(II)

Sampling Distribution

chi-Squared Distribution : For a positive integer 'n' a $G(n/2, 2)$ distⁿ is called the chi-squared distⁿ with n-degrees of freedom (d.f.) and denoted by χ^2_n.

The pdf of $Y \sim \chi^2_n$ is given by

$$f_Y(y) = \begin{cases} \frac{1}{2^{n/2}\Gamma(\frac{n}{2})} e^{-y/2} y^{n/2-1} & \text{if } y > 0 \\ 0 & \text{otherwise} \end{cases}$$

(i) If $Y \sim \chi^2_n$, then
$$E(Y) = n$$
$$V(Y) = 2n$$

(ii) The m.g.f. of $Y \sim \chi^2_n$ is given by

$$M_Y(t) = E(e^{tY}) = (1-2t)^{-n/2}, \quad t < 1/2$$

(iii) Let $Y_i \sim \chi^2_{n_i}$, $i = 1, 2, .., K$ are indep. Then $\sum_{i=1}^{K} Y_i \sim \chi^2_{\sum_{i=1}^{K} n_i}$

Result: (i) If $Z \sim N(0,1)$ then $Y = Z^2 \sim \chi^2_1$

(ii) Let X_1 and X_2 be Independent and identically distributed $N(0,1)$ r.v.s. Then $Y = X_2/X_1$ has a p.d.f.

$$f_Y(y) = \frac{1}{\pi}\frac{1}{1+y^2}, \quad -\infty < y < \infty$$

which is a Cauchy Distⁿ.

Theorem: Let $X_1, X_2, \ldots, X_n$ ($n \geq 2$) be a random sample from $N(\mu, \sigma^2)$ dist^n, where $\mu \in \mathbb{R}$, $\sigma > 0$.

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

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

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

(ii) $\bar{X}$ and S^2 are independent random variables

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

(iv) $E(S^2) = \sigma^2$, $Var(S^2) = \frac{2\sigma^4}{n-1}$, $E(S) = \sqrt{\frac{2}{n-1}} \frac{\Gamma(n/2)\sigma}{\Gamma(\frac{n-1}{2})}$

Proof: (i) This follows from linearity property of normal distribution. i.e. If $X_i \sim N(\mu, \sigma)$

$$\sum X_i \sim N(n\mu, n\sigma^2) \Rightarrow \bar{X} \sim N(\mu, \sigma^2/n).$$

(ii) Let $Y_i = X_i - \bar{X}$, $i = 1, 2, \ldots, n$ and $\underline{Y} = (Y_1, \ldots, Y_n)$

Then $\sum_{i=1}^{n} Y_i = \sum_{i=1}^{n} X_i - n\bar{X} = 0$

$$(n-1) S^2 = \sum_{i=1}^{n} (X_i - \bar{X})^2 = \sum_{i=1}^{n} Y_i^2 \text{ a fun}^n \text{ of } \underline{Y}.$$

The joint m.g.f of $(\underline{Y}, \bar{X})$ is given by

$$M_{\underline{Y}, \bar{X}}(\underline{u}, v) = E\left(e^{\sum_{i=1}^{n} u_i Y_i + v\bar{X}}\right), \quad \underline{u} \in \mathbb{R}^n, \; v \in \mathbb{R}.$$

Let fix $\underline{u} \in \mathbb{R}^n$ and $v \in \mathbb{R}$. Then we have

$$\sum_{i=1}^{n} u_i Y_i + v\bar{X} = \sum_{i=1}^{n} u_i (X_i - \bar{X}) + v\bar{X}$$

$$= \sum_{j=1}^{n} u_j X_j + \frac{\left(v - \sum_{i=1}^{n} u_i\right)}{n} \sum_{j=1}^{n} X_j$$

$$= \sum_{j=1}^{n} \left(u_j - \bar{u} + \frac{v}{n}\right) X_j \quad \left[\bar{u} = \frac{\sum u_i}{n}\right]$$

$$= \sum t_j X_j, \text{ where } t_j = \left(u_j - \bar{u} + \frac{v}{n}\right)$$

Now we have $\sum_{j=1}^{n} (u_j - \bar{u}) = 0$ so

$$\sum_{j=1}^{n} t_j = \sum_{j=1}^{n} \left(u_j - \bar{u} + \frac{v}{n}\right) = v \text{ and}$$

$$\sum_{j=1}^{n} t_j^2 = \sum_{j=1}^{n} \left(u_j - \bar{u} + \frac{v}{n}\right)^2 = \sum_{j=1}^{n} (u_j - \bar{u})^2 + v^2/n$$

Consequently $M_{\underline{Y}, \bar{X}}(\underline{u}, v) = E\left(e^{\sum t_j X_j}\right)$

$$= E\left(\prod_{j=1}^{n} e^{t_j X_j}\right) = \prod_{j=1}^{n} E\left(e^{t_j X_j}\right)$$

$$M_{\underline{Y}, \bar{X}}(\underline{u}, v) = \prod_{j=1}^{n} M_{X_j}(t_j) = \prod_{j=1}^{n} e^{\mu t_j + \frac{\sigma^2 t_j^2}{2}}$$

$$= \cancel{e^{\mu}} \cancel{\sum t_j} = \text{Exp}\left\{\mu \sum t_j + \frac{\sigma^2}{2} \sum t_j^2\right\}$$

$$= \text{Exp}\left\{\mu v + \frac{\sigma^2}{2}\left[\sum_{j=1}^{n}(u_j - \bar{u})^2 + v^2/n\right]\right\}$$

$$= \text{Exp}\left\{\mu v + \frac{\sigma^2 v^2}{2n}\right\} \text{Exp}\left\{\frac{\sigma^2 \sum (u_j - \bar{u})^2}{2}\right\}, \quad u \in \mathbb{R}^n, \ v \in \mathbb{R}.$$

$$= M_{\bar{X}}(v) \, M_{\underline{Y}}(u), \quad v \in \mathbb{R}, \ u \in \mathbb{R}^n.$$

So we ~~to.~~ have $\underline{Y} = (X_1 - \bar{X}, \dots, X_n - \bar{X})$ and $\bar{X}$ are independent. This implies for any Borel function $\psi_1(\cdot)$ and $\psi_2(\cdot)$, $\psi_1(Y)$ and $\psi_2(\bar{X})$ are indepen

$\Rightarrow$ S^2 (a funⁿ of $\underline{Y}$) and $\bar{X}$ are Indepen.

(iii) $X_i \sim N(\mu, \sigma^2) \Rightarrow Z_i = \dfrac{X_i - \mu}{\sigma} \sim N(0,1)$

and Z_i's are iid.

~~A~~gain $\bar{X} \sim N(\mu, \sigma^2/n)$ so $Z = \dfrac{\sqrt{n}(\bar{X} - \mu)}{\sigma} \sim N(0,1)$

$W = Z^2 = \dfrac{n(\bar{X} - \mu)^2}{\sigma^2}, \quad Y = \dfrac{(n-1)S^2}{\sigma^2}$

$W \sim \chi^2_1, \quad T = \sum_{i=1}^{n} Z_i^2 \sim \chi^2_n$

Thus M.g.f. of W and T are

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

$$M_T(t) = (1-2t)^{-n/2} , \quad t < 1/2$$

Also $T = \sum z_i^2 = \sum_{i=1}^{n} \left(\frac{x_i - \mu}{\sigma} \right)^2$

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

$$= Y + W$$

Since Y and W are independent random variable we have

$$M_t(t) = M_Y(t) M_W(t) \Rightarrow M_Y(t) = \frac{M_T(t)}{M_W(t)}$$

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

Which is m.g.f. of χ^2_{n-1} distribution. Now by uniqueness of m.g.f it ~~fowse~~ follows that

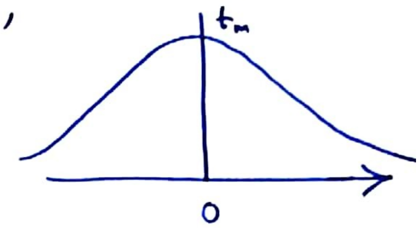
$$Y = \frac{(n-1)s^2}{\sigma^2} \sim \chi^2_{n-1}$$

(iv) $Y = \frac{(n-1)s^2}{\sigma^2} \sim \chi^2_{n-1}$

~~$E(s^r)$~~ $E(s^r) = \frac{\sigma^2}{(n-1)^{r/2}} E\left(Y^{r/2}\right) = \frac{\left(\frac{2}{n-1}\right)^{r/2} \Gamma\left(\frac{n-1+r}{2}\right)\sigma^r}{\Gamma\left(\frac{n-1}{2}\right) \cdot r > -(n-1)}$

Student t-distribution: For a given integer m a random variable X is said to have the Student t-dist^n with m degrees of freedom ($X \sim t_m$) if the p.d.f of X is given by

$$f_X(x) = \frac{\Gamma\left(\frac{m+1}{2}\right)}{\sqrt{m\pi}\,\Gamma\left(\frac{m}{2}\right)} \left(1+\frac{x^2}{m}\right)^{-\frac{m+1}{2}}, \quad -\infty < x < \infty$$



(The Probable Error of a Mean (1908)
By student. William S. Gosset)

Note: (i) The Student-t-dist^n with 1 degree of freedom is also called the standard cauchy dist^n

(ii) If $X \sim t_m$ then $f_X(x) = f_X(-x)$ then $X \stackrel{d}{=} -X$ i.e. the dist^n is symmetric about 0. Hence all odd order moments vanish (provided they exists). Even order moment exist of order $< n$.

Theorem: (i) Let $Z \sim N(0,1)$ and $Y \sim \chi^2_m$ $(m = 1, 2, \cdots)$ be independent random variables then

$$T = \frac{Z}{\sqrt{Y/m}} \sim t_m.$$

(ii) If $X \sim t_m$ then $E(X) = 0$

$$Var(X) = \frac{m}{m-2}, \quad \text{for } m \in \{3, 4, \cdots\}$$

Skewness is $= 0$

Kurtosis $= 3(m-2)/(m-4)$, $m \in \{5, 6, \cdots\}$

Theorem

Let $X_1, X_2, \ldots, X_n$ $(n \geq 2)$ be a random sample from $N(\mu, \sigma^2)$ dist^n, where $\mu \in (-\infty, \infty)$ and $\sigma > 0$.

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i, \qquad S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$$

then $$\frac{\sqrt{n}(\bar{X}-\mu)}{S} \sim t_{n-1}$$

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

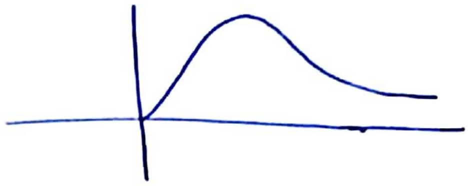
$$W = \frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{n-1}$$

Now $$T = Z \Big/ \sqrt{\frac{W}{n-1}} \sim t_{n-1}$$

$$T = \frac{\sqrt{n}(\bar{X}-\mu)}{S} \sim t_{n-1}$$

F-distribution: For any positive integers n_1 and n_2 a random variable X is said to have the Snedecor F dist^n with (n_1, n_2) degrees of freedom $(X \sim F_{n_1, n_2})$ if the p.d.f. of X is given by

$$f_X(x) = \frac{\left(\frac{n_1}{n_2}\right)\left(\frac{n_1}{n_2}x\right)^{\frac{n_1}{2}-1}}{B\left(\frac{n_1}{2}, \frac{n_2}{2}\right)} \left(1 + \frac{n_1}{n_2}x\right)^{-\frac{n_1+n_2}{2}}, \quad x > 0.$$



__Theorem__: (i) For positive integers n_1, n_2, let $X_1 \sim \chi^2_{n_1}$ and $X_2 \sim \chi^2_{n_2}$ be independent random variables.

Then

$$U = \frac{X_1/n_1}{X_2/n_2} \sim F_{n_1, n_2}$$

(ii) $$E(U) = \frac{n_1}{n_2 - 2} \quad \text{if } n_2 \in \{3, 4, \cdots\}$$

$$Var(U) = \frac{2n_2^2(n_1 + n_2 - 2)}{n_1(n_1-2)^2(n_2-4)} \quad \text{if } n_2 \in \{5, 6, \cdots\}$$

__Theorem__: $X_1, X_2, \ldots, X_m \overset{iid}{\sim} N(\mu_1, \sigma_1^2)$, $Y_1, \ldots, Y_n \overset{iid}{\sim} N(\mu_2, \sigma_2^2)$ independ.

$$S_1^2 = \frac{1}{m-1}\sum_{i=1}^{m}(X_i - \bar{X})^2 \Rightarrow W_1 = \frac{(m-1)S_1^2}{\sigma_1^2} \sim \chi^2_{m-1}$$

$$S_2^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{Y})^2 \Rightarrow W_2 = \frac{(n-1)S_2^2}{\sigma_2^2} \sim \chi^2_{n-1}$$

Now $$\frac{W_1/m-1}{W_2/n-1} = \frac{\sigma_2^2 S_1^2}{\sigma_1^2 S_2^2} \sim F_{m-1, n-1}.$$

Note: (I) $Z \sim N(0,1)$, $Y \sim \chi^2_m$ are independent

$$X = \frac{Z}{\sqrt{Y/m}} \sim \cancel{\chi^2_m}\,.\ t_m$$

Now consider $X^2 = \dfrac{Z^2}{Y/m}$

Since $Z \sim N(0,1) \Rightarrow Z^2 \sim \chi^2_1$. so $X^2 = \dfrac{Z^2/1}{Y/m}$

$$\Rightarrow X^2 \sim F_{1,m}$$

(ii) If $X \sim F_{n_1,n_2}$ then $\frac{1}{X} \sim F_{n_2,n_1}$

$X \sim F_{n_1,n_2} \Rightarrow X \overset{d}{=} \dfrac{X_1/n_1}{X_2/n_2}$, where

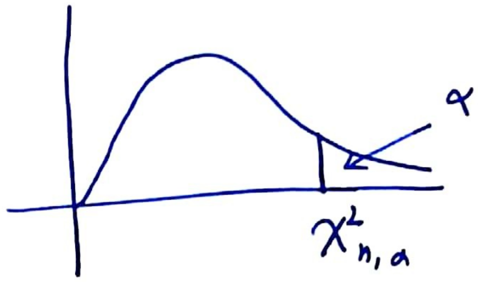
$X_1 \sim \chi^2_{n_1}$, $X_2 \sim \chi^2_{n_2}$ and independent

$$\frac{1}{X} \overset{d}{=} \frac{X_2/n_2}{X_1/n_1} \sim F_{n_2,n_1}$$

Some important facts

(1) We will write $\chi^2_{n,\alpha}$ for upper α percent point of the χ^2_n distn, that is,

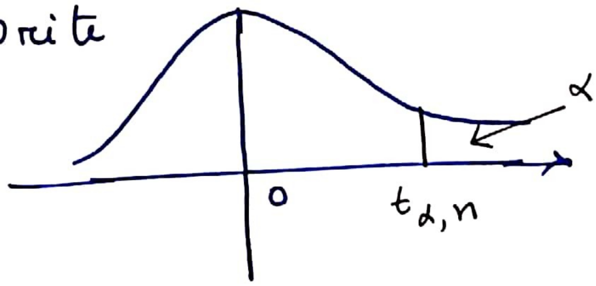
$$P(\chi^2_n > \chi^2_{n,\alpha}) = \alpha$$



For given values of n & α we can find the $\chi^2_{n,\alpha}$ from the χ^2 table.

(ii) For $n > 30$ it is possible to use normal approximation.

(iii) For t-distribution we write $t_{\alpha,n}$ be the upper $100\alpha\%$ point of t-distn, i.e,



$$P(T \geq t_{\alpha,n}) = \alpha.$$

Theorem: Let $T \sim t_n$. As $n \to \infty$ the pdf of T converge to $\phi(t) = \frac{1}{\sqrt{2\pi}} e^{-t^2/2}$