Sampling Distributions: ①

Chi-Square Distribution: A r.v X is said to have χ^2 dist^n of n degree of freedom if its pdf is given as.

$$f_X(x) = \frac{x^{n/2-1}\, e^{-x/2}}{2^{n/2}\, \Gamma\frac{n}{2}} \; ; \quad x > 0$$

$E(X) = n$

$V(X) = 2n$

$M_X(t) = \frac{1}{(1-2t)^{n/2}}$

Degree of freedom: No of free components or no. of pieces of information required to estimate population values

Note: It is special case of Gamma dist^n.

Take $\alpha = \frac{n}{2}$, $\beta = 2$

<u>Note:</u> If $X_1, X_2, \ldots, X_n \overset{iid}{\sim} N(0,1)$

then $S_n = \sum_{i=1}^{n} X_i^2$ has $\chi^2_{(n)}$.

<u>Note:</u> (i) If $X \sim N(0,1)$ then $X^2 \sim \chi^2_{(1)}$

(ii) If $X_1, X_2 \overset{iid}{\sim} N(0,1)$ then $X_1^2 + X_2^2 \sim \chi^2_{(2)}$

and $E(X_1^2 + X_2^2) = 2$

$V(X_1^2 + X_2^2) = 2 \cdot 2 = 4$

(iii) If $X \sim N(\mu, \sigma^2)$ then

$$Z = \left(\frac{X-\mu}{\sigma}\right)^2 \sim \chi^2_{(1)}.$$

<u>Note:</u> If $X_1, X_2 \ldots, X_n \overset{iid}{\sim} N(\mu, \sigma^2)$. then

$$\sum_{i=1}^{n} \left(\frac{x_i - \mu}{\sigma}\right)^2 \sim \chi^2_{(n)}.$$

(3)

Prove that $\frac{(n-1)s^2}{\sigma^2} \sim \chi^2_{(n-1)}$ where $X_i \overset{iid}{\sim} N(\mu, \sigma^2)$ and s^2 is sample variance

Proof: $\because$ $X_i \sim N(\mu, \sigma^2)$

$$\frac{X_i - \mu}{\sigma} \sim N(0,1)$$

$$\bar{X} \sim N(\mu, \frac{\sigma^2}{n})$$

$$\sqrt{n}\left(\frac{\bar{X} - \mu}{\sigma}\right) \sim N(0,1)$$

Now consider

$$W = \sum_{i=1}^{n}\left(\frac{X_i - \mu}{\sigma}\right)^2 \sim \chi^2_{(n)}$$

$$= \sum_{i=1}^{n}\left(\frac{(X_i - \bar{X}) + (\bar{X} - \mu)}{\sigma}\right)^2$$

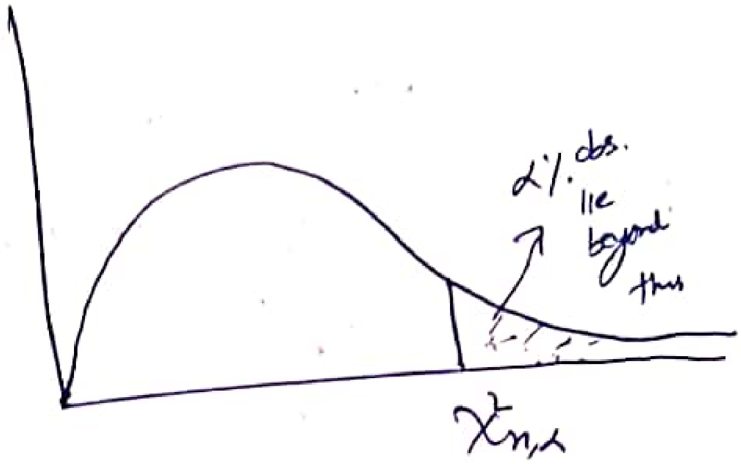
$$= \sum_{i=1}^{n}\left(\frac{X_i - \bar{X}}{\sigma}\right)^2 + \sum_{i=1}^{n}\left(\frac{\bar{X} - \mu}{\sigma}\right)^2 + \frac{2(\bar{X}-\mu)}{\sigma}\sum_{i=1}^{n}(X_i - \bar{X})$$

$$= \frac{1}{\sigma^2}\sum_{i=1}^{n}(X_i - \bar{X})^2 + \frac{n}{\sigma^2}(\bar{X} - \mu)^2 + \frac{2}{\sigma}(\bar{X} - \mu)(n\bar{X} - n\bar{X})$$

$$\underset{\chi^2_{(n)}}{W} = \frac{1}{\sigma^2}\cdot\frac{(n-1)}{(n-1)}\cdot\sum_{i=1}^{n}(X_i - \bar{X})^2 + \underbrace{\left(\frac{\bar{X} - \mu}{\sigma/\sqrt{n}}\right)^2}_{\chi^2_{(1)}}$$

$$\Rightarrow \frac{(n-1)\cdot s^2}{\sigma^2} \sim \chi^2_{(n-1)}.$$

Curve of $\chi^2_{(n)}$:

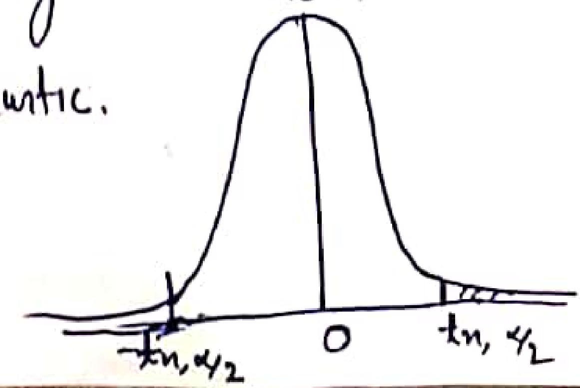


Student - t distribution: Let $X \sim N(0,1)$ and $Y \sim \chi^2_{(n)}$ and X and Y are independent then $T = \dfrac{X}{\sqrt{Y/n}}$ is said to have Student - t dist^n with p.d.f

$$f_T(t) = \frac{\Gamma\frac{n+1}{2}}{\sqrt{n\pi}\,\Gamma\frac{n}{2}\left(\frac{t^2}{n}+1\right)^{\frac{n+1}{2}}} \quad ; -\infty < t < \infty$$

Properties: (i). $f_T(t)$ is symmetric about 0.

(ii) It is leptokurtic.



Note: If $X \sim$ t-dist^n then $E(X) = 0$

F- distribution: Let X and Y be independant χ^2 r.vs with m and n degree of freedom respectively then r.v

$$F = \frac{X/m}{Y/n}$$ is said to have a

F-distribution with (m, n) degree of freedom. We write $F \sim F(m,n)$ and pdf is given as

$$g_F(f) = \frac{\Gamma\frac{m+n}{2} \cdot \left(\frac{m}{n}\right) \left(\frac{m}{n} \cdot f\right)^{\frac{m}{2}-1} \left(1+\frac{m}{n} f\right)^{-\left(\frac{m+n}{2}\right)}}{\Gamma\frac{m}{2} \cdot \Gamma\frac{n}{2}} \; ; f>0$$

Note! If $X \sim F(m,n)$ then $\frac{1}{X} \sim F(n,m)$

Note! If $m = 1$ then $F(1,n) \sim \chi^2_{(n)}$

(6)

(Ex): $\bar{X}$ and S^2 are independent, where $X_i \overset{iid}{\sim} N(\mu, \sigma^2)$ for $i=1,2 \cdot n$.

$$\bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right) \Rightarrow \sqrt{n}\left(\frac{\bar{X}-\mu}{\sigma}\right) \sim N(0,1)$$

Also $$\frac{(n-1) \cdot S^2}{\sigma^2} \sim \chi^2_{(n-1)}$$

then $$\frac{\sqrt{n}\left(\frac{\bar{X}-\mu}{\sigma}\right)}{\sqrt{\frac{(n-1) \cdot S^2}{\sigma^2 \cdot (n-1)}}} \sim t_{(n-1)}$$

$$\Rightarrow \sqrt{n}\left(\frac{\bar{X}-\mu}{S}\right) \sim t_{(n-1)}.$$

(Ex): Indep. $\begin{cases} X_1, X_2, \dots, X_m \sim N(\mu_1, \sigma_1^2), \bar{X}, S_1^2 \\ Y_1, Y_2, \dots, Y_n \sim N(\mu_2, \sigma_2^2), \bar{Y}, S_2^2 \end{cases}$

$$S_1^2 = \frac{1}{m-1} \sum_{i=1}^{m} (X_i - \bar{X})^2, \quad S_2^2 = \frac{1}{n-1} \sum_{i=1}^{n} (Y_i - \bar{Y})^2$$

$$\frac{(m-1) \cdot S_1^2}{\sigma_1^2} \sim \chi^2_{(m-1)} \text{ and } \frac{(n-1) \cdot S_2^2}{\sigma_2^2} \sim \chi^2_{(n-1)}$$

$$\Rightarrow \frac{(m-1) \cdot S_1^2}{\sigma_1^2 \cdot (m-1)} \Big/ \frac{(n-1) \cdot S_2^2}{\sigma_2^2 (n-1)} \sim F_{(m-1, n-1)}$$

$$\Rightarrow \frac{\sigma_2^2 S_1^2}{\sigma_1^2 S_2^2} \sim F_{(m-1, n-1)}.$$

__Note:__ If $\sigma_1^2 = \sigma_2^2 = \sigma^2$

$$\Rightarrow \frac{S_1^2}{S_2^2} \sim F_{(m-1, n-1)}.$$

Result $\because$ $\frac{(n-1)s^2}{\sigma^2} \sim \chi^2_{(n-1)}$ (7)

$$E\left(\frac{(n-1)s^2}{\sigma^2}\right) = n-1$$

$$\frac{(n-1)}{\sigma^2} E(s^2) = (n-1)$$

$$\Rightarrow \boxed{E(s^2) = \sigma^2}$$

$$V\left(\frac{(n-1)s^2}{\sigma^2}\right) = 2(n-1)$$

$$\Rightarrow \frac{(n-1)^2}{\sigma^4} \cdot V(s^2) = 2(n-1)$$

$$\Rightarrow V(s^2) = \frac{2(n-1)\cdot\sigma^4}{(n-1)^2}$$

$$\Rightarrow \boxed{V(s^2) = \frac{2\cdot\sigma^4}{(n-1)}}$$

Ordered Statistics

Let $X_1, X_2, \ldots, X_n$ be iid (continuous) random variables with same pdf and CDF $f_X(x)$ and $F_X(x)$ respectively,

$$X_{(1)} = \min\{X_1, X_2, \ldots, X_n\}$$

$$X_{(2)} = 2^{nd} \min\{X_1, X_2, \ldots, X_n\}$$

$$\vdots$$

$$X_{(n)} = \max\{X_1, X_2, \ldots, X_n\}$$

The random variables $X_{(1)}, X_{(2)}, \ldots, X_{(n)}$ are called ordered statistics of $(X_1, X_2, \ldots, X_n)$

Distribution function of $X_{(1)}$:-

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

$$= 1 - P(X_{(1)} > x_{(1)})$$

$$= 1 - P(X_{(1)} > x_{(1)}, X_{(2)} > x_{(1)}, \ldots, X_{(n)} > x_{(1)})$$

$$= 1 - \prod_{i=1}^{n} P(X_{(i)} > x_{(1)})$$

$$= 1 - \prod_{i=1}^{n} [1 - P(X_{(i)} \leq x_{(1)})]$$

$$= 1 - [1 - F_{X_{(i)}}(x_{(1)})]^n$$

$$f_{X_{(1)}}(x_{(1)}) = n\,[1 - F(x_{(1)})]^{n-1} \cdot f(x_{(1)}).$$

$$; \quad -\infty < x_{(1)} < x_{(2)} < \cdots < x_{(n)} < \infty$$

Distribution function of $X_{(n)}$

$$F_{X_{(n)}}(x_{(n)}) = P(X_{(n)} \leq x_{(n)})$$

$$= P(X_{(1)} \leq x_{(n)}, X_{(2)} \leq x_{(n)}, \ldots X_{(n)} \leq x_{(n)})$$

$$= \prod_{i=1}^{n} P(X_{(i)} \leq x_{(n)})$$

$$= [F_{X_{(i)}}(x_{(n)})]^n$$

$$= [F(x_{(n)})]^n$$

$$f_{X_{(n)}}(x_{(n)}) = n\,[F(x_{(n)})]^{n-1} \cdot f(x_{(n)}); \quad -\infty \leq x_{(n)} < \infty$$

r^{th} order Statistic pdf

$$f_{Y_{(r)}}(y_{(r)}) = \frac{n!\,[F(y_{(r)})]^{n-1}\,[1-F(y_{(r)})]^{n-r}}{(r-1)!\,(n-r)!}\,f(y_{(r)})$$

Ex: Let $X_1, X_2, \ldots, X_n \overset{iid}{\sim} U(0,\theta)$, $\theta > 0$ find $F_{X_{(n)}}(x_{(n)})$?

Sol^n

$$X_i \sim U(0,\theta)$$

$$f_{X_i}(x_i) = \frac{1}{\theta} ; \quad 0 < x_i < \theta ; \forall i = 1, 2 \ldots n$$

$$F_{X_i}(x_i) = \int_0^{x_i} \frac{1}{\theta} \cdot dx$$

$$= \frac{x_i}{\theta} \qquad \forall i = 1, 2 \ldots n$$

$$F_{X_{(n)}}(x_{(n)}) = [F(x_{(n)})]^n$$

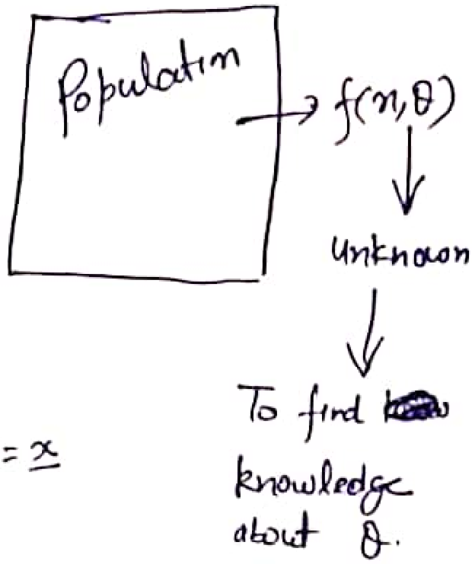
$$= \left[\frac{x_{(n)}}{\theta}\right]^n = \left(\frac{x}{\theta}\right)^n$$

$$f_{X_{(n)}}(x_{(n)}) = n \cdot \left(\frac{x}{\theta}\right)^{n-1} \cdot \frac{1}{\theta} ; \quad 0 < x < \theta$$

Point Estimation:

Estimator:

Estimator is a function of sample which is use to estimate the unknown value of given parametric function $g(\theta)$.



If $(X_1, X_2, \ldots, X_n) = \underline{X}$ is random sample from $f(x,\theta)$ then a function $T(\underline{X})$ used for estimating $g(\theta)$ is known as **estimator**. Let $(x_1, x_2, \ldots, x_n) = \underline{x}$ is observed values of $\underline{X}$ then $T(\underline{x})$ is known as **estimate** for $g(\theta)$.

Point Estimator: A point estimator is any function $T(X_1, X_2, \ldots, X_n)$ of a sample; i.e

any statistic is a point estimator.

Parameter Space: is the set of all possible values of parameters. and denoted as Θ

Ex $X \sim N(\mu, \sigma^2)$, μ is unknown
σ^2 is known

$\Theta = \{\mu : -\infty < \mu < \infty\}$

$\Theta = \{\sigma^2 : 0 < \sigma^2 < \infty\} \to \sigma^2$ unknown

$\Theta = \{ (\mu, \sigma^2) : \mu \in \mathbb{R}, \infty > \sigma^2 > 0 \}$ if both unknown

Desired properties of Estimators :-

(i) Unbiasedness property :- Let $(X_1, X_2, \ldots, X_n)$ be a r.s from a population with pdf $f(x, \theta), \theta \in \Theta$.

An estimator $T(\underline{X}) = T(X_1, X_2, \ldots, X_n)$ is said to be unbiased for estimating $g(\theta)$ if

$$E(T(\underline{X})) = g(\theta) \quad \forall \theta$$

Ex: $X_1, X_2, \ldots, X_n \overset{iid}{\sim} N(\mu, 1)$

Consider $T(\underline{X}) = \overline{X}$

$$E(T(\underline{X})) = E(\overline{X}) = \mu$$

Hence sample mean is unbiased estimator for population mean.

Note: Here $X_1, X_2, \ldots, X_n$ all are unbiased estimator for μ as

$$E(X_i) = \mu \quad \forall i = 1, 2, \ldots, n.$$

(2) Let $X \sim Bin(n, p)$, n is known.

$\because E(X) = np$

$E\left(\frac{X}{n}\right) = p$

$\Rightarrow \frac{X}{n}$ is unbiased estimator for p.

(3) If $X_1, X_2, \ldots X_n \sim P(\lambda)$

$E(\bar{X}) = \lambda$

$E(X_i) = \lambda$

<u>Note:</u> Unbiased estimator may not be unique.

<u>Result</u> Sample variance is unbiased estimator of population variance.

<u>Proof</u> $\because \frac{(n-1)s^2}{\sigma^2} \sim \chi^2_{(n-1)}$

$E\left(\frac{(n-1) \cdot s^2}{\sigma^2}\right) = (n-1)$

$\Rightarrow \frac{\cancel{(n-1)}}{\sigma^2} E S^2 = \cancel{(n-1)}$

$\Rightarrow \boxed{E(S^2) = \sigma^2}$ #

(Ex) If X is rv having pdf

$$f_X(x) = \frac{1}{\beta} e^{-x/\beta} \quad ; \ \beta > 0 \ , x > 0$$

then find the unbiased estimator for β

Soln

$$E(X) = \int_0^\infty x \cdot \frac{1}{\beta} e^{-x/\beta} \, dx$$

$$\boxed{E(X) = \beta}$$

(Ex): $X_1, X_2 \ldots, X_n \overset{iid}{\sim} N(\overset{\text{known}}{\mu}, \sigma^2)$ with unknown σ^2. find unbiased estimator for σ^2?

Soln

$$\cancel{\sum_{i=1}^{n} \left(\frac{X_i - \mu}{\sigma}\right)^2 \sim \chi^2_{(n)}}$$

$$\bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$$

$$V(\bar{X}) = \frac{\sigma^2}{n}$$

$$\Rightarrow E(\bar{X})^2 - (E(\bar{X}))^2 = \frac{\sigma^2}{n}$$

$$\Rightarrow E(\bar{X}^2) - \mu^2 = \frac{\sigma^2}{n}$$

$$\Rightarrow E(\bar{X}^2 - \mu^2) = \frac{\sigma^2}{n}$$

$$\Rightarrow \boxed{E(n(\bar{X}^2 - \mu^2)) = \sigma^2} \ ; \ T(X) = \underline{\underline{n(\bar{X}^2 - \mu^2)}}$$

Note: Unbiased estimator may not exist.

Ex: $X \sim Ber(p)$ then p^2 does not have unbiased estimator.

Let $T(X)$ is unbiased estimator for p^2

$$p_X(x) = p^x \cdot (1-p)^{1-x} \quad ; \quad x = 0,1$$

$$E(T(X)) = p^2$$

$$\Rightarrow \sum_{x=0}^{1} T(x) \cdot p_X(x) = p^2$$

$$\Rightarrow T(0)(1-p) + T(1) \cdot p = p^2$$

$$\Rightarrow \underbrace{p[T(1) - T(0)] + T(0)}_{\text{poly. of degree 1}} = \underbrace{p^2}_{\text{polynomial of degree 2}}$$

Contradiction.

So $T(X)$ is not an unbiased estimator for p^2.

#

Q: Q If $X \sim Bin(n, p)$

then find the unbiased estimator for $p(1-p)$

Sol^n

$\because V(X) = E(X(X-1)) + E(X) - (EX)^2$

$= n(n-1) \cdot p^2 + np - n^2 p^2$

$= npq$

Thus $E(X(X-1)) = n(n-1)p^2$

$\Rightarrow E\left(\frac{X(X-1)}{n(n-1)}\right) = p^2$

$E\left[\frac{X(X-1)}{n(n-1)} - \frac{X}{n}\right] = p^2 - p$

$= p(1-p)$

Thus $T(X) = \frac{X(X-1)}{n(n-1)} - \frac{X}{n}$ is unbiased estimator for $p(1-p)$.

$E\left(\frac{X}{n} - \frac{X(X-1)}{n(n-1)}\right) = p - p^2$

$= p(1-p)$

Thus $T(X) = \frac{X}{n} - \frac{X(X-1)}{n(n-1)}$ is unbiased estimator for $p(1-p)$.

Note: Even if the unbiased estimator exist, that might not be good

(ex) $X \sim P(\lambda)$, λ is unknown

find an unbiased estimator of $e^{-3\lambda}$.

Sol^n Consider $T(X) = (-2)^X$

$$p_X(x) = \frac{e^{-\lambda}\lambda^x}{x!} \; ; \quad x = 0, 1, 2 \ldots \quad \lambda > 0$$

$\cancel{E((-2)^X)}$

$$E(T(X)) = \sum_{x=0}^{\infty} (-2)^x \cdot \frac{e^{-\lambda}\lambda^x}{x!}$$

$$= e^{-\lambda} \sum_{x=0}^{\infty} \frac{(-1)^x \cdot (2\lambda)^x}{x!}$$

$$= e^{-\lambda}\left[1 - \frac{(2\lambda)}{1!} + \frac{(2\lambda)^2}{2!} - \frac{(2\lambda)^3}{3!} + \cdots\right]$$

$$= e^{-\lambda} \cdot e^{-2\lambda}$$

$$= e^{-3\lambda}$$

Thus $T(X) = (-2)^X$ is unbiased estimator for $e^{-3\lambda}$

But it is absurd estimator.

$T(X) = (-2)^X$

$T(0) = 1$, $T(1) = -2$, $T(2) = 4$, $T(3) = -8$.