \longleftrightarrow F$

$$P((X_1, X_2, X_3, \ldots, X_n) \in A) = \int_A f(x_1, x_2, \ldots, x_n) dx_1 dx_2 \ldots dx_n,$$

where $f : \mathbb{R}^n \to \mathbb{R}$ is the joint density.

Independence

- A finite collection $(X_1, X_2, X_3, \ldots, X_n)$ are mutually independent continuous random variables with marginal densities f_{X_i} then their joint density is given by

$$f(x_1, x_2, \ldots, x_n) = \prod_{i=1}^{n} f_{X_i}(x_i),$$

for $x_i \in \mathbb{R}$ and $1 \leq i \leq n$.

Joint p.d.f is Product of the marginal p.d.f

Order Statistics

$X_1, \ldots X_n$ are i.i.d X
and $F(x) = \mathbb{P}(X \le x)$

Let $n \geq 1$ and let $X_1, X_2, \ldots, X_n$ be a i.i.d random sample from a population. Let F be the common distribution function. Let the X's be arranged in increasing order of magnitude denoted by

2nd lowest value

minimum value

$$X_{(1)} \leq X_{(2)} \leq \cdots \leq X_{(n)}.$$

largest value

These ordered values are called the order statistics of the sample $X_1, X_2, \ldots, X_n$.

- $F_{(1)}(x) = P(X_{(1)} \leq x) =$ $\mathbb{P}(\min_{1\le i\le n} X_i \le x) = 1 - \mathbb{P}(\min_{1\le i\le n} X_i > x)$

$$= 1 - \mathbb{P}(X_i > x, \ 1 \le i \le n)$$

$$= 1 - \prod_{i=1}^{n} \mathbb{P}(X_i > x)$$

$$= 1 - \prod_{i=1}^{n} (1 - \mathbb{P}(X_i \le x)) = 1 - (1 - F(x))^n$$

Order Statistics

Let $n \geq 1$ and let $X_1, X_2, \ldots, X_n$ be a i.i.d random sample from a population. Let F be the common distribution function. Let the X's be arranged in increasing order of magnitude denoted by

$$X_{(1)} \leq X_{(2)} \leq \cdots \leq X_{(n)}.$$

These ordered values are called the order statistics of the sample $X_1, X_2, \ldots, X_n$.

- $F_{(n)}(x) = P(X_{(n)} \leq x) = \mathbb{P}(\max_{1 \leq i \leq n} X_i \leq x) = \prod_{i=1}^{n} \mathbb{P}(X_i \leq x)$

$$= (F(x))^n$$

Order Statistics

Let $n \geq 1$ and let $X_1, X_2, \ldots, X_n$ be a i.i.d random sample from a population. Let F be the common distribution function. Let the X's be arranged in increasing order of magnitude denoted by

$$X_{(1)} \leq X_{(2)} \leq \cdots \leq X_{(n)}.$$

These ordered values are called the order statistics of the sample $X_1, X_2, \ldots, X_n$.

- for $1 < r < n$, $F_{(r)}(x) = P(X_{(r)} \leq x) = \mathbb{P}(\text{at least } r \text{ elements of the sample are } \leq x)$

$$= \sum_{j=r}^{n} \mathbb{P}(\text{exactly } j \text{ elements of the sample are } \leq x)$$

$$= \sum_{j=r}^{n} \binom{n}{j} \mathbb{P}(\text{chosen } j \text{ elements } \leq x)\, \mathbb{P}(\text{rest } n-j \text{ elements } > x)$$

$$= \sum_{j=r}^{n} \binom{n}{j} (F(x))^{j} (1 - F(x))^{n-j}$$

Exercises to try

$X \sim \text{Uniform}(0,1)$

$X \sim \text{Exp}(\lambda)$

- Compute explicitly distribution function of $X_{(r)}$, $1 \le r \le n$

  r^{th} record. of the sample

- $R = X_{(n)} - X_{(1)} \equiv$ Compute distribution function of R

Sample Mean

Let $n \geq 1$ and $(X_1, X_2, X_3, \ldots, X_n)$ be a collection of independent Normal random variables with mean 0 and variance 1. Then the joint density is given by

$$f(x_1, x_2, \ldots x_n) = \quad ?$$

for $x_i \in \mathbb{R}$ and $1 \leq i \leq n$. Let $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$. Find the distribution of $\bar{X}$.

- $f(x_1, \ldots x_n) = \prod_{i=1}^{n} \dfrac{\exp\left(-\frac{x_i^2}{2}\right)}{\sqrt{2\pi}} = \left(\dfrac{1}{\sqrt{2\pi}}\right)^n \exp\left(-\sum_{i=1}^{n} \frac{x_i^2}{2}\right)$

- $\bar{X} = \frac{1}{n}X_1 + \frac{1}{n}X_2 + \ldots + \frac{1}{n}X_n =$ linear combination of independent Normal random variables

$\bar{X} \equiv$ distributed as Normal$\left(\dfrac{0+0\ldots+0}{n}, \dfrac{1}{n^2}\cdot n\right) \stackrel{d}{=} N\left(0, \frac{1}{n}\right)$

- [Central limit Theorem] $\equiv$ $\bar{X} \equiv$ "assymptotically" Normal for general distribution.

χ^2

- arises naturally as a function of i.i.d normal random variables
- function are "sum of squares" $\equiv$ arise naturally in statistics.

Let X_1 be a Normal random variable with mean 0 and variance 1. Let $Z = X_1^2$. Find the distribution of Z.

Step 1 :. Find the distribution function of Z.

$$\mathbb{P}(Z \le z) = \mathbb{P}(X_1^2 \le z) = \begin{cases} 0 & z \le 0 \\ \ldots ? & z > 0 \end{cases}$$

$z > 0$, $\quad \mathbb{P}(Z \le z) = \mathbb{P}(X_1^2 \le z)$

$$\mathbb{P}(Z \le z) = \mathbb{P}(-\sqrt{z} \le X_1 \le \sqrt{z})$$

$$= 2\,\mathbb{P}(0 \le X_1 \le \sqrt{z})$$

$$= 2\int_0^{\sqrt{z}} \frac{e^{-\frac{x^2}{2}}}{\sqrt{2\pi}}\,dx$$

$$\boxed{\begin{aligned} u &= x^2 \\ du &= 2x\,dx \end{aligned}} \longrightarrow = 2\int_0^{z} \frac{e^{-u/2}}{\sqrt{2\pi}} \cdot \frac{1}{2} u^{-1/2}\,du$$

$$= \int_0^{z} \frac{e^{-u/2}\,u^{-1/2}}{\sqrt{2\pi}}\,du$$

Step2 :- Differentiate if possible to get p.d.f.

$$f_Z(z) = F'(z) = \frac{d}{dz}\mathbb{P}(Z \le z)$$

So
$$f_Z(z) = \begin{cases} \dfrac{e^{-z/2}\, z^{-1/2}}{\sqrt{2\pi}} & z > 0 \\ 0 & z \le 0 \end{cases}$$

$$Z \sim \text{Gamma}\left(\frac{1}{2}, \frac{1}{2}\right)$$

χ^2-with 2 degrees of freedom

name / nomenclature will become clear later in the course

Let (X_1, X_2) be a collection of independent Normal random variables with mean 0 and variance 1. Let $Z = X_1^2 + X_2^2$. Find the distribution of Z.

Exercise

{ one can do the same procedure as before

Previous step

$$\boxed{\begin{array}{l} X_1^2 \stackrel{d}{=} \text{Gamma}\left(\frac{1}{2}, \frac{1}{2}\right) \\ X_2^2 \stackrel{d}{=} \text{Gamma}\left(\frac{1}{2}, \frac{1}{2}\right) \\ X_1^2 \text{ and } X_2^2 \text{ are independent} \end{array}}$$

Sum of $X_1^2 + X_2^2$

$$X_1^2 + X_2^2 \stackrel{d}{=} \text{Gamma}\left(\frac{1}{2} + \frac{1}{2}, \frac{1}{2}\right)$$

$$\stackrel{d}{=} \text{Exp}\left(\frac{1}{2}\right)$$

χ^2 - n degrees of freedom

Inductively argue

Let $n \geq 1$ and $(X_1, X_2, X_3, \ldots, X_n)$ be a collection of independent Normal random variables with mean 0 and variance 1. Then the joint density is given by

$$f(x_1, x_2, \ldots x_n) =$$

for $x_i \in \mathbb{R}$ and $1 \leq i \leq n$. Let $Z = \sum_{i=1}^{n} X_i^2$. Find the distribution of Z.

$$\left.\begin{array}{l} X_i^2 \stackrel{d}{=} \text{Gamma}\left(\frac{1}{2}, \frac{1}{2}\right) \\ \{X_i^2\}_{i=1}^{n} \text{ are independent} \end{array}\right\} \equiv X_1^2 + \cdots + X_n^2 \stackrel{d}{=} \text{Gamma}\left(\frac{n}{2}, \frac{1}{2}\right)$$

(Chi-Square with n degrees of freedom)

$\chi^2_n \equiv$ arises naturally as a sum of squares of n independent Normal(0,1) random variables.

A random variable X whose distribution is Gamma $(\frac{n}{2}, \frac{1}{2})$ is said to have Chi-square distribution with n-degrees of freedom (i.e number of parameters). Gamma $(\frac{n}{2}, \frac{1}{2})$ is denoted by χ^2_n and as discussed earlier it has density given by

$$f(x) = \frac{2^{-\frac{n}{2}}}{\Gamma(\frac{n}{2})} x^{\frac{n}{2}-1} e^{-\frac{x}{2}}$$

$$= \begin{cases} \frac{2^{-\frac{n}{2}}}{(\frac{n}{2}-1)!} x^{\frac{n}{2}-1} e^{-\frac{x}{2}} & \text{when } n \text{ is even.} \\ \\ \frac{2^{n-\frac{n}{2}-1}(\frac{n-1}{2})!}{(n-1)!\sqrt{\pi}} x^{\frac{n}{2}-1} e^{-\frac{x}{2}} & \text{when } n \text{ is odd.} \end{cases}$$

when $x > 0$.

F-distribution : "arises naturally" in statistics as a ratio of sum of squares.

later in the course

- $X_1, X_2, \ldots, X_m$ i.i.d Normal $(0, \sigma_1^2)$
  $Y_1, Y_2, \ldots, Y_n$ i.i.d Normal $(0, \sigma_2^2)$

- $U_i = \dfrac{X_i}{\sigma_1}$, $1 \le i \le m$ i.i.d. Normal $(0,1)$
  $V_i = \dfrac{Y_i}{\sigma_2}$, $1 \le i \le n$ i.i.d. Normal $(0,1)$

- $$U = \sum_{i=1}^{m} \left(\frac{X_i}{\sigma_1}\right)^2 \sim \chi^2_m \text{ degrees of freedom}$$
  and independent
  $$V = \sum_{i=1}^{n} \left(\frac{Y_i}{\sigma_2}\right)^2 \sim \chi^2_n \text{ degrees of freedom}$$

- $$Z = \frac{\frac{U}{m}}{\frac{V}{n}} \sim F(m, n)$$

Exercise – to complete the calculation

F-Distribution

Suppose $X_1, X_2, \ldots, X_{n_1}$ be an i.i.d. random sample from a Normal mean 0 and variance σ_1^2 population and $Y_1, Y_2, \ldots, Y_{n_2}$ be an i.i.d. random sample from a Normal mean 0 and variance σ_2^2 population. Let $U = \sum_{i=1}^{n_1} \left(\frac{X_i}{\sigma_1}\right)^2$ and $V = \sum_{i=1}^{n_2} \left(\frac{Y_i}{\sigma_2}\right)^2$.Let $Z = \frac{U}{n_1} / \frac{V}{n_2}$.

The density of Z, for $z > 0$ is given by

$$f(z) = \left(\frac{n_2}{n_1}\right)^{\frac{n_1}{2}} \frac{z^{\frac{n_1}{2}-1}}{(1 + \frac{n_1}{n_2} z)^{\frac{n_1+n_2}{2}}} \frac{\Gamma(\frac{n_1+n_2}{2})}{\Gamma(\frac{n_1}{2})\Gamma(\frac{n_2}{2})}.$$

Z is said to have $\mathrm{F}(n_1, n_2)$ distribution.

Z is close to a widely used distribution in statistics called F- distribution.

Sample $\quad X_1, X_2, \ldots, X_n$ i.i.d Assume Normal(μ, σ^2)

Sample mean $\quad \bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$

Sample variance $\quad S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$

(1) $\quad \bar{X} \overset{d}{=} \text{Normal}\left(\mu, \frac{\sigma^2}{n}\right)$ — $\boxed{\text{Result}}$

(2) [NOT A PROOF] $\quad$ "$X_i - \bar{X}$" = linear combination of $X_1, \ldots, X_n$ $\overset{d}{=}$ Normal(... , ...)

$(X_i - \bar{X})^2 \sim$ "chi-square"

$S^2 \equiv$ "sum of chi squares"

$$\frac{n-1}{\sigma^2} S^2 \sim \chi^2_{n-1} \text{ distribution}$$

— $\boxed{\text{Result}}$

(3)

$$\bar{X} = \frac{\sum_{i=1}^{n} X_i}{n} \quad \text{and} \quad S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$$

are independent random variables

— $\boxed{\text{Result}}$

(1) $\equiv$ $\boxed{\dfrac{\bar{X} - \mu}{\sigma/\sqrt{n}} \overset{d}{=} N(0,1)}$ $\quad$ In ...... Statistic σ - not known $\quad \boxed{\dfrac{\bar{X} - \mu}{S/\sqrt{n}}}$

t_n-distribution

Example from Statistics : $X_1, X_2, \ldots X_n$ i.i.d Normal with mean μ and unknown σ : $\boxed{\dfrac{\bar{X}-\mu}{S/\sqrt{n}}}$

Let X_1 be a Normal random variable with mean 0 and variance 1. Let X_2 be an independent χ^2_n random variable. Let

$$Z = \frac{X_1}{\sqrt{\frac{X_2}{n}}}.$$

The density of Z is given by

$$\begin{aligned} f_Z(z) &= \mid z \mid f_U(z^2) \\ &= \mid z \mid \frac{\Gamma\left(\frac{n+1}{2}\right)}{\sqrt{n\pi}\,\Gamma\left(\frac{n}{2}\right)} \frac{z^{2^{-\frac{1}{2}}}}{\left(1+\frac{u}{n}\right)^{\frac{n+1}{2}}} \\ &= \frac{\Gamma\left(\frac{n+1}{2}\right)}{\sqrt{n\pi}\,\Gamma\left(\frac{n}{2}\right)} \left(1+\frac{z^2}{n}\right)^{-\frac{n+1}{2}} \end{aligned}$$

Z is said to have t-distribution with n-degrees of freedom. We will denote this by the notation $Z \sim t_n$.

Result

Let $n \geq 1$, $X_1, X_2, \ldots X_n$, be an i.i.d random sample with distribution $X \sim$ Normal(μ, σ^2). Let $\overline{X}$ and S^2 be as above. Then

$$\frac{\sqrt{n}(\overline{X} - \mu)}{S}$$

has the t_{n-1}