Example: Let a random sample of size 54 be taken from a discrete distn with pmf $p(x) = \frac{1}{3}$, $x = 2, 4, 6$.
Find the prob that the sample mean will lie between 4.1 to 4.4.

Soln: $n = 54$, $\mu = \frac{1}{3}(2+4+6) = 4$

$$E(X^2) = \frac{1}{3}(4+16+36) = \frac{56}{3}, \quad \sigma^2 = \frac{8}{3}.$$

We apply central limit theorem (CLT)

$$Z = \frac{\sqrt{54}\,(\bar{X}_{54} - 4)}{\sqrt{8/3}} \rightarrow N(0,1) \qquad \bar{X}_{54} = \frac{1}{54}\sum_{i=1}^{54} X_i$$

$$P(4.1 \leq \bar{X}_{54} \leq 4.4)$$

$$\simeq P\left(\frac{\sqrt{54}\,(4.1-4)}{\sqrt{8/3}} \leq Z \leq \frac{\sqrt{54}\,(4.4-4)}{\sqrt{8/3}} \right)$$

$$= P(0.45 \leq Z \leq 1.8)$$

$$= \Phi(1.8) - \Phi(0.45) = 0.9641 - 0.6736$$

$= 0.2905$

2. The TV picture tubes of manufacturer A have a mean life of 6.5 years and s.d. 0.9 years. Those from manufacturer B have a mean life of 6 years and s.d. 0.8 years. A random sample 36 tubes is taken from A and 49 tubes is taken B. What is the prob that $\bar{X} - \bar{Y}$ exceeds 1 years?

<u>Sol</u>. We apply CLT for two samples.

$$m = 36, \quad n = 49, \quad \mu_1 = 6.5, \quad \sigma_1 = 0.9$$

$$\mu_2 = 6, \quad \sigma_2 = 0.8$$

$$Z = \frac{\bar{X} - \bar{Y} - 0.5}{\sqrt{\frac{(0.9)^2}{36} + \frac{(0.8)^2}{49}}} \longrightarrow N(0,1)$$

$$\longrightarrow 0.189$$

$$P(\bar{X} - \bar{Y} > 1) \approx P\left(Z > \frac{1 - 0.5}{0.189}\right)$$

$$= P(Z > 2.65) = 0.004.$$

Let $X_1, \ldots, X_n$ be a random sample from a $N(\mu, \sigma^2)$ population.

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

So the distn of the sample mean from a normal population is again normal.

Let us consider the sample variance

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

We want dist^n of S^2 ??

<u>Chi-square (χ^2) dist^n</u> :

A continuous r.v. W is said to have a Chi-square dist^n on n degrees of freedom if it has pdf

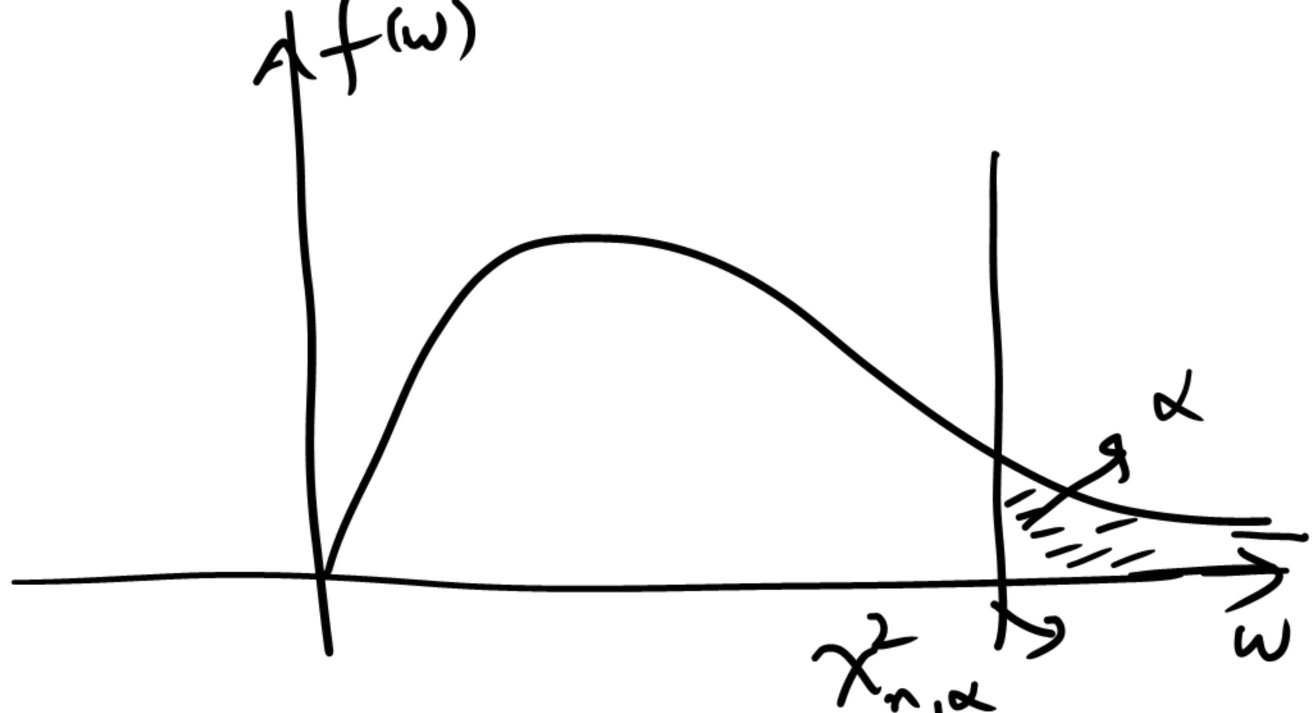
$$f_W(w) = \frac{1}{2^{n/2}\,\Gamma\frac{n}{2}}\, e^{-w/2}\, w^{\frac{n}{2}-1}, \quad \begin{array}{l} w>0 \\ n>0 \end{array}$$

Actually this is Gamma$\left(\frac{n}{2}, \frac{1}{2}\right)$ dist^n.

$$E(W) = n, \quad V(W) = 2n$$

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

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



$$\mu_3 = 8n > 0,$$

χ^2 distⁿ is always +vely skewed

$$\beta_1 = \frac{\mu_3}{\sigma^3} = \frac{8n}{(2n)^{3/2}} = \sqrt{\frac{8}{n}} \to 0 \text{ as } n \to \infty$$

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

$\to 0$ as $n \to \infty$

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

$\downarrow$

upper $100(1-\alpha)\%$ point of χ^2_n distn.

<u>Additive Property of χ^2 distn.</u>

Let $W_1, \ldots, W_k$ be independent r.v's with $W_i \sim \chi^2_{n_i}$. Then

$$U = \sum_{i=1}^{k} W_i \sim \chi^2_{\sum_{i=1}^{k} n_i}.$$

<u>Pf</u>. $M_U(t) = \prod_{i=1}^{k} M_{W_i}(t)$

$$= \prod_{i=1}^{k} \frac{1}{(1-2t)^{n_i/2}} = \frac{1}{(1-2t)^{\sum n_i/2}}$$

Let $X \sim N(0,1)$ and $Y = X^2$

$$f_Y(y) = \frac{1}{\sqrt{2\pi}} e^{-y/2} \cdot \frac{1}{\sqrt{y}}, \quad y > 0$$

$$\sim \chi^2_1 .$$

So if $X \sim N(0,1)$, $X^2 \sim \chi^2_1$.

Let $X_1, \ldots, X_n \overset{i.i.d.}{\sim} N(0,1)$

Then $W = \sum_{i=1}^{n} X_i^2 \sim \chi^2_n$

Now let $X_1, \dots, X_n \overset{i.i.d.}{\sim} N(\mu, \sigma^2)$

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

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

Let $U_i = X_i - \bar{X}$, $i = 1, \dots, n$

$\underline{U} = (U_1, \dots, U_n)$.

<u>Theorem</u>: Y and $\underline{U}$ are independently distributed.

<u>Proof</u>: We will prove independence using

mgf approach. i.e. to show that

$$M_{Y, \underline{U}}(s, \underline{t}) = M_Y(s)\, M_{\underline{U}}(\underline{t})$$

for all $(s, \underline{t}) \in \mathbb{R}^{n+1}$

$\underline{t} = (t_1, \ldots, t_n)$.

$\bar{t} = \frac{1}{n} \sum t_i$

$$M_{\underline{U}}(\underline{t}) = E\left\{ e^{\sum_{i=1}^{n} t_i U_i} \right\}$$

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

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

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

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

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

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

$$= e^{\mu \underbrace{\sum (t_i - \bar{t})}_{0} + \frac{1}{2}\sigma^2 \sum (t_i - \bar{t})^2}$$

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

$$= e$$

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

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

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

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

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

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

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

$$= M_Y(s)\, M_{\underline{U}}(\underline{t}).$$

<u>Corollary</u>: $\bar{X}$ and S^2 are independently dist^n.

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

$$W_1 = \frac{\sum (X_i - \bar{X})^2}{\sigma^2}, \qquad W_2 = \frac{n(\bar{X} - \mu)^2}{\sigma^2}$$

Then $W = W_1 + W_2$

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

So $W_2 \sim \chi^2_1$

Also W_1 and W_2 are independently dist^d. So $M_W(t) = M_{W_1}(t)\, M_{W_2}(t)$

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

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

which is mgf of χ^2_{n-1} distr^n

So $W_p = \dfrac{\Sigma(X_i - \bar{X})^2}{\sigma^2} = \dfrac{(n-1)\, S^2}{\sigma^2} \sim \chi^2_{n-1}$

$E(W_1) = (n-1) \;\Rightarrow\; E\left(\dfrac{(n-1)S^2}{\sigma^2}\right) = (n-1)$

$\Rightarrow \; E(S^2) = \sigma^2.$

Next we consider another sampling dist^n called Student's-t dist^n.

Let X and Y be independent r.v's.

Let $X \sim N(0,1)$ and $Y \sim \chi^2_n$

Then $T = \dfrac{X}{\sqrt{Y/n}}$ is said to follow

t- dist^n on n degrees of freedom

$$T \sim t_n$$

The pdf of T can be derived as

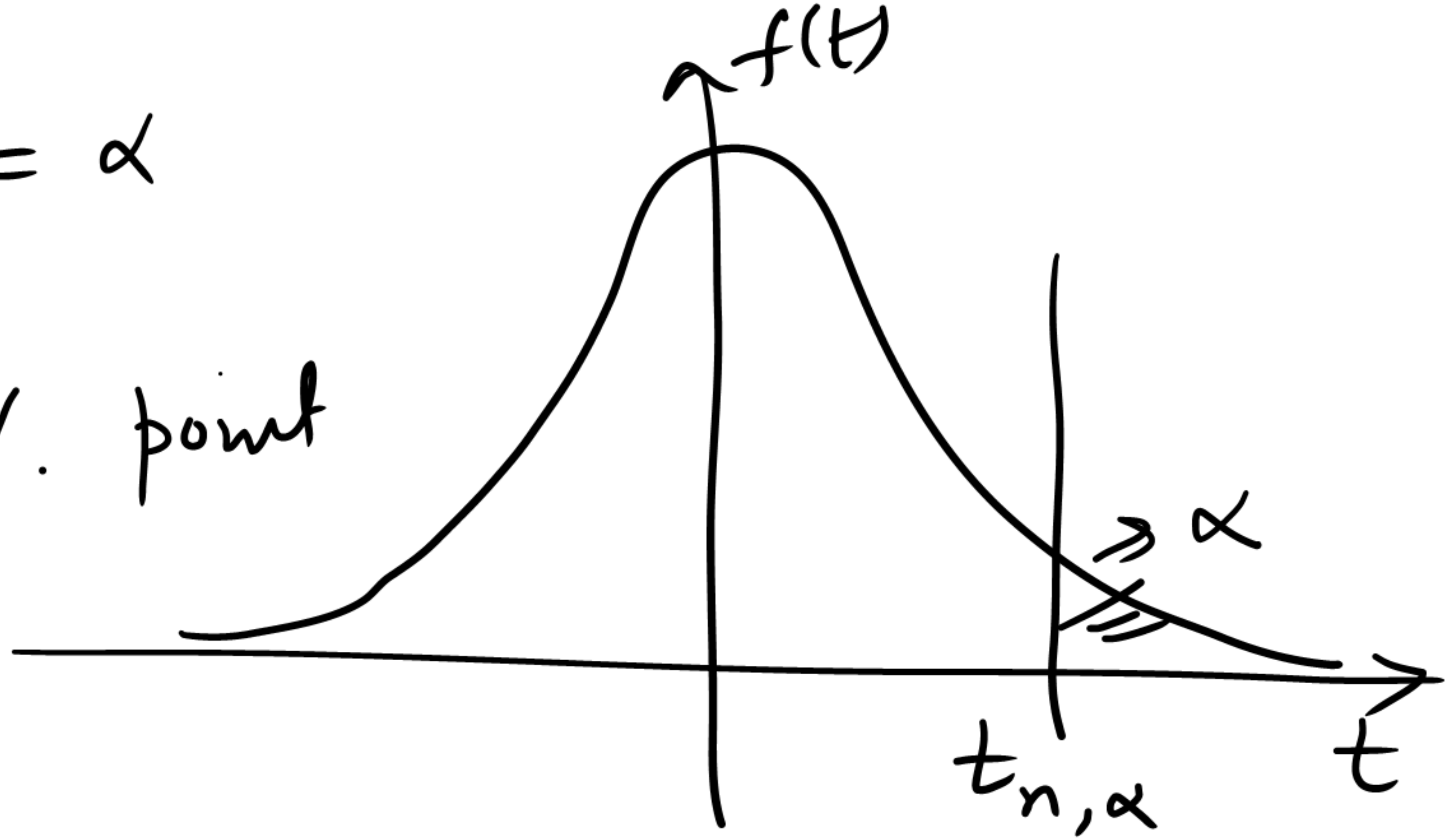
$$f_T(t) = \frac{1}{\sqrt{n}\, B\left(\frac{n}{2}, \frac{1}{2}\right)} \left(1 + \frac{t^2}{n}\right)^{-\left(\frac{n+1}{2}\right)}, \quad -\infty < t < \infty$$

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

$\downarrow$

upper $100(1-\alpha)\%$ point

of t_n distn.



The density is symmetric about 0.

So $E(T) = 0$.

$$\mu_k' = E(T^k) = \frac{n^{k/2}\,\Gamma\left(\frac{k+1}{2}\right)\Gamma\left(\frac{n-k}{2}\right)}{\Gamma\frac{1}{2}\;\Gamma\frac{n}{2}},\quad k<n$$

$$E(T^2) = V(T) = \frac{n}{n-2}, \quad n > 2$$

$$\mu_4 = \frac{3n^2}{(n-2)(n-4)}, \quad n > 4$$

$$\beta_2 = \frac{\mu_4}{\mu_2^2} - 3 = \frac{6}{n-4} > 0$$

So T distn is leptokurtic

Suppose $X_1, \ldots, X_n \overset{i.i.d.}{\sim} N(\mu, \sigma^2)$

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

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

$\bar{X}$ & S^2 are indept.

Then $$\frac{\sqrt{n}(\bar{X}-\mu)/\cancel{\sigma}}{\sqrt{\frac{\cancel{(n-1)}S^2}{\cancel{\sigma^2}}/\cancel{(n-1)}}} \sim t_{n-1}$$

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

Theorem: Let $T \sim t_n$. As $n \to \infty$ the $f_T(t) \to \phi(t) \; \forall t$

Usually for $n > 30$, t-tables & normal tables have approximately same values.

Next we consider Fisher's F-distⁿ.

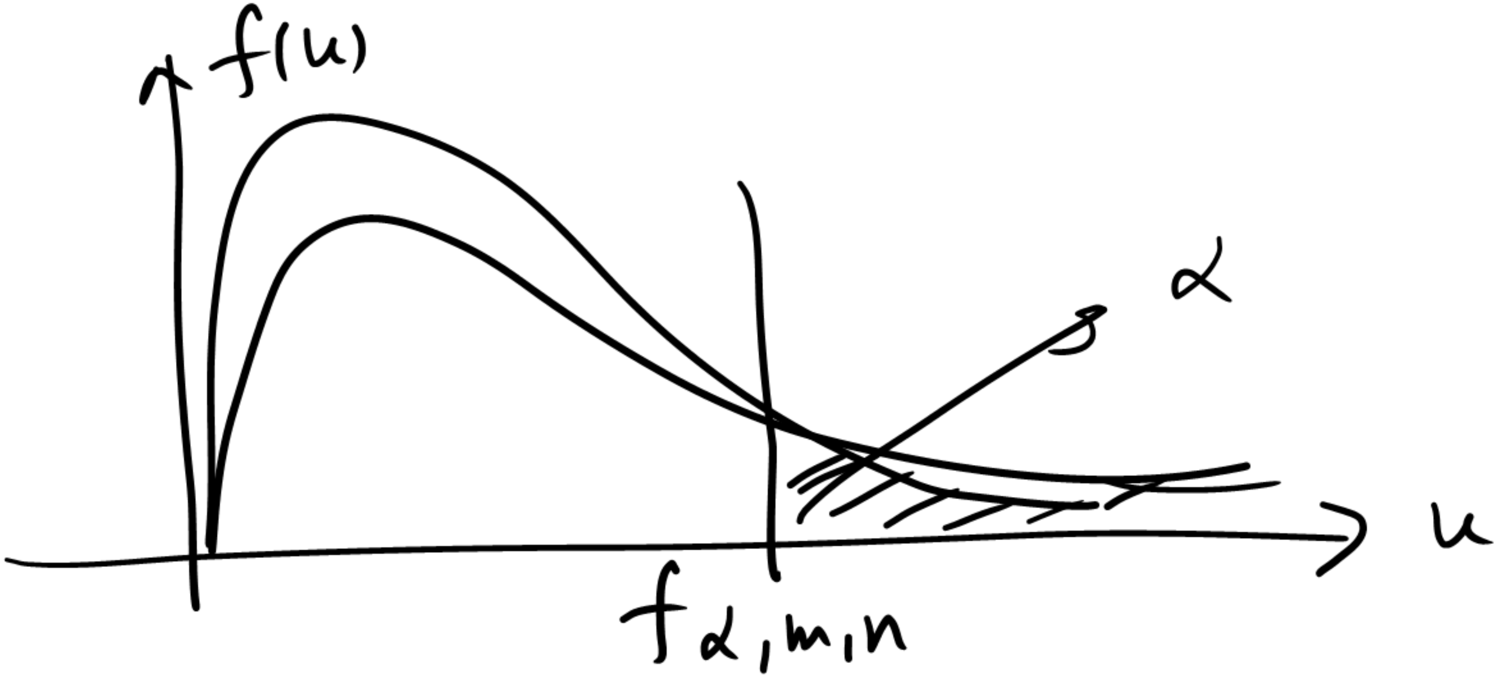
Let Y_1 and Y_2 be independent r.v's with $Y_1 \sim \chi^2_m$, $Y_2 \sim \chi^2_n$.

Then $U = \dfrac{(Y_1/m)}{(Y_2/n)} = \dfrac{nY_1}{mY_2}$ is said

to have F-distⁿ on (m, n) d.f.

The pdf of $F_{m,n} = U$ is

$$f_U(u) = \frac{(m/n)^{m/2}}{B\left(\frac{m}{2}, \frac{n}{2}\right)} \frac{u^{m/2-1}}{\left(1+\frac{m}{n}u\right)^{\frac{m+n}{2}}}, \quad u > 0$$



$$P(U > f_{\alpha,m,n}) = \alpha$$

$\downarrow$

upper $100(1-\alpha)\%$ point of $F_{m,n}$.

If $U \sim F_{m,n} \Rightarrow \frac{1}{U} \sim F_{n,m}$

$$P(U > f_{\alpha,m,n}) = \alpha$$

$$\Rightarrow P\left(\frac{1}{U} < \frac{1}{f_{\alpha,m,n}}\right) = \alpha$$

$$\Rightarrow P\left(F_{n,m} < \frac{1}{f_{\alpha,m,n}}\right) = \alpha$$

$$\Rightarrow P\left(F_{n,m} \geq \frac{1}{f_{\alpha,m,n}}\right) = 1-\alpha$$

$$\Rightarrow \quad \frac{1}{f_{\alpha,m,n}} = f_{1-\alpha,n,m}$$

$$E(U) = \frac{n}{n-2}, \quad n > 2$$

$$V(U) = \frac{2n^2(m+n-2)}{m(n-2)^2(n-4)}, \quad n > 4.$$

The moments upto order $k \; (< \frac{n}{2})$ exist.

Let $X_1, \dots, X_m$ be a random sample from $N(\mu_1, \sigma_1^2)$ and let $Y_1, \dots, Y_n$ be another independent random sample from $N(\mu_2, \sigma_2^2)$.

$$S_1^2 = \frac{1}{(m-1)} \sum (X_i - \bar{X})^2, \quad \frac{(m-1)S_1^2}{\sigma_1^2} \sim \chi^2_{m-1}$$

indept

$$S_2^2 = \frac{1}{(n-1)} \sum_{j=1}^{n} (Y_j - \bar{Y})^2, \quad \frac{(n-1)S_2^2}{\sigma_2^2} \sim \chi^2_{n-1}$$

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

<u>Example</u>: Let the time to failure of a bulb be a r.v. with mean μ and s.d. 100 hrs. If the failure times of n bulbs are recorded, how large n should be so that the prob of the sample mean differing from μ be less than 50 hrs. is at least 0.95?

$$P(|\bar{X}-\mu| \le 50) \geq 0.95 \quad ??$$

$$\approx P\left(\left|\frac{\sqrt{n}(\bar{X}-\mu)}{100}\right| \leq \frac{50\sqrt{n}}{100}\right) \geq 0.95$$

$$\overset{CLT}{\approx} P\left(|Z| \leq \frac{\sqrt{n}}{2}\right) \geq 0.95$$

$$Z \sim N(0,1)$$

$$\Rightarrow 2\Phi\left(\frac{\sqrt{n}}{2}\right) - 1 \geq 0.95$$

$$\Rightarrow \Phi\left(\frac{\sqrt{n}}{2}\right) \geq 0.975 \Rightarrow \frac{\sqrt{n}}{2} \geq 1.96$$

$$\Rightarrow n \geq 16.$$

Confidence Intervals

Let $X_1 \dots X_n$ be a random sample from a population with distⁿ $P_{\underline{\theta}}$, $\underline{\theta} \in \Theta$

For parametric function $g(\underline{\theta}) = \eta$, we say $(T_1(\underline{X}), T_2(\underline{X}))$ is $100(1-\alpha)\%$ confidence interval if

$$P\left(T_1(\underline{X}) \leq g(\underline{\theta}) \leq T_2(\underline{X}) \right) = 1-\alpha$$

Method of Pivots for developing confidence intervals.

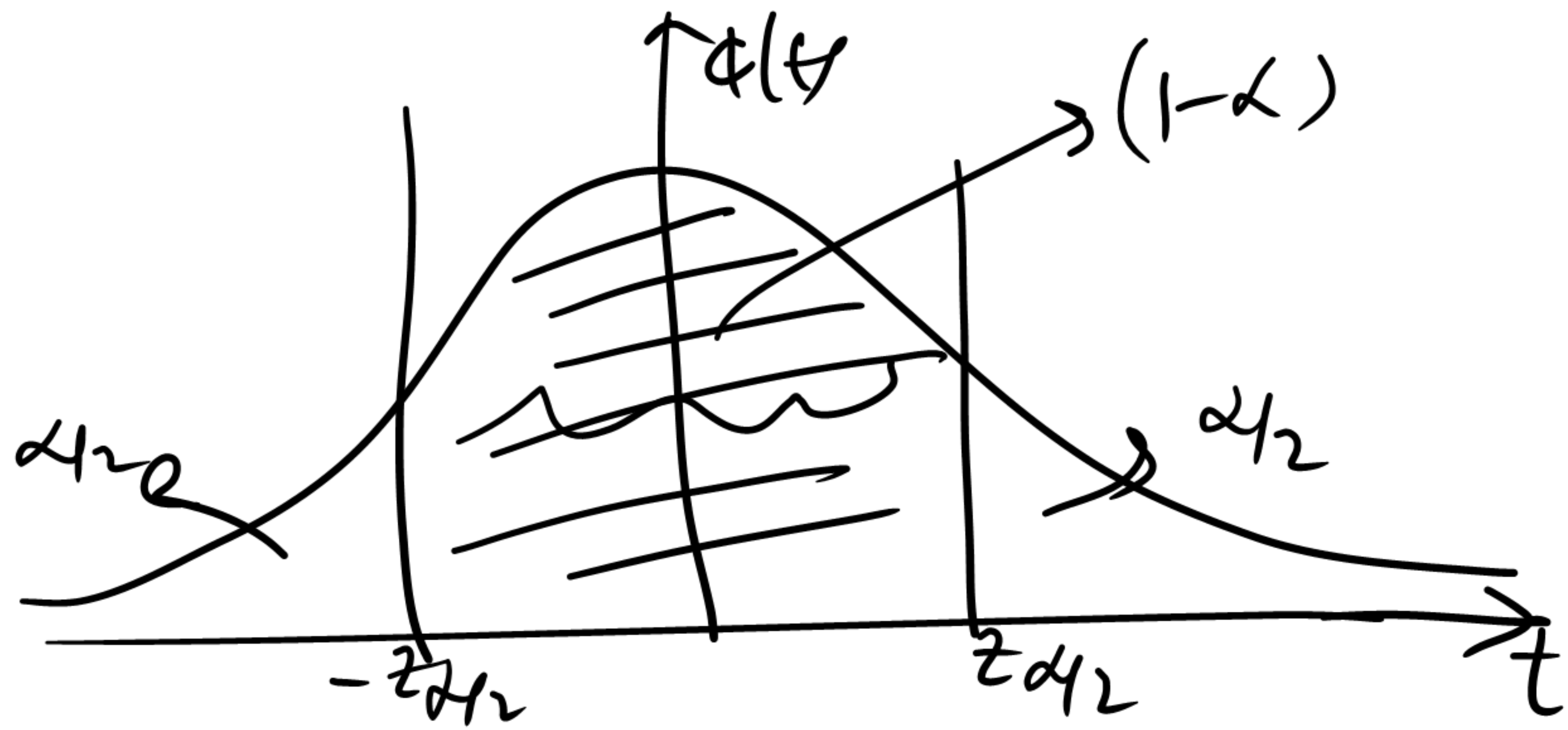
Confidence Interval for Mean of a Normal distⁿ.

Let $X_1, \ldots, X_n$ be a random sample from $N(\mu, \sigma^2)$.

For finding confidence interval for μ, we have two cases:

1. σ^2 is known, 2. σ^2 is unknown

Case 1: σ^2 is known. In this case we use pivot as
$$Z = \frac{\sqrt{n}(\bar{X} - \mu)}{\sigma} \sim N(0,1)$$



$$P\left(-z_{\alpha/2} \le Z \le z_{\alpha/2}\right) = 1-\alpha$$

$$\Rightarrow P\left(-z_{\alpha/2} \le \frac{\sqrt{n}\,(\bar{X}-\mu)}{\sigma} \le z_{\alpha/2}\right) = 1-\alpha$$

$$\Rightarrow P\Big(\underbrace{\bar{X} - \frac{\sigma}{\sqrt{n}} z_{\alpha/2}}_{T_1(\underline{x})} \le \mu \le \underbrace{\bar{X} + \frac{\sigma}{\sqrt{n}} z_{\alpha/2}}_{T_2(\underline{x})}\Big) = 1-\alpha$$

So $\underbrace{\left(\bar{X} - \frac{\sigma}{\sqrt{n}} z_{\alpha/2},\ \bar{X} + \frac{\sigma}{\sqrt{n}} z_{\alpha/2}\right)}$ is

$100(1-\alpha)\%$ confidence integral for μ

when σ^2 is known

Example: Suppose a r.s of size $n=4$ from $N(\mu, 1)$ has sample mean $\bar{x} = 2$

To find 95% confidence interval for μ,

$z_{0.025} = 1.96$,

$$\left(\bar{x} - \frac{\sigma}{\sqrt{n}} z_{\alpha/2}, \quad \bar{x} + \frac{\sigma}{\sqrt{n}} z_{\alpha/2}\right)$$

$$\equiv \left(2 - \frac{1}{2} \cdot 1.96, \quad 2 + \frac{1}{2} \cdot 1.96\right)$$

$= \quad \underbrace{(1.02,\ 2.98)}$ is 95% C.I for μ

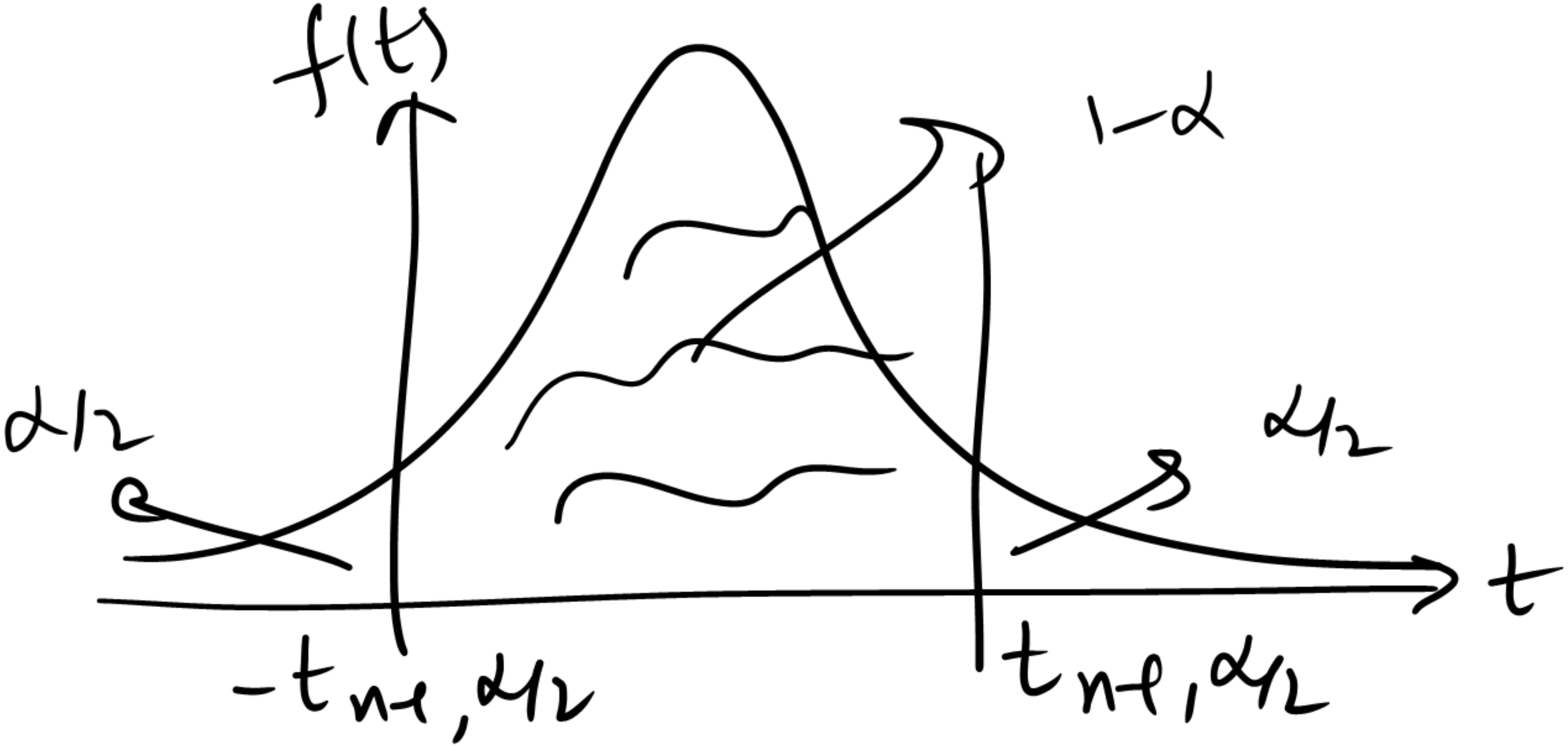
Suppose $\underline{\underline{n = 25}}$. Then C.I is

$$\left(2 - \frac{1}{5} \times 1.96 \ ,\ 2 + \frac{1}{5} 1.96\right)$$

$$= \underbrace{(1.608,\ 2.392)}$$

<u>Case 2</u>: σ^2 is unknown. We have

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



$$P\left(-t_{n-1,\alpha/2} \le T \le t_{n-1,\alpha/2}\right) = 1-\alpha$$

$$\Rightarrow P\left(-t_{n-1,\alpha/2} \le \frac{\sqrt{n}(\bar{X}-\mu)}{S} \le t_{n-1,\alpha/2}\right) = 1-\alpha$$

$$\Rightarrow P\left(\bar{X} - \frac{S}{\sqrt{n}} t_{n-1, \alpha/2} \le \mu \le \bar{X} + \frac{S}{\sqrt{n}} t_{n-1, \alpha/2}\right) = 1-\alpha$$

So $\left(\bar{X} - \frac{S}{\sqrt{n}} t_{n-1, \alpha/2}, \ \bar{X} + \frac{S}{\sqrt{n}} t_{n-1, \alpha/2}\right)$

is $100(1-\alpha)\%$ C.I. for μ when σ^2 is unknown.

Example: 10 ball bearing made from a certain process have a mean diameter

0.0506 cm & s.d. 0.004 cm. Find a 95% C.I for mean diameters of ball bearings.

Solⁿ $n=10, \quad \bar{x} = 0.0506, \quad s = 0.004$

$$t_{9,0.025} = 2.262$$

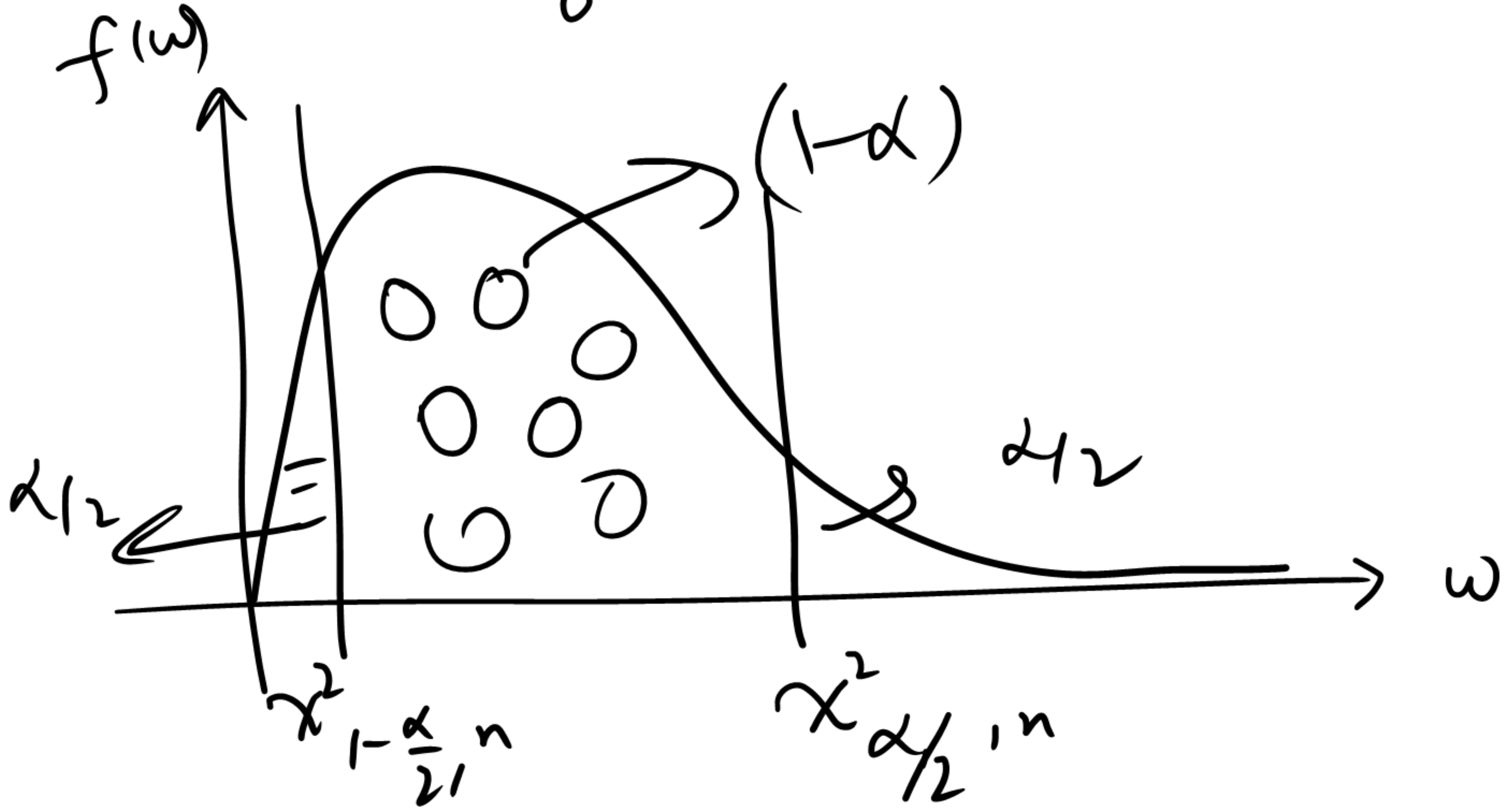
$$\bar{x} \pm \frac{s}{\sqrt{n}} t_{9,0.025} = (0.0477, 0.0535)$$

is 95% C.I for μ.

C.I. for σ^2.

Case I : μ is known

$$W = \frac{\sum (X_i - \mu)^2}{\sigma^2} \sim \chi^2_n$$



$$P\left(\chi^2_{1-\frac{\alpha}{2},n} \le W \le \chi^2_{\frac{\alpha}{2},n}\right) = 1-\alpha$$

$$\Rightarrow P\left(\chi^2_{1-\frac{\alpha}{2},n} \le \frac{\sum (X_i-\mu)^2}{\sigma^2} \le \chi^2_{\alpha/2,n}\right) = 1-\alpha$$

$$\Rightarrow P\left(\frac{\sum (X_i-\mu)^2}{\chi^2_{\frac{\alpha}{2},n}} \le \sigma^2 \le \frac{\sum (X_i-\mu)^2}{\chi^2_{1-\frac{\alpha}{2},n}}\right) = 1-\alpha$$

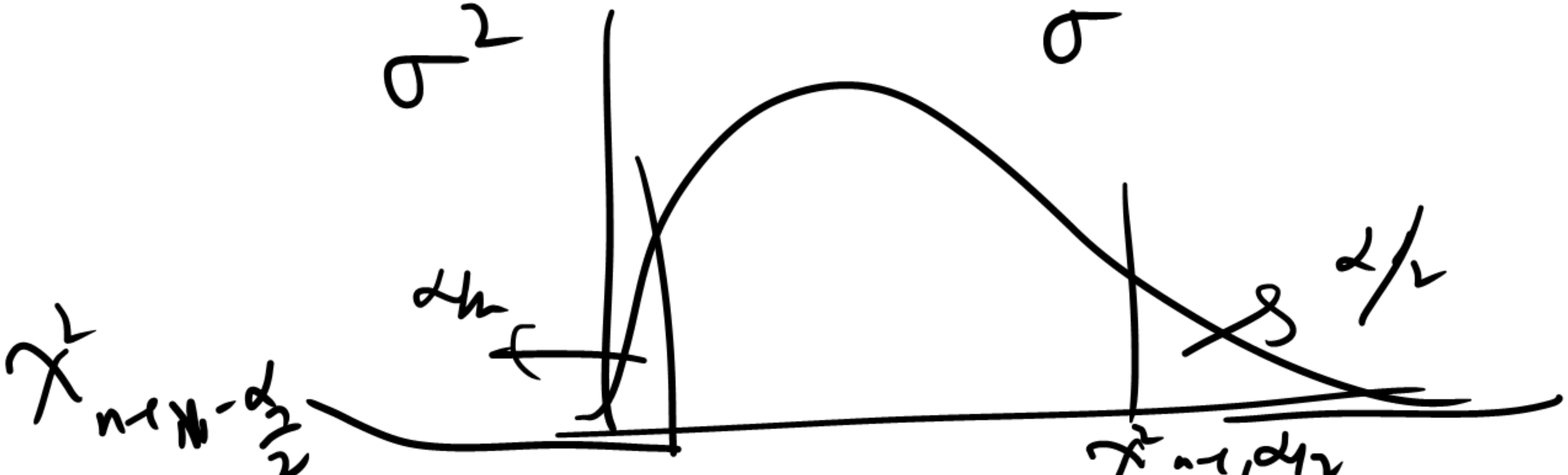
So $\left(\frac{\sum (X_i-\mu)^2}{\chi^2_{\alpha/2,n}}, \frac{\sum (X_i-\mu)^2}{\chi^2_{1-\alpha/2,n}}\right)$ is $100(1-\alpha)\%$ C.I for σ^2.

Also $\left(\sqrt{\frac{\sum(X_i-\mu)^2}{\chi^2_{\alpha/2,n}}}\ ,\ \sqrt{\frac{\sum(X_i-\mu)^2}{\chi^2_{1-\frac{\alpha}{2},n}}}\right)$ is

$100(1-\alpha)\%$ C.I. for σ. when μ is known

Case II: μ is unknown

$$W_1 = \frac{(n-1)S^2}{\sigma^2} = \frac{\sum(X_i-\bar{X})^2}{\sigma^2} \sim \chi^2_{n-1}$$



$$P\left(\chi^2_{n-1,\, 1-\frac{\alpha}{2}} \le \frac{(n-1)S^2}{\sigma^2} \le \chi^2_{n-1,\, \alpha/2} \right) = 1-\alpha$$

$$\Rightarrow P\left(\frac{(n-1)S^2}{\chi^2_{n-1,\, \alpha/2}} \le \sigma^2 \le \frac{(n-1)S^2}{\chi^2_{n-1,\, 1-\alpha/2}} \right) = 1-\alpha$$

So $\left(\frac{(n-1)S^2}{\chi^2_{n-1,\, \alpha/2}} , \frac{(n-1)S^2}{\chi^2_{n-1,\, 1-\alpha/2}} \right)$ is

$100(1-\alpha)\%$ CI for σ^2 when μ is unknown.

Ex. If 31 measurements of boiling point of sulphur have a s.d. $s = 0.83^\circ C$

find 98% C.I for the true s.d. σ.

$$\left(\sqrt{\frac{30 s^2}{\chi^2_{0.01,30}}} , \sqrt{\frac{30 s^2}{\chi^2_{0.99,30}}} \right)$$

$$\equiv \left(\sqrt{\frac{30 \times (0.83)^2}{50.89}} , \sqrt{\frac{30 \times (0.83)^2}{14.95}} \right)$$

$$= (0.6373, 1.1756)$$ is 98% C.I for σ