STA260 - Lec 3 July 9 2025

reminder: $\frac{n-1}{\sigma^2}S^2 = \sum\left(\frac{Y_i - \bar{Y}}{\sigma}\right)^2$

Test goes up to todays lecture

$$Y_1, \dots, Y_n \overset{iid}{\sim} N(\mu, \sigma^2) \Rightarrow \begin{cases} \dfrac{\bar{Y}-\mu}{\sigma/\sqrt{n}} \sim N(0,1) \\ \dfrac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1) \end{cases}$$

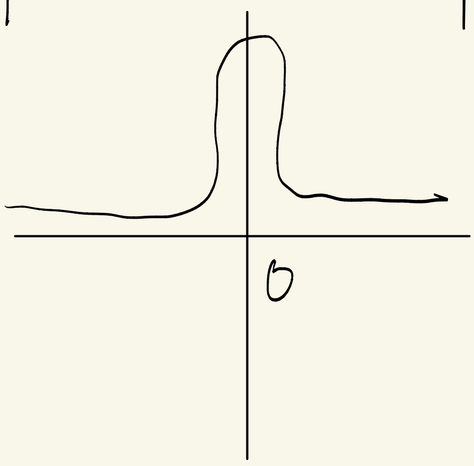
In reality, variance is not known, we must estimate μ

must use pop. std. dev. to approximate

$$\frac{\bar{Y}-\mu}{S/\sqrt{n}} \sim \text{t-distribution (Student t-dist)}$$

$$T \sim t(r): \quad f_T(t) = \frac{\Gamma\left(\frac{r+1}{2}\right)}{\Gamma\left(\frac{r}{2}\right)\sqrt{\pi r}}\left(1+\frac{t^2}{r}\right)^{-\left(\frac{r+1}{2}\right)}$$

$-\infty < t < \infty$



$r \to \infty \Rightarrow t(r) \to N(0,1)$

How to derive t random variable

$$\left.\begin{array}{l} Z \sim N(0,1) \\ W \sim \chi^2(r) \end{array}\right\} \xRightarrow{\text{indep.}} T = \frac{Z}{\sqrt{W/r}} \sim t(r)$$

$$\frac{\sqrt{n}(\bar{y}-\mu)}{S} = \frac{\frac{\sqrt{n}(\bar{y}-\mu)}{\sigma}}{S/\sigma} = \frac{\frac{\sqrt{n}(\bar{y}-\mu)}{\sigma}}{\sqrt{\frac{S^2}{\sigma^2}}} = \frac{Z}{\sqrt{\frac{W}{n-1}}}$$

$$= \frac{N(0,1)}{\sqrt{\frac{\chi^2_{(n-1)}}{n-1}}} \sim t_{(n-1)}$$

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

$$\therefore \frac{\bar{y}-\mu}{S/\sqrt{n}} \sim t(n-1)$$

$Z \sim N(0,1)$
$W \sim \chi^2(r)$
independent

$$\Longrightarrow \begin{cases} E[Z]=0 \\ Var(Z)=E[X^2]-\left(E[X]\right)^2 \\ \Longrightarrow E[Z^2]=Var(Z)+\left(E[Z]\right)^2 \\ \Longrightarrow E[Z^2]=1 \end{cases}$$

$$E[T]=E\left[\frac{Z}{\sqrt{\frac{W}{r}}}\right]$$

$$=\overset{0}{\cancel{E[Z]}}\cdot E\left[\frac{1}{\sqrt{\frac{W}{r}}}\right]=0$$

In general!
$E[XY]\neq E[X]E[Y]$

$$\Longrightarrow E[T]=0$$

$$Var(T)=E[T^2]-\left(\overset{0}{\cancel{E[T]}}\right)^2=E[T^2]=E\left[\frac{Z^2}{W/r}\right]$$

$$=E[Z^2]\cdot E\left[\frac{1}{W/r}\right]$$

$$=E\left[\frac{1}{W/r}\right]$$

$$=rE\left[\frac{1}{W}\right]$$

$$E\left[\frac{1}{X}\right]\neq\frac{1}{E[X]}$$ Jensen's Inequality

$$E[W^k]=\frac{\Gamma\left(\frac{r}{2}+k\right)2^k}{\Gamma\left(\frac{r}{2}\right)}$$

$$=rE[W^{-1}]$$

$$=r\,\frac{\Gamma\left(\frac{r}{2}-1\right)2^{-1}}{\Gamma\left(\frac{r}{2}\right)}$$

$$=\frac{r}{2}\left(\frac{\Gamma\left(\frac{r}{2}-1\right)}{\Gamma\left(\frac{r}{2}\right)}\right)$$

STAC56 identity: $\Gamma(z) = (z-1)\Gamma(z-1)$ {always imagine works}

$\Gamma((z-1)+1) =$

$$\neg \ldots\ldots = \frac{\left(\frac{r}{2}\right) \cancel{\Gamma\left(\frac{r}{2}-1\right)}}{\left(\frac{r}{2}-1\right)\cancel{\Gamma\left(\frac{r}{2}-1\right)}} = \frac{r}{r-2}$$

Ex:

$$Y_1, Y_2, \ldots, Y_5 \overset{iid}{\sim} N(0,1)$$
$$Y_6 \overset{iid}{\sim} N(0,1)$$
indep $\quad W = \sum_{i=1}^{5} Y_i^2$

Find dist of $\dfrac{\sqrt{5}\, Y_6}{\sqrt{W}}$

$$Y_i \sim N(0,1) \Rightarrow (Y_i)^2 \sim \chi^2(1) \Rightarrow \sum_{i=1}^{5} Y_i^2 \sim \chi^2(5)$$

So $\dfrac{\sqrt{5}\, Y_6}{\sqrt{W}} = \dfrac{Y_6}{\sqrt{\frac{W}{5}}} \equiv \dfrac{N(0,1)}{\sqrt{\frac{\chi^2(5)}{5}}} \sim t(5) \Rightarrow \dfrac{\sqrt{5}\, Y_6}{\sqrt{W}} \sim t(5)$

deg of freedom

Next up: F distribution -> underlying dist of ratios of variables

$\sigma^2 = 1$ -> whether its high or low depends

absolute quantities : Ex: μ, Mean, etc

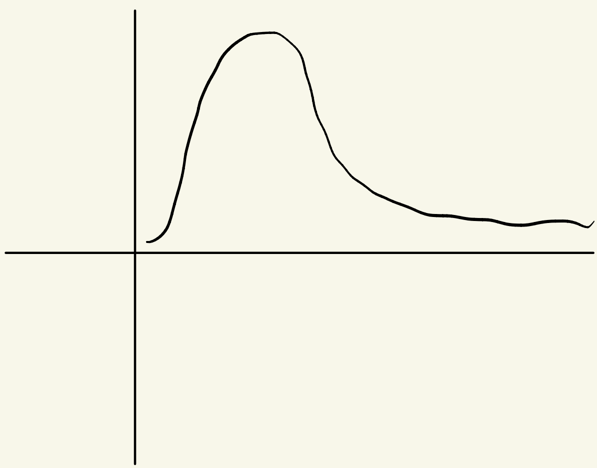
relative quantities : variance is a relative quantity

$\frac{X (m)}{Y (m)}$ } compare r.v. = unitless number $\therefore$ they cancel out

numerator, denominator

$X \sim F(r_1, r_2)$; $f(x) = \dfrac{\Gamma\left(\frac{r_1+r_2}{2}\right)\left(\frac{r_1}{r_2}\right)^{\frac{r_1}{2}} x^{\frac{r_1}{2}-1}}{\Gamma\left(\frac{r_1}{2}\right)\Gamma\left(\frac{r_2}{2}\right)\left(1+\frac{r_1 x}{r_2}\right)^{\frac{(r_1+r_2)}{2}}}$

$x > 0$



$$\left.\begin{array}{l} W_1 \sim \chi^2(r_1) \\ W_2 \sim \chi^2(r_2) \end{array}\right\rangle \text{independent} \Bigg\} \Rightarrow F = \frac{W_1/r_1}{W_2/r_2} \sim F(r_1, r_2)$$

$$\frac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2} \sim ? \qquad W_1 = \frac{(n_1-1)S_1^2}{\sigma_1^2} \sim \chi^2(n_1-1)$$

we have two independent populations with variances σ_1^2 and σ_2^2 with sample sizes n_1 and n_2

$$W_2 = \frac{(n_2-1)S_2^2}{\sigma_2^2} \sim \chi^2(n_2-1)$$

$$\Rightarrow \frac{W_1}{n_1-1} = \frac{S_1^2}{\sigma_1^2}, \quad \frac{W_2}{n_2-1} = \frac{S_2^2}{\sigma_2^2}$$

$$\therefore \frac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2} \sim \frac{\frac{W_1}{n_1-1}}{\frac{W_2}{n_2-1}} \equiv \frac{\frac{\chi^2(n_1-1)}{n_1-1}}{\frac{\chi^2(n_2-1)}{n_2-1}}$$

$$\sim F(n_1-1, n_2-1)$$

Exp. Two indep. samples with $n_1=6$ and $n_2=10$, which come from normal populations.

Find b such that $P\left(\frac{S_1^2}{S_2^2} \le b\right) = 0.95$

Assume that $\sigma_1^2 = \sigma_2^2$

$$\frac{\frac{S_1^2}{\sigma_1^2}}{\frac{S_2^2}{\sigma_2^2}} \sim F(n_1-1, n_2-1) \Rightarrow \frac{S_1^2}{S_2^2} \sim F_{(n_1-1, n_2-1)}$$

$$\Rightarrow \frac{S_1^2}{S_2^2} \sim F(5,9)$$

$$P(F \le b) = 0.95 \Rightarrow P[F \geq b] = 0.05$$

After break: $P[F \geq b] = 0.05 \Rightarrow b = 3.48$

Q1: $Y \sim F(r_1, r_2) \Rightarrow \frac{1}{Y} \sim ?$

$$Y \sim F(r_1, r_2) \Rightarrow Y = \frac{W_1/r_1}{W_2/r_2}$$

$$\frac{1}{Y} = \frac{W_2/r_2}{W_1/r_1} \sim F(r_2, r_1)$$

$W_1 \sim \chi^2(r_1)$

$W_2 \sim \chi^2(r_2)$

indep

Q. $T \sim t(r) \Rightarrow T^2 \sim ?$

$$T \sim t(r) \Rightarrow T = \frac{Z^2}{\sqrt{\frac{W}{r}}} \Rightarrow T = \frac{Z^2}{W/r}$$

$W \sim \chi^2(r)$

Z and W are indep

$$\equiv \frac{Z^2(1)/1}{\chi^2(r)/r} \sim F(1, r)$$

Q. $E[F] = ?$ $\quad F = \dfrac{W_1/r_1}{W_2/r_2}$

$$= \left(\frac{r_2}{r_1}\right) E\left[\frac{W_1}{W_2}\right]$$

W_1 and W_2 indep.

$$= \left(\frac{r_2}{r_1}\right) E[W_1] E\left[\frac{1}{W_2}\right]$$

$$= \left(\frac{r_2}{r_1}\right)(r_1)\left(\frac{1}{r_2 - 1}\right)$$

$$= \frac{r_2}{r_1 - 1}$$

$E[F^k]$ kth moment $\rightarrow$ useful for finding estimators

$F \sim F(r_1, r_2)$

$$E[F^k] = \left[\left(\frac{W_1/r_1}{W_2/r_2}\right)^k\right]$$

$$= \left(\frac{r_2}{r_1}\right)^k E\left[\frac{W_1^k}{W_2^k}\right] = \left(\frac{r_2}{r_1}\right)^k E[W_1^k] E[W_2^{-k}]$$

$$= \left(\frac{r_2}{r_1}\right)^k \frac{\Gamma\left(\frac{r_1}{2}+k\right)2^k}{\Gamma\left(\frac{r_1}{2}\right)} \frac{\Gamma\left(\frac{r_2}{2}-k\right)2^{-k}}{\Gamma\left(\frac{r_2}{2}\right)}$$

$$= \left(\frac{r_2}{r_1}\right)^k \frac{\Gamma\left(\frac{r_1}{2}+k\right)\Gamma\left(\frac{r_2}{2}-k\right)}{\Gamma\left(\frac{r_1}{2}\right)\Gamma\left(\frac{r_2}{2}\right)}$$

$$\left.\begin{array}{l} Y_1, Y_2, \dots, Y_5 \overset{iid}{\sim} N(0,1) \\ Y_6 \overset{iid}{\sim} N(0,1) \end{array}\right\} \text{indep} \quad U = \sum_{i=1}^{5} (Y_i - \bar{Y})^2$$

what is dist of $\dfrac{2(5\bar{Y}^2 + Y_6^2)}{U}$

$$\bar{Y} \sim N\left(0, \frac{1}{5}\right)$$

$$X \sim N(\mu, \sigma^2) \Rightarrow$$
$$kX \sim N(k\mu, k^2\sigma^2)$$

k scale

$$\sqrt{5}\,\bar{Y} \sim N(0,1)$$

$$(\sqrt{5}\,\bar{Y})^2 = 5\bar{Y}^2 \sim \chi^2_{(1)}$$

$$5\bar{Y}^2 + Y_6^2 \equiv \chi^2_{(1)} + \chi^2_{(1)} \sim \chi^2_{(2)}$$

$$S^2 = \frac{1}{n-1} \sum (Y_i - \bar{Y})^2$$

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

$$(n-1)S^2 \sim \chi^2(n-1)$$

$$\Rightarrow (n-1)\frac{1}{(n-1)} \sum_{i=1}^{5} (Y_i - \bar{Y})^2$$

$$\sim \chi^2(4)$$

$$\frac{2(5\bar{Y}^2 + Y_6^2)}{U} \equiv \frac{2\chi^2(2)}{\chi^2(4)}$$

$$\equiv \frac{\chi^2(2)/2}{\chi^2(4)/4} \sim F(2,4)$$

In Summary: $Y_1, Y_2, \ldots, Y_n \overset{iid}{\sim} N(\mu, \sigma^2) \Rightarrow \begin{cases} \dfrac{\bar{Y}-\mu}{\sigma/\sqrt{n}} \sim N(0,1) \\ \dfrac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1) \\ \dfrac{\bar{Y}-\mu}{S/\sqrt{n}} \sim t(n-1) \end{cases}$

$$\frac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2} \sim F_{(n_1-1, n_2-1)}$$

Final topic: Central Limit Theorem

$Y_1, Y_2, \ldots, Y_n \overset{iid}{\sim} \mu, \sigma^2 \rightarrow$ don't care about distributions

$$\frac{Y_n - E[Y_n]}{\sqrt{\operatorname{var}(Y_n)}} \sim N(0,1) \quad n \rightarrow \infty$$

$E[\bar{Y}] = \mu$

$\operatorname{var}(\bar{Y}) = \dfrac{\sigma^2}{n}$ $\Rightarrow \dfrac{\bar{Y} - E[\bar{Y}]}{\sqrt{\operatorname{var}(\bar{Y})}} = \dfrac{\bar{Y}-\mu}{\sigma/\sqrt{n}} \sim N(0,1)$

One of the most elegant theorems in mathematics | Read the proof at least once in your lifetime

Normal Dist. to Binomial Dist

$X \sim \text{Binomial}(n, p): p(x) = \binom{n}{x} p^x (1-p)^{n-x}$

$\binom{1000}{100}$ ← hard to compute

Corrections for approximating binomial

Rule 1: $P(Y=b) = P(\underbrace{b-0.5 \leq y \leq b+0.5}_{\text{Care about intervals only}})$

Rule 2:

$P(Y \leq b) = P(Y \leq b+0.5)$

Rule 3: $P(Y \geq b) = P(Y \geq b-0.5)$

Q. we have 200 files, each gets damaged with prob 0.2 (independent of each other), what is the prob that fewer than 50 files get damaged

A: Let $A :=$ r.v. denoting files that got damaged

$$A \sim \text{Binomial}(200, 0.2)$$

$$P[X < 50] = P[X \leq 49]$$

$$E[X] = np = (200)(0.2) = 40$$

$$\text{Var}(X) = np(1-p) = 40(1-0.2) = (0.8)(40) = 32$$

$$\to N(40, 32)$$

$$\simeq P[Y \leq 49] \simeq P[Y \leq 49 + 0.5]$$

$$\simeq P[Y \leq 49.5]$$

$$= P\left[\frac{Y - E[Y]}{\sqrt{\text{Var}(Y)}} \leq \frac{49.5 - 40}{\sqrt{32}}\right]$$

$$= P\left[Z \leq \frac{49.5 - 40}{\sqrt{32}}\right] = 0.9525$$

$X_1, X_2, \ldots, X_n \overset{iid}{\sim} N(\mu, \sigma^2) \quad S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$

Show that $E[(cS^2 - \sigma^2)^2]$ is minimized only when $c = \frac{n-1}{n+1}$

$$\mathrm{var}((cS^2 - \sigma^2)) = E[(cS^2 - \sigma^2)^2] - (E[cS^2 - \sigma^2])^2$$

$$E[((cS^2 - \sigma^2))^2] = \mathrm{var}[(cS^2 - \sigma^2)] + (E[cS^2 - \sigma^2])^2$$

$$= E[(cS^2 - \sigma^2)^2] = c^2\mathrm{var}(S^2) + (cE[S^2] - \sigma^2)^2$$

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

$$W \sim \chi^2(r) \Rightarrow \begin{cases} E[W] = r \\ \mathrm{var}(W) = 2r \end{cases}$$

$$\mathrm{var}\left(\frac{(n-1)S^2}{\sigma^2}\right) = 2(n-1) \Rightarrow \frac{(n-1)^2}{\sigma^4}\mathrm{var}(S^2) = 2(n-1)$$

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

$$\Rightarrow E\left[(cS^2-\sigma^2)^2\right] = c^2\left(\frac{2\sigma^4}{n-1}\right) + (c\sigma^2-\sigma^2)^2$$

$$= \left(\frac{2c^2}{n-1}\right)\sigma^4 + (c-1)^2\sigma^4$$

$$= \sigma^4\left(\frac{2c^2}{n-1} + (c-1)^2\right)$$

$$\frac{\partial}{\partial c} E\left[(cS^2-\sigma^2)^2\right] = \left[\frac{4c}{n-1} + 2(c-1)\right]\sigma^4$$

$$\frac{\partial}{\partial c} E\left[(cS^2-\sigma^2)^2\right] = 0 \Rightarrow \left[\frac{4c}{n-1} + 2(c-1)\right]\sigma^4 = 0$$

$$\Rightarrow \frac{4c}{n-1} + 2(c-1) = 0 \qquad \star \quad C = \frac{n-1}{n+1}$$

$$\Rightarrow \left(\frac{4}{n-1}\right)c - 2 = 0 \Rightarrow$$