## Chi-Square Dist$^n$ ($\chi^2$)

Let W be a continuous r.v. with pdf

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

is said to be a $\chi^2$ r.v. with n degrees of freedom.

This is nothing but Gamma $(\frac{n}{2}, \frac{1}{2})$ dist$^n$.

$$\mu_k' = E(W^k) = n(n+2)\cdots\{n+2(k-1)\}$$

$$E(W) = n,\ E(W^2) = n(n+2),\ V(W) = 2n.$$

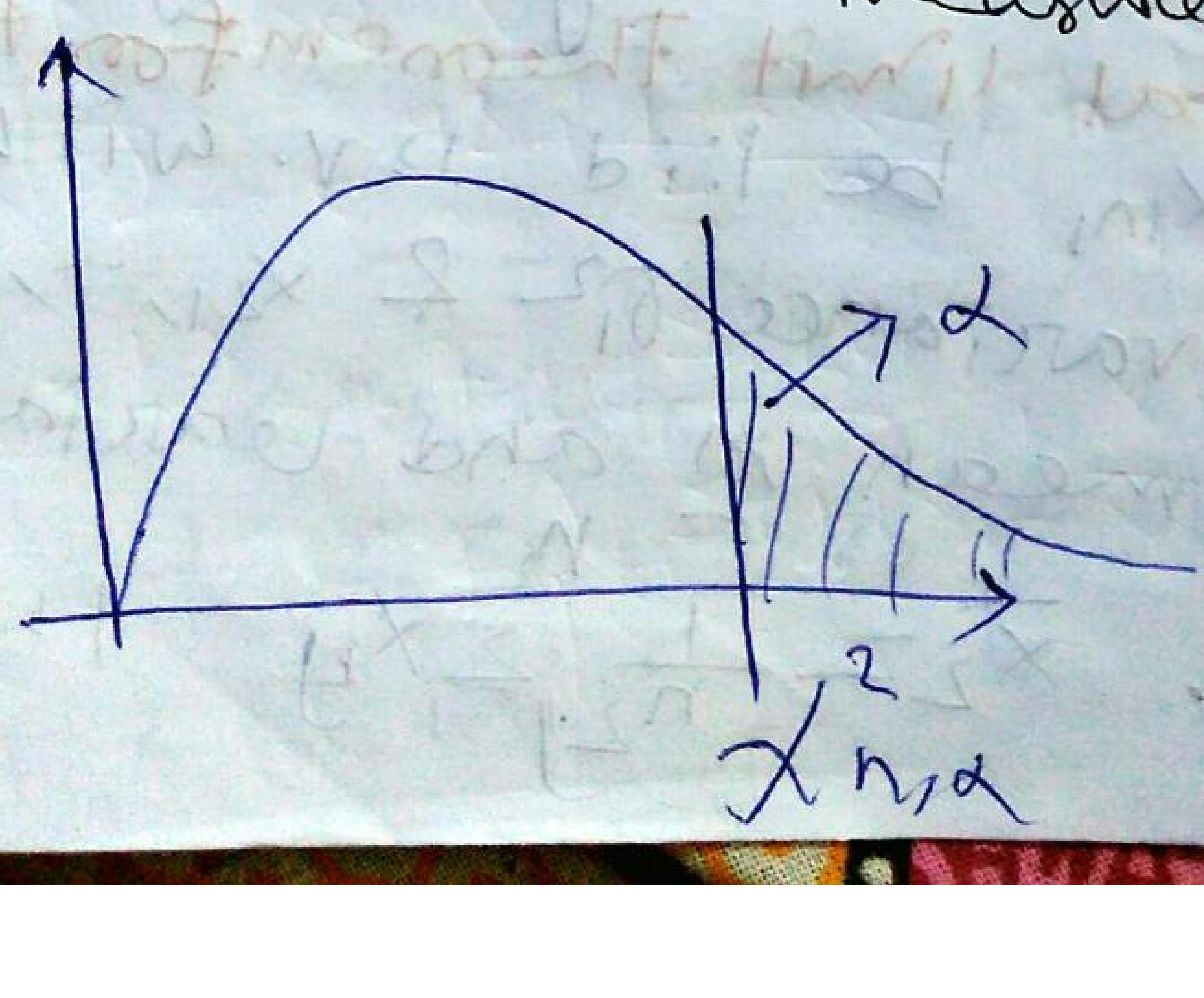
$$M_W(t) = \frac{1}{(1-2t)^{n/2}},\ t<\tfrac{1}{2}$$

$$\mu_3 = 8n>0,\quad \beta_1 = \frac{8n}{(2n)^{3/2}} = \sqrt{\frac{8}{n}}$$ → measure of asymmetry.

→ +vely skewed dist$^n$.

$$\mu_4 = 12n(n+4),\quad \beta_2 = \frac{\mu_4}{\mu_2^2} - 3 = \frac{12}{n} > 0$$

→ measure of kurtosis




If $P(W > \chi^2_{n,\alpha}) = \alpha$, $\chi^2_{n,\alpha}$ is called upper ~~100α%~~ 100(1-α)% point of $\chi^2_n$ dist$^n$.

**Theorem:** Let $W_1, W_2, \cdots, W_k$ be independent chi-square random variables with degrees of freedom $n_1, \cdots, n_k$ respectively. Then $W = \sum_{i=1}^{k} W_i \sim \chi^2_n$, where $n = \sum_{i=1}^{k} n_i$.

**Example:** Let $X \sim N(0,1)$ the $Y = X^2 \sim \chi^2_{(1)}$.

$X \sim N(0,1)$

$f_X(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}, \; -\infty < x < \infty$

$y = x^2$

$\Rightarrow x = -\sqrt{y}, \; x = \sqrt{y}$.

$\Rightarrow \frac{dx}{dy} = -\frac{1}{2\sqrt{y}}, \; \frac{dx}{dy} = \frac{1}{2\sqrt{y}}$.

$$f_Y(y) = \begin{cases} \frac{1}{\sqrt{2\pi}} e^{-y/2} \frac{1}{2\sqrt{y}} + \frac{1}{\sqrt{2\pi}} e^{-y/2} \frac{1}{2\sqrt{y}}, & y > 0 \\ 0, & \text{otherwise} \end{cases}$$

$$= \frac{1}{\sqrt{2\pi y}} e^{-y/2} = \frac{1}{2^{1/2}\Gamma\frac{1}{2}} e^{-y/2} y^{1/2 - 1}, \; y > 0$$

So, $Y = X^2 \sim \chi^2_{(1)}$.

- Let $X_1, X_2, \cdots, X_n$ be i.i.d $N(0,1)$. Then $W = \sum X_i^2 \sim \chi^2_n$.
- Let $X_1, X_2, \cdots, X_n \overset{\text{i.i.d}}{\sim} N(\mu, \sigma^2)$. $W = \sum_{i=1}^{n} \left(\frac{X_i - \mu}{\sigma}\right)^2 \sim \chi^2_n$.
- ~~Let $Y = \bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \sim N(\mu, \frac{\sigma^2}{n})$~~

Theorem:- Let $X_1, X_2, \dots, X_n \overset{i.i.d}{\sim} N(\mu, \sigma^2)$, Let $V_i = X_i - \bar{X}$, $i = 1, 2, \dots, n$ and $\underline{V} = (V_1, \dots, V_n)$. Then $Y = \bar{X}$ and $\underline{V}$ are independently distributed.

Proof:- We will show that the joint mgf of $(Y, \underline{V})$ is product of mgfs of $Y$ and $\underline{V}$.

$$M_Y(s) = e^{\mu s + \frac{1}{2}\frac{\sigma^2 s^2}{n}}$$

$$M_{\underline{V}}(\underline{t}) = E(e^{\Sigma t_i v_i}), \quad \underline{t} = (t_1, t_2, \dots t_n)$$

$$= E\left\{ e^{\Sigma t_i (x_i - \bar{x})} \right\}$$

$$= E\left\{e^{\sum t_i x_i - \bar{x}\sum t_i}\right\}$$

$$= \cancel{E\left(e^{\sum t_i x_i - n\bar{x}\bar{t}}\right)}$$

$$= \cancel{E\left(e^{\sum t_i x_i - \bar{t}\sum_{i=1}^{n} x_i}\right) = E\left(e^{\sum_{i=1}^{n}}\right.}$$

$$= \cancel{E\left\{e^{\sum t_i x_i}\right.} \quad = E\left\{e^{\sum t_i x_i - n\bar{x}\bar{t}}\right\}$$

$$= E\left\{e^{\sum t_i x_i - \bar{t}\sum_{i=1}^{n} x_i}\right\}$$

$$= E\left\{e^{\sum (t_i-\bar{t})x_i}\right\}$$

$$= E\left(\prod_{i=1}^{n} e^{(t_i-\bar{t})x_i}\right)$$

$$= \prod_{i=1}^{n} E\left\{e^{(t_i-\bar{t})x_i}\right\} = \prod_{i=1}^{n} M_{x_i}(t_i-\bar{t})$$

(∵ $x_i$'s are independent)

$$= \prod_{i=1}^{n}\left[e^{\mu(t_i-\bar{t})+\frac{1}{2}\sigma^2(t_i-\bar{t})^2}\right]$$

[If $x_i \overset{iid}{\sim} N(\mu,\sigma^2)$ then $M_{x_i}(t) = e^{\mu t+\frac{1}{2}\sigma^2 t^2}$]

$$= e^{\mu\sum_{i=1}^{n}(t_i-\bar{t})+\frac{1}{2}\sigma^2\sum_{i=1}^{n}(t_i-\bar{t})^2}$$

[∵ $\sum_{i=1}^{n}(t_i-\bar{t}) = 0$]

$$= e^{\frac{1}{2}\sigma^2\sum_{i=1}^{n}(t_i-\bar{t})^2}$$

Now, the joint mgf of $(Y, \underline{V})$ is

$$M_{(Y,\underline{V})}(s,t) = E\left(e^{sY+\sum t_i V_i}\right)$$

$$= E\left(e^{s\bar{x}+\sum_{i=1}^{n}(t_i-\bar{t})x_i}\right)$$

$$= E\left(e^{\frac{1}{n}s\sum_{i=1}^{n}x_i+\sum_{i=1}^{n}(t_i-\bar{t})x_i}\right)$$

$$= E\left(e^{\sum(t_i-\bar{t}+\frac{s}{n})x_i}\right)$$

$$= E\left(\prod_{i=1}^{n} e^{(t_i-\bar{t}+\frac{s}{n})x_i}\right)$$

$$= \prod_{i=1}^{n} E\left\{e^{(t_i-\bar{t}+\frac{s}{n})x_i}\right\} = \prod_{i=1}^{n} M_{x_i}\left(t_i-\bar{t}+\frac{s}{n}\right)$$

Scanned by TapScanner

$$= \prod_{i=1}^{n} \left[ e^{\mu(t_i - \bar{t} + \frac{s}{n}) + \frac{1}{2}\sigma^2 (t_i - \bar{t} + s/n)^2} \right]$$

$$= e^{\mu \sum_{i=1}^{n} (t_i - \bar{t} + \frac{s}{n}) + \frac{1}{2}\sigma^2 \sum_{i=1}^{n} (t_i - \bar{t} + s/n)^2}$$ [$\because$ $x_i$'s follows $N(\mu, \sigma^2)$]

$$= e^{\mu s + \frac{1}{2}\sigma^2 \sum_{i=1}^{n} (t_i - \bar{t})^2 + \frac{1}{2}\sigma^2 \frac{s^2}{n}}$$

$$= M_Y(s) \cdot M_V(t).$$

So, $Y = \bar{X}$ and $\underline{V} = (X_1 - \bar{X}, \ldots, X_n - \bar{X})$ are independently distributed.

Corollary :- Let $X_1, X_2, \ldots, X_n \overset{iid}{\sim} N(\mu, \sigma^2)$. Then $\overline{X}$ and $S^2$ are independent.

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

We have proved $\overline{X}$ and $\underline{V} = (X_1 - \overline{X}, \ldots, X_n - \overline{X})$ are independent.

$S^2$ is a fun$^n$ of $\underline{V}$, so, $\overline{X}$ & $S^2$ are independent.

Let $Y = \overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \sim N(\mu, \sigma^2/n)$

so, $\frac{Y-\mu}{\sigma/\sqrt{n}} \sim N(0,1)$

$\frac{n(Y-\mu)^2}{\sigma^2} \sim \chi_1^2$ or $\frac{n(\overline{X}-\mu)^2}{\sigma^2} \sim \chi_1^2$

Now, $\sum_{i=1}^{n}(X_i - \mu)^2 = \sum_{i=1}^{n}(X_i - \overline{X} + \overline{X} - \mu)^2$

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

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

e. $W = W_1 + W_2$.

Now, as $\overline{X}$ and ~~$\underline{V} = (X_1 - \overline{X}, \ldots, X_n - \overline{X})$~~ $S^2$ are independent, so $W_1$ & $W_2$ are independent

so, $M_W(t) = M_{W_1}(t) \cdot M_{W_2}(t)$

$\Rightarrow M_{W_1}(t) = \frac{M_W(t)}{M_{W_2}(t)} = \frac{(1-2t)^{-n/2}}{(1-2t)^{-1/2}} = (1-2t)^{-\left(\frac{n-1}{2}\right)}, \; t < \frac{1}{2}$

$[\because W \sim \chi_n^2$ & $W_2 \sim \chi_1^2$

$\therefore W_1 \sim \chi_{(n-1)}^2$

$\cancel{If} \Rightarrow \; S^2 = \frac{1}{(n-1)} \Sigma (x_i - \overline{x})^2$

then, $\frac{(n-1) s^2}{\sigma^2} \sim \chi^2_{(n-1)}$

$$W_1 = \frac{\sum_{i=1}^{n} (x_i - \overline{x})^2}{\sigma^2} = \frac{(n-1) s^2}{\sigma^2} \sim \chi^2_{(n-1)}.$$

So, this means that chisquare is a distribution of the sample variance after certain scaling. This shows that chi-square is a sampling distribution.

Note: In $\sum_{i=1}^{n} (x_i - \overline{x})^2$ ~~or~~ $(x_i - \overline{x})$ are ~~ind~~ not independent as $\sum_{i=1}^{n} (x_i - \overline{x}) = 0$. That is why d.f. is $n-1$.

$W_1 \sim \chi^2_{(n-1)}$

$E(W_1) = n-1$

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

$\Rightarrow E[s^2] = \sigma^2$

$\Downarrow$

$s^2 = \frac{1}{(n-1)} \sum_{i=1}^{n} (x_i - \overline{x})^2$ is an unbiased estimator of $\sigma^2$.

$Var(W_1) = 2(n-1)$

$\Rightarrow Var\left[\frac{(n-1) s^2}{\sigma^2}\right] = 2(n-1)$

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

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