Next lecture - Ch 8

"Parameter" = a value that describes a population

⇒ this is a constant value, but is typically unknown in practice.

e.g., population mean

"unbiased" used when expected value of statistic is equal to value of parameter.

~~Why~~ we not help each other?

Ch. 8: Statistics and Sampling Distribution

§8.2 - Statistics

Def: A statistic is a fcn of observable random variables $T = t(X_1, \dots, X_n)$ that is free of unknown parameters.

function of R.V.s ; a R. sample ;

I.e., T is a fcn of your random sample.

⇒ T is a R variable.

"Statistic" = a value that describes a sample.

⇒ is a random variable.

- values can change from sample to sample.

common examples of statistics are sample mean, sample variance

Ex. $T = t(X_1, \dots, X_n)$

$$= \frac{\sum_{i=1}^{n} X_i}{n} = \bar{X} \text{ is a r.v.}$$

little $\rightarrow \bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}$ = a single value (of $\bar{X}$), not a R. variable.

- Properties of Sample mean, $\bar{X}$, and Sample Variance, S^2

size of n irrelevant

I) If $X_1, X_2, \dots, X_n$ form a r.s. from a population w/ pdf $f(x)$ with mean μ and variance σ^2, then $E(\bar{X}) = \mu$ and $Var(\bar{X}) = \frac{\sigma^2}{n}$ = parameter (a constant)

$E(\bar{X})$ = statistic ; μ = population mean = parameter ; $Var(\bar{X})$ = statistic

mean of all the sample means

Note: Since $E(\bar{X}) = \mu$, $\bar{X}$ is unbiased for μ.

no tendency to over-/under-estimate pop. mean μ

II) Sample Variance $S^2 = \dfrac{\sum_{i=1}^{n}(X_i - \overline{X})^2}{n-1}$ ← measures average spread

the mean, cuz is a sample

$x_i - \overline{X}$ is distance from mean

$$= \frac{\sum_{i=1}^{n} X_i^2 - \frac{\left(\sum_{i=1}^{n} X_i\right)^2}{n}}{n-1}$$

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

$\dfrac{\sum_{i=1}^{n} X_i}{n} = \overline{X}$

$\Rightarrow \sum_{i=1}^{n} X_i = n\overline{X}$

"from $f(x)$" (AKA from some population)

Thm: If $X_1, \dots, X_n$ form a r.s. from $f(x)$ with mean μ and variance σ^2, then

$$\underbrace{E(S^2)}_{\text{statistic}}^{\text{sample var}} = \underbrace{\sigma^2}_{\text{parameter}}^{\text{pop. var}} \text{ and } Var(S^2) = \frac{\mu_4 - \frac{n-3}{n-1}\sigma^4}{n}, \quad n > 1$$

where $\mu_4 = E(X-\mu)^4 = 4^{th}$ moment about the mean

note: Since $E(S^2) = \sigma^2$, S^2 is unbiased for σ^2.

(if can take all possible values of size n for every sample you compute sample variance, then take the mean of all those sample variances, should get the pop. variance)

Conclusion: sample mean and sample variance are unbiased for μ and σ^2, respectively

Pf: $E(S^2) = \overset{\text{lin operator}}{E}\left(\dfrac{\sum_{i=1}^{n} X_i^2 - n(\overline{X})^2}{n-1}\right)$

$$= \frac{1}{n-1}\left[E\left(\sum_{i=1}^{n} X_i^2\right) - E(n\overline{X}^2)\right]$$

$$= \frac{1}{n-1}\left[\sum_{i=1}^{n} E(X_i^2) - n\,E(\overline{X}^2)\right]$$

cuz X_i's identical

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

constants

$$= \frac{1}{n-1}\left[n(\sigma^2 + \mu^2) - \sigma^2 - n\mu^2\right] = \frac{1}{n-1}\left[n\sigma^2 + \cancel{n\mu^2} - \sigma^2 \dots\right]$$

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

we use n-1 in denom cuz gives us unbiasedness in the end.

Recall:

$Var(X) = E(X^2) - [E(X)]^2$

$\Rightarrow E(X^2) = Var(X) + [E(X)]^2 = \sigma^2 + \mu^2$

some R.V.

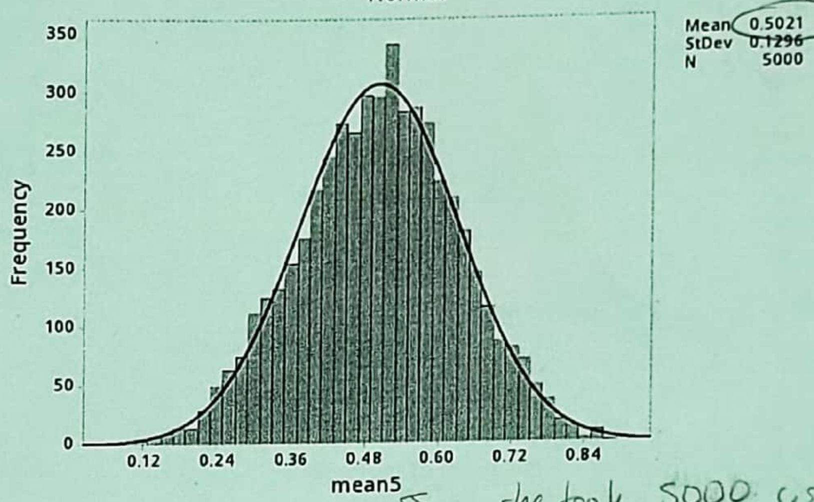
$Var(\overline{X}) = E(\overline{X}^2) - [E(\overline{X})]^2$

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

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

Distribution of Mean
Sampling from Uniform (0,1)

Histogram of mean5
Normal

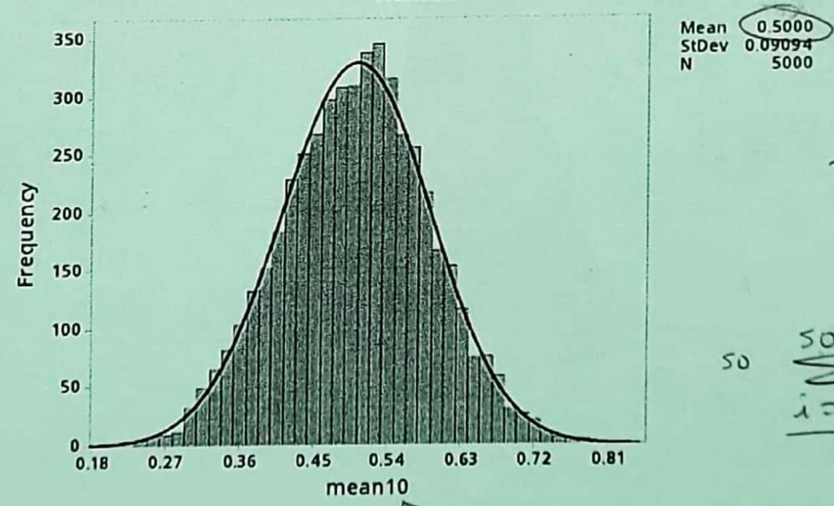


exact value is 0.5

~~get closer to it as n↑'s.~~

she took 5000 r.smpls of size 5 and for each r.s., she computed $\bar{x}_1, \bar{x}_2, \ldots \bar{x}_{5000}$

Histogram of mean10
Normal



...of size 10

where

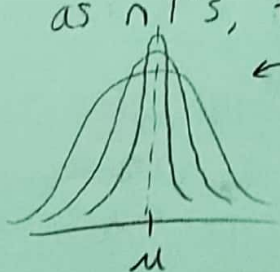
$$\bar{x}_1 = \frac{x_{11} + x_{12} + x_{13} \cdots x_{15}}{5}$$

$$\bar{x}_2 = \frac{x_{21} + x_{22} + \cdots + x_{25}}{5}$$

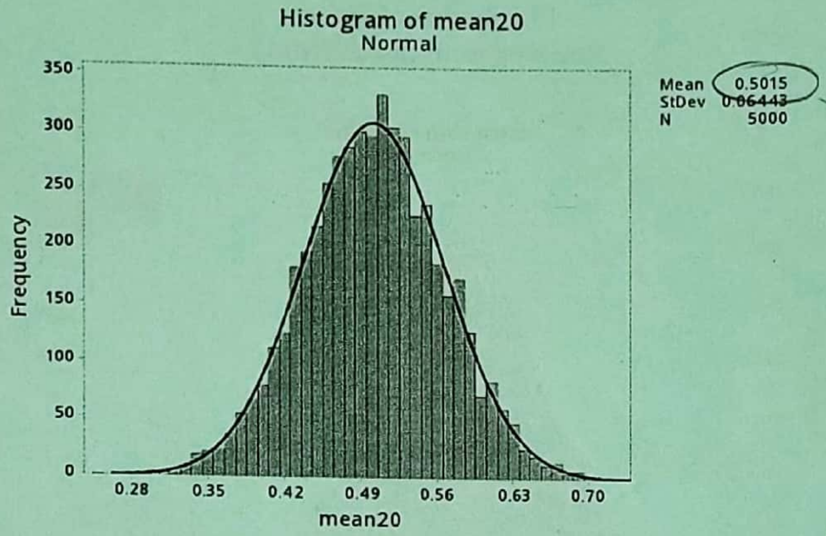
$\vdots$

so $$\frac{\sum_{i=1}^{5000} \bar{x}_i}{5000} = \mu$$

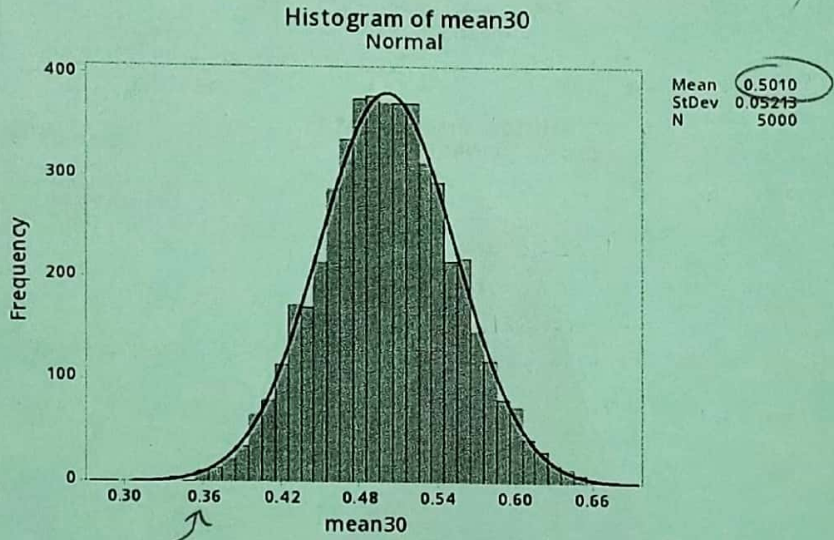
as n↑'s, the variance of $\bar{X}$ gets smaller.



gets taller & narrower as n↑'s; closing in on the actual value.



size of n irrelevant for getting closer to exact value



With larger n, can no longer get something as low as 0.2, for instance. i.e., our std dev gets smaller & smaller as n ↑'s.

taking $n \to \infty$ implies talking about whole population, so there should no longer be any variability/variance

Ch. 8: Start w/ Normal $X_i \sim N(\mu_i, \sigma_i^2)$ indep., a_i are constants

§8.3 – Sampling Distribution

Def: The distribution of a statistic is called a <u>sampling distribution.</u>

– Focus on distributions derived from functions of Normal R.V.s.

Linear combination of normal r.v.'s: each X_i has its own mean and variance (are not identical) at least, not necessarily

- **Thm:** If $X_i \sim N(\mu_i, \sigma_i^2)$, $i = 1, 2, \dots, n$ and X_i's are <u>independent</u>,

  then $Y = \sum_{i=1}^n a_i X_i \sim N\left(\sum_{i=1}^n a_i \mu_i, \sum_{i=1}^n a_i^2 \sigma_i^2\right)$ where a_i's are constants.

linear combination of normal R.V.s. ✩✩ exact dist, not limiting dist., cuz start with NORMAL, end w/ NORMAL.

Pf: Use MGF technique. Find $M_{\sum_{i=1}^n a_i X_i}(t)$.

$M_{aX+b}(t) = e^{bt} M_X(at)$

$$M_{\sum_{i=1}^n a_i X_i}(t) = \prod_{i=1}^n M_{a_i X_i}(t) \quad \text{since } X_i\text{'s indep.}$$

$$= \prod_{i=1}^n M_{X_i}(a_i t)$$

$M_{X_i}(t) = e^{\mu_i t + \frac{\sigma_i^2 t^2}{2}}$ (know cuz each is normal)

$$= \prod_{i=1}^n e^{\left[\mu_i a_i t + \frac{\sigma_i^2}{2}(a_i t)^2\right]}$$

$$= e^{\left[t\underbrace{\left(\sum_{i=1}^n a_i \mu_i\right)}_{\mu} + \frac{t^2}{2}\underbrace{\left(\sum_{i=1}^n a_i^2 \sigma_i^2\right)}_{\sigma^2}\right]}$$ which is the MGF of

a normal dist. with $N\left(\sum a_i \mu_i, \sum a_i^2 \sigma_i^2\right)$.

∎

[See first green handout]

- Corollary: $X_i \sim N(\mu, \sigma^2)$, independent, (and identical) (i.e., have a r.s.) $i = 1, 2, \dots, n$.

  Then $\bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$. $\left[\text{cuz } \bar{X} = \frac{\sum X_i}{n} = \sum \frac{X_i}{n} \text{ (or } a_i\text{'s are } \frac{1}{n}\text{)}\right]$

start w/ Normal?

Any lin. comb. will be Normal.

- Some functions of normal R.V.s can result in other distributions.

- Recall: $X \sim GAM(\underset{\text{scale}}{\theta}, \underset{\text{shape}}{K})$, where $\leftarrow$ (exponential when $K=1$)

pdf is $f(x) = \dfrac{1}{\theta^K \Gamma(K)} x^{K-1} e^{-\frac{x}{\theta}}$, $\quad x>0,\ K>0,\ \theta>0$,

where $\Gamma(K) = \int_0^\infty t^{K-1} e^{-t}\,dt$ which has properties

i) $\Gamma(K) = (K-1)\Gamma(K-1)$ for $K>1$,

ii) if $n \in \mathbb{Z}$, then $\Gamma(K) = (K-1)!$,

iii) $\Gamma(\frac{1}{2}) = \sqrt{\pi}$.

- Special Case: $\theta = 2,\ K = \frac{v}{2}$, $\leftarrow v \in \mathbb{Z}$

where in this case we will get a Chi-square dist. with v degrees of freedom. I.e.,

$Y \sim \chi^2(v)$. $\leftarrow$ have readily-available tables for the χ^2-dist CDF, but not for the GAM dist.

$\sim GAM(2, \frac{v}{2})$ — same thing

[See green 2nd handout for properties]

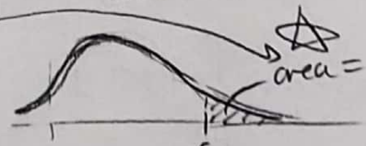
Ex: $Y \sim \chi^2(6)$. χ^2 is cont. and its skewness depends

1) Find $P(Y \leq 5)$ (use table; $v=6$, $c=5$)

$= \boxed{0.456}$ by Table 5. (pale yellow)

2) $P(Y > 8.6) = 1 - P(Y \leq 8.6) = 1 - (0.803) = \boxed{0.197}$

3) Find c such that $P(Y > c) = 0.05$.



$= \chi^2_{0.95}(6)$

area to the left of $0.95 = c$ w/ 6 deg. of freedom.

AKA 95th percentile with $v=6$.

- Table 4 (blue) has percentiles.
- find the 95th percentile. $\leftarrow$ AKA

$\Rightarrow \chi^2_{0.95}(6) = \boxed{12.59}$

4) Find the median of Y.

$= \chi^2_{0.5}(6) = \boxed{5.35}$

§8.3 <u>first handout</u>

STAT 480B

Example for Section 8.3

Suppose $X_1 \sim N(5,2)$, $X_2 \sim N(2,3)$, $X_3 \sim N(-1,4)$, and $X_i, i = 1,2,3$, are independent. Determine the probability that X_1 is more than twice the sum of X_2 and X_3.

$$\left.\begin{array}{l} X_1 \sim N(\overset{\mu_i}{5},\overset{\sigma_i^2}{2}) \\ X_2 \sim N(2,3) \\ X_3 \sim N(-1,4) \end{array}\right\} \text{independent}$$

· Find $P(X_1 > 2(X_2+X_3))$.

write R.V.s on one side (everything w/ an X), constants on other side.

$= P(\underbrace{X_1 - 2X_2 - 2X_3}_{\text{call this } Y} > 0)$ what is the distribution of this? (treat like one R.V.)

· I.e., what is the dist. of Y? (Then find $P(Y>0)$.)

$Y \sim N$ with a_i constants are $1, -2$, and -2 → i.e., $\underline{a_1 = 1}$, $\underline{a_2 = -2}$, $\underline{a_3 = -2}$

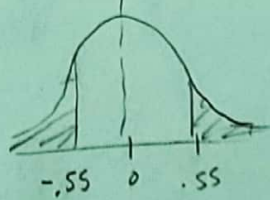
mean, $\mu = 5 - 2(2) - 2(-1) = 3$,

variance, $\sigma^2 = \overset{a_1^2}{(1)^2}\overset{\sigma_1^2}{(2)} + \overset{a_2^2}{(-2)^2}\overset{\sigma_2^2}{(3)} + \overset{a_3^2}{(-2)^2}\overset{\sigma_4^2}{(4)} = 30$.

$= P\left(Z > \frac{0-(3)}{\sqrt{(30)}}\right)$

$\frac{-3}{\sqrt{30}} =$

$= 1 - P(Z < -0.55)$



$= \cancel{1} - [\cancel{1} - P(Z < 0.55)]$

$= P(Z < 0.55)$

$= \boxed{0.7088}$.

$\sqrt{30}$

§8.3 2nd handout

STAT 480B

Properties of χ^2 Distribution

prove by taking MGF of a gamma dist and letting $\theta = 2$, $K = \frac{v}{2}$

Theorem 1: *If $Y \sim \chi^2(v)$, then*

~~Ask~~ Can we use $M_Y^{(r)}(t)\big|_{t=0}$ to prove ? Yeah but it's probably harder.

1) $M_Y(t) = (1-2t)^{-v/2}$

2) $\mathbf{E}(Y^r) = \dfrac{2^r\Gamma(\frac{v}{2}+r)}{\Gamma(\frac{v}{2})}$ try to prove? (practice)

prove for practice..?

3) $\mathbf{E}(Y) = v$ i.e., expected value is equal to its parameter v, the deg. of freedom.

4) $\mathbf{Var}(Y) = 2v$

Y is χ^2 dist, which is a gamma dist, whose support is 0 to ∞.

2) $E(Y^r) = \int_0^\infty y^r f_Y(y)\,dy$ [using def.]

$$= \int_0^\infty y^r \frac{1}{2^{v/2}\Gamma(\frac{v}{2})} y^{\frac{v}{2}-1} e^{-\frac{y}{2}}\,dy$$

$$= \frac{1}{2^{v/2}\Gamma(\frac{v}{2})}\int_0^\infty y^{r+\frac{v}{2}-1}\cdot e^{-\frac{y}{2}}\,dy$$

use change of variables: Let $u = \frac{y}{2}$. Then $du = \frac{1}{2}dy$

$\Rightarrow 2du = dy$

$\Rightarrow y = 2u$

so $u_1 = \frac{0}{2} = 0$

$u_2 = \frac{\infty}{2} = \infty$

$$= \frac{1}{2^{v/2}\Gamma(\frac{v}{2})}\int_0^\infty (2u)^{r+\frac{v}{2}-1}\cdot e^{-(u)}(2du)$$

$$= \frac{2^{r+\cancel{v/2}}}{\cancel{2^{\frac{v}{2}}}\Gamma(\frac{v}{2})}\underbrace{\int_0^\infty u^{r+\frac{v}{2}-1} e^{-u}\,du}_{=\Gamma(r+\frac{v}{2})}$$

force the integral to look like something we recognize.

o.w., int. by parts indefinitely,

$$= \boxed{\frac{2^r\Gamma(r+\frac{v}{2})}{\Gamma(\frac{v}{2}).}}$$

So $E(Y) = E(Y^{(1)}) = \dfrac{2^1\Gamma(1+\frac{v}{2})}{\Gamma(\frac{v}{2})} = \dfrac{2\,\frac{v}{2}\cancel{\Gamma(\frac{v}{2})}}{\cancel{\Gamma(\frac{v}{2})}} = \frac{2v}{2} = v.$

Pf of 4) $Var(Y) = E(Y^2) - [E(Y)]^2$.

Using (2), $$E(Y^2) = \frac{2^2 \cdot \Gamma(\frac{v}{2}+2)}{\Gamma(\frac{v}{2})} = \frac{4 \cdot [(\frac{v}{2}+2)-1] \cdot \Gamma(\frac{v}{2}+2-1)}{\Gamma(\frac{v}{2})}$$

$$= \frac{4(\frac{v}{2}+1) \cdot \Gamma(\frac{v}{2}+1)}{\Gamma(\frac{v}{2})} = \frac{4(\frac{v}{2}+1) \cdot [(\frac{v}{2}+1)-1] \cdot \Gamma(\frac{v}{2}\cancel{+1}\cancel{-1})}{\Gamma(\frac{v}{2})}$$

$$= \frac{4(\frac{v}{2}+1) \cdot (\frac{v}{2}) \cdot \cancel{\Gamma(\frac{v}{2})}}{\cancel{\Gamma(\frac{v}{2})}} = 4(\frac{v}{2}+1)(\frac{v}{2})$$

$$= (\frac{4v}{2}+4) \cdot (\frac{v}{2}) = \frac{4v^2}{4} + \frac{4v}{2} = v^2 + 2v;$$

by (3) we already have (and otherwise can derive the fact that)

$E(Y) = v$.

Thus,

$$Var(Y) = E(Y^2) - [E(Y)]^2$$
$$= \cancel{v^2} + 2v - \cancel{(v)^2}$$
$$= 2v.$$

∎

☆

<u>Thm:</u> If $X \sim GAM(\theta, \kappa)$, then $Y = \frac{2X}{\theta} \sim \chi^2(\underbrace{2\kappa}_{=\nu \text{ (deg. of freedom)}})$.

How to transform ~~any~~ any $X \sim GAM$ into a χ^2 dist equivalent.

<u>Pf:</u> Use MGF technique.

$$M_Y(t) = M_{\frac{2}{\theta}X}(t)$$

$$= M_X\left(\frac{2t}{\theta}\right) \qquad \text{Recall } M_X(t) = \left(\frac{1}{1-\theta t}\right)^{\kappa},$$

$$\text{so } M_Y(t) = \left(\frac{1}{1-\theta\left(\frac{2t}{\theta}\right)}\right)^{\kappa}$$

$$= \left(\frac{1}{1-2t}\right)^{\kappa}$$

make look like MGF of a χ^2-dist.

$$= \left(\frac{1}{1-2t}\right)^{\frac{2\kappa}{2}}.$$

ν

Hence $Y \sim \chi^2(2\kappa)$. ▨

<u>Example:</u> Sps $X \sim GAM(\overset{\theta}{4}, \overset{\kappa}{2})$.

(a) $P(X < 10)$.

- can't get a CDF if stick w/ X, integrating sucks and no table.
- transform to χ^2-dist:

$$P(X < 10) = P\left(\frac{2X}{4} < \frac{2(10)}{4}\right)$$

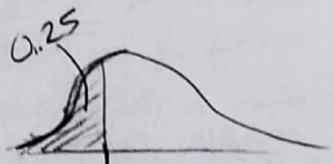
$$= P(Y < 5) \text{ where } Y \sim \chi^2(4) \quad \checkmark$$

$=2\kappa = 2(2) = 4$

- By Table 5, $P(Y<5)$ for $\nu = 4$ is $\boxed{0.713.}$

(b) Find c s.t. $P(X<c) = 0.25$.

$$P(X < c) = P\Big(\underbrace{\frac{2X}{4}}_{Y} < \frac{2(c)}{4}\Big) = 0.25$$



$\frac{c}{2} = \chi^2_{0.25}(4)$

25th percentile according to table for $\nu = 4$ is 1.92,

so $\frac{c}{2} = 1.92$

$\Rightarrow \boxed{c = 3.84.}$

Pg 284, # 5)

if take sum of iid exponentials, w/ param θ then the dist of the sum ends up being

- Given T_i = time to failure, in days, of component i.

$\sim EXP(\theta = 100)$

(a) Find dist. of $\sum_{i=1}^{10} T_i$, ← T_i's are a random sample. Call $W = \sum_{i=1}^{10} T_i$.

use MGF technique. cuz T_i's are indep.

$$M_W(t) = M_{\sum_{i=1}^{10} T_i}(t) = \prod_{i=1}^{10} M_{T_i}(t)$$ → MGF for an exponential dist given by..

$$= \prod_{i=1}^{10} \left[\frac{1}{1-\theta t}\right]$$

$$= \left[\frac{1}{1-100t}\right]^{10}$$ ← cuz identical

$\Rightarrow W \sim GAM(\theta = 100, K = 10)$.

(b) Find $P(W > 1.5(365))$. change into χ^2

Do for HW,

also do sugg. exercises

Start of Lecture on 1/31/22

- $X \sim GAM(\theta, \kappa)$, then $Y = \frac{2X}{\theta} \sim \chi^2(2\kappa)$.

sum of indep. χ^2 will also be χ^2.

- <u>Thm:</u> If $Y_i \sim \chi^2(\nu_i)$, $i = 1, 2, \ldots, n$ independent,

then $V = \sum_{i=1}^{n} Y_i \sim \chi^2\left(\sum_{i=1}^{n} \nu_i\right)$ sum of degrees of freedom...

<u>Pf:</u> Use mgf technique.

- <u>Note:</u> Unlike the normal dist., only the sum of chi-squared r.v.s turn out to be chi-square distributed.
  - so $\bar{X} = \sum \frac{X_i}{n} \not\sim \chi^2$ (cuz no longer <u>just</u> a sum of χ^2 dists)
  - however, we could still find probabilities...

e.g., Let $Y_1 \sim \chi^2(2)$, $Y_2 \sim \chi^2(3)$, $Y_3 \sim \chi^2(6)$ with Y_i, $i=1,2,3$ indep.

Find $P(\bar{Y} < 5)$.

we DO know this dist!

11 deg. of freedom

- $P(\bar{Y} < 5) = P\left(\frac{\sum Y_i}{3} < 5\right) = P\left(\sum Y_i < 15\right)$, $\sum Y_i \sim \chi^2(11)$.
- Can still use our χ^2 table to get probs of things other than <u>just</u> a sum of χ^2-dists.

<u>The point:</u> the only time a lin combination of χ^2-distributed R.V.s is itself going to be χ^2-distributed is if the coefficients are all positive 1. $(X+Y+Z+\ldots)$
The Normal dist. is much more flexible in comparison; coefficients may be $>0, <0$, etc.

- <u>Ex #8, Pg 284)</u>
  - Let $X \sim \chi^2(m)$, $Y \sim \chi^2(n)$ both indep.
  - <u>Question:</u> If $Y - X \sim \chi^2$ if $n > m$?
  - Use MGF technique.

$$M_{Y-X}(t) = M_Y(t) \cdot M_{-X}(t)$$
$$= (1-2t)^{-\frac{n}{2}} \cdot M_X(-t)$$
$$= (1-2t)^{-\frac{n}{2}} \cdot (1-2(-t))^{-\frac{m}{2}} = (1-2t)^{-\frac{n}{2}}(1+2t)^{-\frac{m}{2}}$$

no way to make this look like the MGF of a χ^2 dist.

In fact, can't make it look like <u>any</u> of our known dists!

- So the <u>difference</u> of two χ^2-dists is not one of our known special dists.

· **Thm:** If $Z \sim N(0,1)$, then $Z^2 \sim \chi^2(1)$.

Pf: Use MGF technique.

Want $M_{Z^2}(t)$. Need we definition cuz no rule for products of MGFs.

$M_{Z^2}(t) = E(e^{tZ^2})$ ← conti. R.V., so need integral

$$= \int_{-\infty}^{\infty} e^{tz^2} \cdot \underbrace{\frac{1}{\sqrt{2\pi}} e^{-z^2/2}}_{\text{pdf of } N(0,1)} \, dz$$

For $X \sim N(\mu, \sigma^2)$, $f_X(x) = \frac{1}{\sqrt{2\pi}\,\sigma} e^{-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}}$

$$= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} \cdot \underbrace{e^{-\frac{z^2}{2}(1-2t)}}_{} \, dz$$

looks like exponential part of the pdf of a $N(\underset{\mu}{0}, \underset{\sigma^2}{\frac{1}{1-2t}})$

$$= \sqrt{\frac{1}{1-2t}} \underbrace{\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sqrt{\frac{1}{1-2t}}} e^{-\frac{z^2}{2}(1-2t)} \, dz}_{}$$

pdf of $N(0, \frac{1}{1-2t})$, so integrating it over $\mathbb{R}$ means $= 1$.

$= (1-2t)^{-\frac{1}{2}}$, which is the mgf of $\chi^2(1)$. ∎

· **Cor:** If $X_1, X_2, \ldots, X_n$ form a random sample from $N(\mu, \sigma^2)$, then

$$\sum_{i=1}^{n} \frac{(X_i - \mu)^2}{\sigma^2} \sim \chi^2(n),$$ and

sum of Z^2's.

$$\frac{n(\bar{X} - \mu)^2}{\sigma^2} \sim \chi^2(1).$$

$Z_i = \frac{X_i - \mu}{\sigma} \sim N(0,1)$
$Z_i^2 \sim \chi^2(1)$,
$\sum^n Z_i \sim N(0, n)$
$\frac{1}{\sqrt{n}} \sum Z_i \sim N(0,1)$
$\sum^n Z_i^2 \sim \chi^2(n)$

Pf: $Z_i = \frac{X_i - \mu}{\sigma} \sim N(0,1)$, so $\boxed{Z_i^2 = \frac{(X_i - \mu)^2}{\sigma^2} \sim \chi^2(1)}$.

each has deg. of freedom $v = 1$, and there is a sum of n of these, so the degrees of freedom of $\sum Z_i^2$ is n

Also,

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

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

$Z^2 = \frac{(\bar{x} - \mu)^2}{\sigma^2/n} = \frac{n(\bar{X} - \mu)^2}{\sigma^2} \sim \chi^2(1)$. ∎

• If have a prob. involving Z^2, can answer using $N(0,1)$ or $\chi^2(1)$.

Ex) Have $Z \sim N(0,1)$.

Find $P(Z^2 < 4.6)$.

• First way: using std normal table: [need go from Z^2 to Z]

$$P(Z^2 < 4.6) = P(|Z| < \sqrt{4.6})$$

sqrt both sides

$$= P(-\sqrt{4.6} < Z < \sqrt{4.6})$$

$$= P(-2.14 < Z < 2.14)$$ now use normal table

$$= P(Z < 2.14) - P(Z < -2.14)$$

$$= (0.9838) - (1 - 0.9838)$$

$$= \boxed{0.9676.}$$ This is too much work.....

• Second way: using $\chi^2(1)$:

use $v=1$ cuz only have one Z r.v. in this problem.

$$P(Z^2 < 4.6) = \boxed{0.968}$$

use χ^2 table for $v=1$

independence of χ^2's crucial, not necessarily need identical.

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

Note: $\sum_{i=1}^{n} (X_i - \bar{X}) = 0 \Rightarrow$ only have $(n-1)$ indep. terms, so don't have all indep. terms. Hence why we haven't talked about S^2 at all yet.

• we get a nice result about the significance of the n-1 also representing deg. of freedom in Normal case.

[see Handout #3]

§8.3 Handout #3

STAT 480B

Sample Mean and Sample Variance from Normal Distribution

Theorem: If $X_1, X_2, \ldots, X_n$ form a random sample from $N(\mu, \sigma^2)$, then

let's us talk about dists w/ S^2 in them.

1. $\overline{X}$ and the terms $X_i - \overline{X}, i = 1, 2, \ldots, n$, are independent.
2. $\overline{X}$ and S^2 are independent. the sample mean & deviations from the mean are indep.
3. $\frac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1)$. S^2 only dep. on the deviations. σ^2

Pf: Practice; most of proof in book on Pg 272, 273. Makes use of transformation technique.

Example: Consider a random sample of size 10 from a normal distribution with mean 6 and variance 25. Determine the probability that the sample variance exceeds 18.3.

S^2

Given that $X_i \overset{iid}{\sim} N(6, 25)$, $i = 1, 2, \ldots, 10$.

Want $P(S^2 > 18.3)$.

$$P(S^2 > 18.3) = P\Big(\underbrace{\frac{(n-1)S^2}{\sigma^2}}_{Y \sim \chi^2(9)} > \frac{\overset{n-1}{(9)}\overset{S^2>18.3}{(18.3)}}{\underset{\sigma^2}{(25)}}\Big)$$

don't know dist of S^2 itself, but can use the Theorem above; we DO know how to find the dist. of $\frac{(n-1)S^2}{\sigma^2}$, so we transform S^2 into $\frac{(n-1)S^2}{\sigma^2}$.

$= P(Y > 6.588)$

$= 1 - P(Y < 6.588)$ use χ^2 table, $v = 9$

$= 1 - (0.321) = \boxed{0.679.}$

S^2 is unbiased for σ^2, meaning $E(S^2) = \sigma^2$.

statistic parameter

S^2's dist. dep. only on σ^2

$\bar{X}$'s dist dep. on μ and σ^2; if one is unknown, then cannot parameterize. Next best thing to σ^2 is S^2.

§ 8.4 – The t, F, and Beta Distributions

HW due next Mon., feedback by Tues. Exam next Thurs.

- Thm; If $\overbrace{Z \sim N(0,1)}^{\text{starting in Normal Land}}$ and $U \sim \chi^2(v)$ and Z & U are independent,

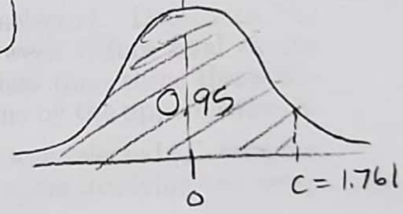
then $$T = \frac{Z}{\sqrt{U/v}} \sim t(v),$$

where $t(v)$ is the Student's t-dist, with v degrees of freedom.

- looks like Normal, is centered @ 0, but is flatter...
- as $v \to \infty$, $t(v) \to$ normal dist.
- have t-tables! Pg 608.

Ex) Sps $T \sim t(14)$.

- Find c such that $P(T < c) = 0.95$ (i.e., find the 95th percentile).
  - Using Pg 608 table, $\boxed{c = 1.761 = t_{0.95}(14)}$.



- Find $P(T < 2.5)$.
  - can't get an exact value cuz table is only in terms of percentiles, but can give bounds.
  - $\boxed{P(T < 2.5) \text{ is between } 0.975 \text{ \& } 0.99.}$

- Motivation to use T: replace σ^2 with s^2.
- Thm: Let $X_1, X_2, \ldots, X_n$ be a r.s. from $N(\mu, \sigma^2)$. Then

$$\frac{\bar{X} - \mu}{s/\sqrt{n}} \sim t(n-1).$$

$n-1$ deg. of freedom.

specific case of $t(v)$; helpful if recognizable

s instead of σ

Pf: Let $Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}}$, $U = \frac{(n-1)s^2}{\sigma^2} \sim \chi^2(n-1)$. Here, Z and U are independent since $\bar{X}$ and s^2 are independent. Now $T = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \Big/ \sqrt{\frac{(n-1)s^2}{\sigma^2} \cdot \left(\frac{1}{n-1}\right)}$

fcn of $\bar{X}$; fcn of s^2

$\Rightarrow T = \frac{\bar{X} - \mu}{s/\sqrt{n}} \sim t(n-1)$. (s can be calculated if given the individual values from each X_i.) ∎

STAT 480B
Problem Set 3
Due Date: Monday, February 7, 2022

Name:

Solve the following problems completely and neatly. You are expected to work **independently** of each other. If the problem involves random variables, make sure that you define the random variables. If you are using approximations, justify why your approximation is valid. Use the appropriate notation and if applicable, encircle your final answer. **No solution, no credit.**

1. Studies have shown that the amount of time, in hours, undergraduate students spend studying the week before final exams can be modeled by a normal distribution with the mean and variance varying depending on the class standing (freshman, sophomore, junior, senior). The means and variances are as follows:

	Class Standing	Mean	Variance
X_1	Freshman	μ_1 25	σ_1^2 49
X_2	Sophomore	μ_2 28	σ_2^2 53
X_3	Junior	μ_3 32	σ_3^2 37
X_4	Senior	μ_4 27	σ_4^2 40

   (a) A student from each class standing was randomly selected. Determine the probability that the total time spent studying the week before final exams by the underclassmen (freshman and sophomore) is less than three times the average time spent studying the week before final exams by the upperclassmen.

   (b) Suppose a random sample of ten sophomore students was selected. Determine the probability that the sample variance of the time spent studying the week before final exams is at most 48 hours^2.

2. You are given that $U \sim \chi^2(n_1)$, $W = U + V \sim \chi^2(n_1 + n_2)$, and U and V are independent. Prove that $W - U \sim \chi^2(n_2)$.

3. The amount of time, in days, that a surveillance camera will run without having to be reset can be modeled by an exponential distribution with mean 50 days. Suppose a random sample of seven surveillance cameras was obtained. Determine the probability that the mean of the sampled cameras will be between 50 and 100 days.

4. Suppose $X_i, i = 1, 2, \ldots, m$, form a random sample from $N(\mu_1, \sigma_1^2)$ and $Y_i, i = 1, 2, \ldots, m$, form a random sample from $N(\mu_2, \sigma_2^2)$. In addition, assume that X_is and Y_is are independent. Determine the distribution of $\dfrac{\sigma_1 \sum_{i=1}^{m} (Y_i - \mu_2)}{\sigma_2 \sqrt{\sum_{i=1}^{m} (X_i - \mu_1)^2}}$. If the resulting distribution is one of the special distributions, identify the distribution and all its parameters; otherwise, state "unknown".

Matthew Miller Problem Set #3, due Mon. Feb 7th @ 6:00 pm $\frac{41}{45}$ very good! (11)

Define your r.v.s properly!

1) · Given that $X_1 \sim N(25, 49)$ [Freshmen], $X_2 \sim N(28, 53)$ [Sophomores], $X_3 \sim N(32, 37)$ [Juniors], $X_4 \sim N(27, 40)$ [Seniors].

X_1 = amt of time, in hrs, freshmen spend studying... & so on...

(a) Find $P\left(\underbrace{X_1 + X_2}_{\text{total time spent studying by the underclassmen}} < 3\underbrace{\left(\frac{X_3 + X_4}{2}\right)}_{\text{average time spent studying by the upperclassmen}}\right)$.

· Rearranging, $P\left(X_1 + X_2 < 3\left(\frac{X_3 + X_4}{2}\right)\right) = P(2X_1 + 2X_2 - 3X_3 - 3X_4 < 0)$.

9/10

· Letting $Y = \overset{a_1}{2}X_1 + \overset{a_2}{2}X_2 \overset{a_3}{-3}X_3 \overset{a_4}{-3}X_4$, Y is a linear combination of independent normal R.V.s, and we want to find $P(Y < 0)$.

· By a theorem, it follows that $Y \sim N\left(\sum_{i=1}^{4} a_i \mu_i, \sum_{i=1}^{4} a_i^2 \sigma_i^2\right)$.

That is, Y is Normally distributed with

mean $= 2(\overset{\mu_1}{25}) + 2(\overset{\mu_2}{28}) - 3(\overset{\mu_3}{32}) - 3(\overset{\mu_4}{27}) = -71$ hours, and

variance $= (2)^2(\overset{\sigma_1^2}{49}) + (2)^2(\overset{\sigma_2^2}{53}) + (-3)^2(\overset{\sigma_3^2}{37}) + (-3)^2(\overset{\sigma_4^2}{40}) = 1101$ hours^2,

so $Y \sim N(\mu = -71, \sigma^2 = 1101)$.

· Thus, $P(Y < 0) = P\left(Z < \frac{0 - (-71)}{\sqrt{(1101)}}\right) = P\left(Z < \frac{71}{33.18...}\right)$

$= P(Z < 2.13975...) = P(Z < 2.14) = \boxed{0.9838.}$

(b) Sps $X_{2,1}, X_{2,2}, \dots, X_{2,10}$ is a random sample of 10 sophomore students such that each $X_{2,i} \sim N(\overset{\mu_2}{28}, \overset{\sigma_2^2}{53})$, $i = 1, 2, \dots, 10$.

· Want to find $P(S^2 \leq 48 \text{ hrs}^2)$; to do so, use theorem:

$\frac{(n-1)S^2}{\sigma^2} \sim \chi^2(v = n-1)$. I.e., $n-1 = 10-1 = 9$

5/5

$P(S^2 \leq 48) = P\left(\underbrace{\frac{(n-1)S^2}{\sigma_2^2}}_{\text{call this } Y} \leq \frac{(9)(48)}{(53)}\right)$

call this Y; then $Y \sim \chi^2(v = (10) - 1 = 9)$.

$= P(Y \leq 8.15)$ [use χ^2-dist. table, $v = 9$] (in table, closest thing to 8.15 is 8.2; use that)

$= \boxed{0.486.}$

2) Pf: Sps U, V, W are R.V.s where $U \sim \chi^2(n_1)$, $W = U+V \sim \chi^2(n_1+n_2)$ and U and V are independent.

First, since $W = U+V$ and U, V indep., it follows that

$$M_W(t) = M_{U+V}(t) = M_U(t) \cdot M_V(t).$$

Thus, $\boxed{M_V(t) = \frac{M_W(t)}{M_U(t)}}$, where $M_W(t) = (1-2t)^{\frac{-(n_1+n_2)}{2}}$ and $M_U(t) = (1-2t)^{-n_1/2}$.

Therefore,

$$M_{\underbrace{W-U}_{V}}(t) = M_{(\cancel{U}+V)-\cancel{U}}(t)$$

$$= M_V(t)$$

$$= \frac{M_W(t)}{M_U(t)}$$

not needed

$$= \frac{(1-2t)^{-(n_1+n_2)/2}}{(1-2t)^{-n_1/2}}$$

$$= (1-2t)^{\cancel{\frac{-n_1}{2}} - \frac{n_2}{2} - \left(\cancel{\frac{-n_1}{2}}\right)}$$

$$= (1-2t)^{-n_2/2},$$

which implies that $\underbrace{W-U}_{=V} \sim \chi^2(n_2)$. [And also that $V \sim \chi^2(n_2)$.] ∎

10/10

3) Let $X_1, \dots, X_7 \overset{iid}{\sim} EXP(50 \text{ days})$. ← θ, also μ

X_i = amt of time, in days, that camera i will run w/out having to be reset

· Let $\bar{X} = \dfrac{\sum_{i=1}^{7} X_i}{7}$ denote the sample mean; need to find $P(50 \le \bar{X} \le 100)$.

– Can't apply CLT to do this because each X_i being exponentially distributed means we have skewed distributions, so we would not have a sufficiently large sample size (need $n \geq 30$, roughly, when don't have symmetric distributions in our r.s.).

– So, try MGF technique:

not needed since we have proven this in class

$$M_{\bar{X}}(t) = M_{\sum_{i=1}^{7} X_i}(t) = \left[M_{X_i}(t)\right]^7 \text{ since each } X_i \text{ is independent \& identical}$$

$$= \left[\frac{1}{1-(50)t}\right]^7 = (1-50t)^{-7}$$

$$\Rightarrow \sum_{i=1}^{7} X_i \; \cancel{\bar{X}} \sim GAM(\theta = 50, \kappa = 7).$$

· To utilize this fact to determine the value of $P(50 \le \bar{X} \le 100)$, we convert $\cancel{\bar{X}}$ from a gamma-dist. to a χ^2-dist:

If $X \sim GAM(\theta, \kappa)$, then $Y = \frac{2X}{\theta} \sim \chi^2(2\kappa)$.

8/10

$$P(50 \le \bar{X} \le 100) = \cancel{P\left(\frac{2(50)}{50} \le \frac{2\bar{X}}{50} \le \frac{2(100)}{50}\right)}$$

$$= \cancel{P(2 \le Y \le 4)} \text{ where } Y \sim \chi^2(v = 2\kappa = 2(7) = 14).$$

$$= P(Y \le 4) - P(Y \le 2) \quad [\text{use } \chi^2\text{-dist table for } v = 14]$$

$$= 0.0045 - 0$$

$$= \boxed{\cancel{0.0045}}$$

for $c = 3.8$, $v = 14$, get 0.003
for $c = 4.2$, $v = 14$, get 0.006.
Taking their average, we get 0.0045

$$= P\left(50 \le \frac{\sum X_i}{7} \le 100\right)$$

$$= P(350 \le \textstyle\sum X_i \le 700)$$

$$= P\left[\frac{2(350)}{50} \le Y \le \frac{2(700)}{50}\right] \text{ where } Y \sim \chi^2(14)$$

$$= P(Y < 28) - P(Y < 14)$$

$$= .986 - .550 = .436$$

10/10

4) Let $X_1, X_2, \dots, X_m \overset{iid}{\sim} N(\mu_1, \sigma_1^2)$, let $Y_1, Y_2, \dots, Y_m \overset{iid}{\sim} (\mu_2, \sigma_2^2)$, and assume the X_i's and Y_i's are independent.

For convenience, let

$$W = \frac{\sigma_1 \sum_{i=1}^{m} (Y_i - \mu_2)}{\sigma_2 \sqrt{\sum_{i=1}^{m} (X_i - \mu_1)^2}}.$$

- First, we rewrite W as

$$W = \frac{\sigma_1 \sum_{i=1}^{m} (Y_i - \mu_2)}{\sigma_2 \sqrt{\sum_{i=1}^{m} (X_i - \mu_1)^2}} = \frac{\frac{1}{\sigma_2} \sum_{i=1}^{m} (Y_i - \mu_2)}{\frac{1}{\sigma_1} \sqrt{\sum_{i=1}^{m} (X_i - \mu_1)^2}} = \frac{\sum_{i=1}^{m} \frac{(Y_i - \mu_2)}{\sigma_2}}{\sqrt{\sum_{i=1}^{m} \frac{(X_i - \mu_1)^2}{\sigma_1^2}}}$$

In doing so, we see that the numerator is a sum of Z_i's, where each $Z_i \sim N(0,1)$ and each Z_i is independent. Meanwhile, we see the denominator involves a sum of Z_i^2's, which is therefore describable as a χ^2-distribution with m degrees of freedom. In other words, let $U = \sum_{i=1}^{m} \frac{(X_i - \mu_1)^2}{\sigma_1^2}$, then $U \sim \chi^2(m)$.

- Now, multiplying the top and bottom by $\frac{1}{\sqrt{m}}$, we have

$$W = \frac{\frac{1}{\sqrt{m}} \sum_{i=1}^{m} Z_i}{\frac{1}{\sqrt{m}} \cdot \sqrt{U}};$$ from exercise 3 of Problem Set 1, we know that

the numerator $\frac{1}{\sqrt{m}} \sum_{i=1}^{m} Z_i$ adheres to a standard normal distribution. So, let $Z = \frac{1}{\sqrt{m}} \sum_{i=1}^{m} Z_i$; then we have, in total, that

$$W = \frac{Z}{\sqrt{U/m}},$$ where $Z \sim N(0,1)$ and $U \sim \chi^2(m)$, Z and U independent.

- Therefore, $$\boxed{W = \frac{\sigma_1 \sum_{i=1}^{m} (Y_i - \mu_2)}{\sigma_2 \sqrt{\sum_{i=1}^{m} (X_i - \mu_1)^2}} \sim t(m).}$$

I.e., W is a Student's t-distribution with m degrees of freedom.

Start of lecture 2/7/22

Review) · Have $Z \sim N(0,1)$, $-\infty < z < \infty$.

· Then $Z^2 \sim \chi^2(1)$, $0 < z^2 < \infty$ ($z^2 > 0$, basically)

· Have $X \sim GAM(\theta, \kappa)$.

· Then $\frac{2X}{\theta} \sim \chi^2(2\kappa)$

· If have $Y_i \sim \chi^2(v_i)$, $i = 1, 2, \ldots, n$, Y_i's indep.,

then $\Sigma Y_i \sim \chi^2(\sum_{i=1}^{n} v_i)$ ← or nv_i if all the v_i's are identical.

· $X_1, X_2, \ldots, X_n \overset{iid}{\sim} N(\mu, \sigma^2) \Rightarrow \sum_{i=1}^{n} X_i \sim N(n\mu, n\sigma^2)$,

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

Ask Example problem where need to do this?

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

· $Z \sim N(0,1)$ and $U \sim \chi^2(v)$, with Z & U indep.,

$\Rightarrow \frac{Z}{\sqrt{U/v}} \sim t(v)$.

· $X_1, X_2, \ldots, X_n \overset{iid}{\sim} N(\mu, \sigma^2) \Rightarrow \frac{\bar{X} - \mu}{s/\sqrt{n}} \sim t(n-1)$.

all the sampling dists (dist. of a statistic — a value derived from a sample.) we know up to this point

— excludes Ch. 7 stuff, which are not sampling dists.

Ex #16, P. 285) Given $X_i \sim N(6, 25)$, $i = 1, 2, \ldots, 9$.

(a) Find $P(3 < \bar{X} < 7)$. ar RV... do we know its dist?

$\cdot\ \bar{X} \sim N(6, \frac{25}{9})$, so standardize:

$$P(3 < \bar{X} < 7) = P\left(\frac{3-6}{\sqrt{25/9}} < Z < \frac{7-6}{\sqrt{25/9}}\right) = P(-\tfrac{9}{5} < Z < \tfrac{3}{5})$$

$$= P(-1.8 < Z < 0.6) = P(Z < 0.6) - P(Z < -1.8)$$

$$= 0.7257 - (1 - 0.9641)$$

$$= \boxed{0.6898}$$

(b) Find $P\left(1.860 < \dfrac{\overset{=\sqrt{9}}{3}(\overset{\text{R.V.}}{\bar{X}} - \overset{\mu}{6})}{\underset{\text{R.V.}}{S}}\right)$; can make look like $\dfrac{\bar{X}-\mu}{S/\sqrt{n}} \sim t(n-1)$.

call this W. Then $W \sim t(8)$.

$n-1 = 9-1 = 8$

$= P(W > 1.860)$

$= 1 - P(W < 1.860)$ — uses percentiles, Pg 608 Table 6

$= 1 - 0.95$

$= \boxed{0.05}$

(c) Find $P(S^2 \leq 31.9375)$.

just have only an S^2, so try make look like $\dfrac{S^2(n-1)}{\sigma^2} \sim \chi^2(n-1)$

$\dfrac{(n-1)S^2}{\sigma^2} = \dfrac{(8)S^2}{(25)} \sim \chi^2(8)$, so

$= P\left(\underbrace{\dfrac{8S^2}{25}}_{Y,\ \sim \chi^2(8)} \leq \dfrac{8(31.9375)}{25}\right) = P(Y \leq 10.22) = \boxed{0.75}$

should use Table 4 instead of Table 5 cuz can get 10.22 exactly using Table 4 in body.

American

Snedecor's F-distribution

named F after Ronald Fisher

- derived from Normal dist.

• If $U_1 \sim \chi^2(v_1)$, $U_2 \sim \chi^2(v_2)$, and U_1 & U_2 independent,

then

$$F = \frac{U_1/v_1}{U_2/v_2} \sim F(v_1, v_2).$$

two params, each is a deg. of freedom, but order matters

v_1: deg. of freedom of numerator

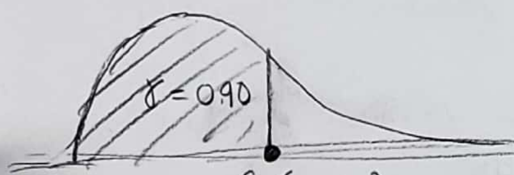
v_2: deg. of freedom of denominator

• Table 7, P. 609-611

- percentiles

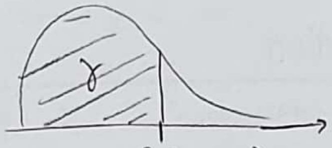
Note, $0 < f < \infty$ (cuz like the χ^2 dists, can only involve positive values)

• F is a skewed dist.



$f_\gamma(v_1, v_2)$

e.g., $f_{0.90}(v_1, v_2)$



$Y \sim F(v_2, v_1)$

$f_\gamma(v_2, v_1)$

so, $\frac{1}{f_{1-\gamma}(v_1,v_2)} = f_\gamma(v_2, v_1)$.

Example) $X \sim F(\overset{v_1}{2}, \overset{v_2}{8})$.

• Want 95th percentile of X:

$f_{.95}(2, 8) = 4.46$.

• What if want $f_{.10}(2, 8)$?

$f_{1-\gamma}(v_1, v_2)$

$$\boxed{f_{1-\gamma}(v_1, v_2) = \frac{1}{f_\gamma(v_2, v_1)}.} \quad \text{Why?}$$

• Let $X \sim F(v_1, v_2)$. Then $Y = \frac{1}{X} = \frac{\frac{U_1/v_1}{U_2/v_2}}{} = \frac{U_2/v_2}{U_1/v_1} \sim F(v_2, v_1)$.

$X = \frac{U_1/v_1}{U_2/v_2}$

So $P(X < f_{1-\gamma}(v_1, v_2)) = 1 - \gamma$,

but $X = \frac{1}{Y}$, so $P(\frac{1}{Y} < f_{1-\gamma}(v_1, v_2)) = 1 - \gamma$

$\Rightarrow P(\frac{1}{f_{1-\gamma}(v_1,v_2)} < Y) = 1 - \gamma$, ie, $P(Y > \frac{1}{f_{1-\gamma}(v_1,v_2)}) = 1 - \gamma$

$= 1 - P(Y < \frac{1}{f_{1-\gamma}(v_1,v_2)}) = 1 - \gamma \Rightarrow P(Y < \underbrace{\frac{1}{f_{1-\gamma}(v_1,v_2)}}_{f_\gamma(v_2,v_1)}) = \gamma$

$\overset{\nu_1\ \nu_2}{\sim F(2,8)}$

- Want $P(X < c) = 0.10$. find c.

$$c = f_{.10}(\overset{\nu_1}{2}, \overset{\nu_2}{8})$$

$$= \frac{1}{f_{.90}(\underset{\nu_2}{8}, \underset{\nu_1}{2})}$$

$$= \frac{1}{9.37}$$

F- dist used when making inferences about sample variance s^2.

- Think of a χ^2 dist when see s^2. [See Green Handout]

Beta Distribution

☆ $Y \sim BETA(a, b)$ if its pdf can be written as

$$f(y; a, b) = \frac{\Gamma(a+b)}{\Gamma(a)\cdot\Gamma(b)} \cdot y^{a-1} \cdot (1-y)^{b-1}, \quad 0 < y < 1, \quad a, b > 0$$

- Special case: $a = 1$, $b = 1$

$$f(y; 1, 1) = \frac{\Gamma(2)}{\Gamma(1)\,\Gamma(1)} y^0 \cdot (1-y)^0 = \frac{(2-1)!}{0! \cdot 0!} = 1, \quad 0 < y < 1$$

So $Y \sim Unif(0,1)$.

- Beta related to F-dist.

→

- If $X \sim F(v_1, v_2)$, then

don't use this often

$$Y = \frac{v_1/v_2 \cdot X}{1 + \frac{v_1}{v_2} X} \sim BETA\left(\frac{v_1}{2}, \frac{v_2}{2}\right).$$

Basically, the Beta dist is a generalization of the Uniform dist. It is continuous.

Cauchy Distribution

- If Z_1 & Z_2 are indep. $N(0,1)$, then

$$Y = \frac{Z_1}{Z_2} \sim \text{Cauchy}, \quad \text{where } -\infty < y < \infty.$$

(the $P(Z_2 = 0) = 0$ cuz Z_2 a continuous R.V.)

Exam 1

[end of Chapter 8]

§7.1 - 7.5, §8.1 - 8.4

- given color-coded tables, also table in cover of book with MGFs, pdfs, etc.

(Relevant to Exam 1)

STAT 480B

Example for the F-Distribution

Let X and Y denote the heights, in centimeters, of men and women, respectively. Suppose $X \sim N(173, 40)$ and $Y \sim N(160, 20)$. Random samples of ten men and seven women were taken. Determine the probability that the sample variance of men's heights is at least six times as much as the sample variance of women's heights.

what's given

- X = ht. in cm. of men $\sim N(\overset{\mu_X}{173}, \overset{\sigma_X^2}{40})$, $n_1 = 10$
- Y = ht. in cm. of women $\sim N(\overset{\mu_Y}{160}, \overset{\sigma_Y^2}{20})$, $n_2 = 7$

so X's & Y's indep., and each is indep. amongst themselves

- Find $P(S_X^2 \geq 6 S_Y^2)$.
  - Start by putting all R.V.s on one side like usual;

$= P\left(\frac{S_X^2}{S_Y^2} \geq 6\right)$; we know $\frac{(n_1-1)S_X^2}{\sigma_X^2} = \underbrace{\frac{(9)S_X^2}{(40)}}_{U_1} \sim \chi^2(9)$

have to form actual χ^2 dists first.

and we know $\frac{(n_2-1)S_Y^2}{\sigma_Y^2} = \underbrace{\frac{(6)S_Y^2}{(20)}}_{U_2} \sim \chi^2(6)$

χ^2/χ^2 => use F-distribution, but first need in proper form.

$$\frac{U_1/v_1}{U_2/v_2} = \frac{\frac{\cancel{9}}{40} S_X^2 \cdot \frac{1}{\cancel{9}}}{\frac{\cancel{6}}{20} S_Y^2 \cdot \frac{1}{\cancel{6}}} = \frac{1}{2}\frac{S_X^2}{S_Y^2} \sim F(9,6).$$

Now,

$$P\left(\frac{S_X^2}{S_Y^2} \geq 6\right) = P\left(\underbrace{\frac{1}{2}\frac{S_X^2}{S_Y^2}}_{\sim F(9,6)} \geq \frac{1}{2}\cdot 6\right)$$

(or whatever you call the R.V.)

$$= 1 - P(F(9,6) < 3)$$

$$\approx 1 - 0.90$$

$$= \boxed{0.10.}$$

Sugg. Exercises...

Chapter 7: 1c, 2, 3a, 5, 7a, 8, 9, 13

✓ 1) given random sample of size n from a dist w/ CDF $F(x) = 1 - \frac{1}{x}$ if $1 \le x < \infty$, and 0 o.w.

$X_1, X_2, \ldots, X_n \overset{iid}{\sim}$ something w/ $F_{X_i}(x) = \begin{cases} 1 - \frac{1}{x}, & 1 \le x < \infty \\ 0, & x < 1 \\ & (\text{o.w.}) \end{cases}$

(c) Find limiting dist. of $X_{1:n}^n$.

· Letting $Y_n = X_{1:n}^n$, we need to first figure out the CDF $G_n(y)$ of Y_n.

$$G_n(y) = P(Y_n \le y) = P(X_{1:n}^n \le y) = P(X_{1:n} \le \sqrt[n]{y})$$

$$= 1 - P(X_{1:n} > \sqrt[n]{y}) = 1 - P(\text{all } X_i\text{'s} > \sqrt[n]{y})$$

$$= 1 - P(X_1 > \sqrt[n]{y}, X_2 > \sqrt[n]{y}, \ldots, X_n > \sqrt[n]{y})$$

$$= 1 - [P(X_i > \sqrt[n]{y})]^n = 1 - [1 - P(X_i \le \sqrt[n]{y})]^n$$

$= 1 - [1 - F_{X_i}(\sqrt[n]{y})]^n$; if $\sqrt[n]{y} < 1$, then have $1 - [1-0]^n = 0$

if $1 \le \sqrt[n]{y} < \infty$, then have $1 - [\cancel{1} - (\cancel{1} - \frac{1}{\sqrt[n]{y}})]^n = 1 - [\frac{1}{\sqrt[n]{y}}]^n$

$= 1 - \frac{1^n}{(y^{1/n})^n} = 1 - \frac{1}{y}$. ~ if each X_i's are $1 \le X_i$, so is $X_{1:n}$, the smallest X_i, so $1 \le X_{1:n} < \infty$.

$\Rightarrow 1^n \le X_{1:n}^n < \infty \Rightarrow 1 \le Y_n < \infty$

$\Rightarrow 1 \le y < \infty$

So $G_n(y) = 1 - \frac{1}{y}$ if $1 \le y < \infty$, 0 o.w.

· Now take $\lim_{n\to\infty} G_n(y)$ to get $G(y)$ (potentially).

- If $y = 1$, $G_n(y) = 1 - 1 = 0$ so $\lim_{n\to\infty} G_n(y) = G(y) = 0$ for $y = 1$

- O.w., $\lim_{n\to\infty} G_n(y) = \lim_{n\to\infty} (1 - \frac{1}{y}) = 1 - \frac{1}{y} = G(y)$ $\forall y > 1$.

· Valid?

$\lim_{y\to\infty} G(y) = \lim_{y\to\infty} (1 - \frac{1}{y}) = 1 - 0 = 1$ ✓

$\lim_{y\to-\infty} G(y) = \lim_{y\to 1^+} (0) = 0$ ✓

So $Y_n = X_{1:n}^n \xrightarrow{d} Y \sim G(y) = \begin{cases} 1 - \frac{1}{y}, & y > 1 \\ 0, & y \le 1 \end{cases}$

2) Consider a r. sample of size n from a dist. with CDF $F(x) = (1+e^{-x})^{-1} \ \forall x \in \mathbb{R}$

✓(a) Does the largest order statistic, $X_{n:n}$, have a limiting distribution?

$G_n(y) = P(Y_n \le y) = P(X_{n:n} \le y) = P(\text{all } X_i\text{'s} \le y) = P(X_1 \le y, X_2 \le y, \ldots, X_n \le y)$

$= [F_{X_i}(y)]^n = [(1+e^{-y})^{-1}]^n = \frac{1}{(1+e^{-y})^n}, \quad y \in \mathbb{R}$

$\forall y \in \mathbb{R}, \ e^{-y} > 0 \Rightarrow 1+e^{-y} > 1$

$\Rightarrow (1+e^{-y})^n > (1)^n$

$\Rightarrow (1+e^{-y})^n > 1$

$\Rightarrow \underbrace{\frac{1}{(1+e^{-y})^n}}_{>0, \text{ also}} < 1$

So $0 < \frac{1}{(1+e^{-y})^n} < 1$.

just needs to be $<$ or ≤ 0, but also needs < 1 specifically, to get desired limit.

Thus $\lim_{n\to\infty} G_n(y) = \lim_{n\to\infty} \frac{1}{(1+e^{-y})^n}$

$= 0 = G(y) \ \forall y \in \mathbb{R}$.

exclude ∞'s when testing

$\boxed{\text{This is an invalid CDF b/c } \lim_{y\to\infty} G(y) \ne 1, \text{ so there is no limiting distribution for } X_{n:n}.}$

correct! No ✓

$e^{-e^{-y}} = e^{-\left(\frac{1}{e^{y}}\right)} = \frac{1}{e^{\frac{1}{e^y}}}$

✓(b) Does $X_{n:n} - \ln(n)$ have a limiting dist.? If so, what is it?

$G_n(y) = P(Y_n \le y) = P(X_{n:n} - \ln(n) \le y) = P(X_{n:n} \le y + \ln(n))$

$= P(\text{all } X_i\text{'s} \le y + \ln(n)) = [F_{X_i}(y+\ln(n))]^n$

$= [(1+e^{-(y+\ln(n))})^{-1}]^n = [(1+e^{-y} \cdot e^{-\ln(n)})^{-1}]^n = [(1+e^{-y} \cdot n^{-1})^{-1}]^n$

$= \left[\frac{1}{1+\frac{1}{ne^{y}}}\right]^n = \frac{1}{(1+\frac{1}{ne^{y}})^n}, \quad y \in \mathbb{R}$

$= \left(1 + \frac{(e^{-y})}{n}\right)^{n(-1)}$ $\qquad \lim_{n\to\infty}\left(1+\frac{c}{n}\right)^{nb} = e^{cb}$

$\Rightarrow G(y) = \lim_{n\to\infty} G_n(y) = \lim_{n\to\infty}\left(1+\frac{e^{-y}}{n}\right)^{n(-1)} = e^{(e^{-y})(-1)} = e^{-e^{-y}} = \frac{1}{e^{\frac{1}{e^y}}}$, $y \in \mathbb{R}$

Valid? $\lim_{y\to\infty} \frac{1}{e^{\frac{1}{e^y}}} = \frac{1}{e^0} = 1$ ✓, $\lim_{y\to-\infty} \frac{1}{e^{\frac{1}{e^y}}} = \frac{1}{e^{\infty}} = 0$ ✓

$\boxed{\text{So } Y_n = X_{n:n} - \ln(n) \xrightarrow{d} Y \sim G(y) = \frac{1}{e^{\frac{1}{e^y}}}, \ y \in \mathbb{R}.}$

3a) Consider r.sample $X_1, \dots, X_n \overset{iid}{\sim} F_{X_i}(x) = \begin{cases} 1 - \frac{1}{x^2}, & x > 1 \\ 0, & \text{o.w.} \end{cases}$ $\leftarrow 1 < X_{1:n} \Rightarrow 1 < y$

$(x \leq 1)$

Determine whether $X_{1:n}$ has limiting dist. ~~Z~~

$G_n(y) = P(Y_n \leq y) = P(X_{1:n} \leq y) = 1 - P(X_{1:n} > y) = 1 - [P(X_i > y)]^n$

$= 1 - [1 - P(X_i \leq y)]^n = 1 - [\cancel{1} - (\cancel{1} - \frac{1}{y^2})]^n = 1 - [\frac{1}{y^2}]^n, \ y > 1$

$G(y) = \lim_{n\to\infty} G_n(y) = \lim_{n\to\infty} (1 - (\frac{1}{y^2})^n) = 1 - \lim_{n\to\infty} (\frac{1}{y^2})^n$ $\Rightarrow y^2 > 1 \Rightarrow 0 < \frac{1}{y^2} < 1$

$= 1 - 0$

$= 1, \ \forall y > 1$

~~Invalid, $\neq 0$ ever, so no limiting dist.~~

and $= 0$, o.w

$\therefore G(y) = \begin{cases} 0, & y \leq 1 \\ 1, & y > 1 \end{cases}$

$\boxed{G(y) \text{ is degenerate at } y = 1, \text{ so can say } Y_n = X_{1:n} \xrightarrow{P} 1. \text{ (conv. stochastically)}}$

$\boxed{(\text{or, can just say } Y_n = X_{1:n} \xrightarrow{d} Y \sim G(y) = \begin{cases} 0, & y \leq 1 \\ 1, & y > 1 \end{cases})}$

I.e., in degen case, $Y_n \xrightarrow{P} 1$ (conv. in ~~prob.~~ (stochastically)) (is degen @ $y = 1$)

✓5) Sps that $Z_i \sim N(0,1)$ and that $Z_1, Z_2, \dots$ are independent. Use moment generating functions to find limiting dist of $\sum_{i=1}^{n} (Z_i + \frac{1}{n}) / \sqrt{n}$.

$M_{aX+b}(t) = e^{bt} M_X(at)$

$M_{\frac{\sum_{i=1}^n (Z_i + \frac{1}{n})}{\sqrt{n}}}(t) = M_{\frac{1}{\sqrt{n}} \sum_{i=1}^n Z_i + \frac{1}{\sqrt{n}}}(t)$

$M_{Z_i}(t) = e^{t^2/2} = M_Z(t)$

$= M_{Z + \underbrace{\frac{1}{n}}_{b}}(t)$

$\sum_{i=1}^n (Z_i + \underset{\text{constant}}{\frac{1}{n}}) = \sum_{i=1}^n Z_i + n(\frac{1}{n}) = (\sum_{i=1}^n Z_i) + 1$

$= e^{\frac{1}{n}t} M_Z(t) = e^{\frac{1}{n}t} \cdot e^{t^2/2}$

So $\lim_{n\to\infty} e^{\frac{1}{n}t} \cdot e^{t^2/2} = e^{0t} \cdot e^{t^2/2} = (1) e^{t^2/2}$

$= e^{t^2/2} \sim N(0,1)$.

so $\frac{\sum_{i=1}^n (Z_i + \frac{1}{n})}{\sqrt{n}} = \frac{\sum_{i=1}^n Z_i + 1}{\sqrt{n}}$

So $M_n(t) = M_{\frac{\sum_{i=1}^n (Z_i + \frac{1}{n})}{\sqrt{n}}}(t) \longrightarrow M(t) \sim N(0,1)$ as $n \to \infty$,

$= \underbrace{\frac{1}{\sqrt{n}} \sum_{i=1}^n Z_i}_{= Z \sim N(0,1)} + \frac{1}{\sqrt{n}}$

So the limiting dist of $\sum_{i=1}^n (Z_i + \frac{1}{n}) / \sqrt{n}$ is standard normal.

$\theta \quad \beta$

7a) Consider a random sample from a Weibull distribution, $X_i \sim WEI(1,2)$,
Find approximate values a and b such that for $n = 35$, $P(a < \bar{X} < b) \approx 0.95$.

- Let $X_1, X_2, \ldots, X_{35} \overset{iid}{\sim} WEI(1, 2)$ and let $\bar{X} = \frac{\sum_{i=1}^{35} X_i}{35}$ denote the sample mean of the random sample. ← MGF has no closed form, so neither will $\bar{X}$.
- Thus, need pdf of each X_i; integrate to get CDF of each X_i. Tho due to the nature of the problem, with a & b unknown, this won't help.
- Actually, $n = 35 > 30$, so can apply CLT. Just need μ_{X_i} and $\sigma^2_{X_i}$.

$$\mu_{X_i} = \theta\Gamma(1+\tfrac{1}{\beta}) = (1)\cdot\Gamma(1+\tfrac{1}{2}) = \Gamma(\tfrac{3}{2}) = (\tfrac{3}{2}-1)\Gamma(\tfrac{3}{2}-1)$$
$$= \tfrac{1}{2}\Gamma(\tfrac{1}{2}) = \tfrac{1}{2}\sqrt{\pi}$$

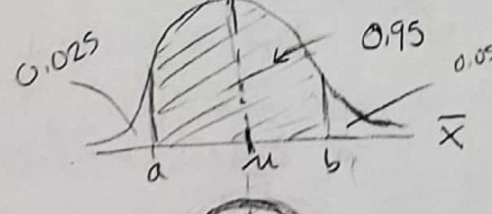
$$\sigma^2_{X_i} = \theta^2\left[\Gamma(1+\tfrac{2}{\beta}) - \Gamma^2(1+\tfrac{1}{\beta})\right] = (1)^2\left[\Gamma(1+\tfrac{2}{2}) - \Gamma^2(1+\tfrac{1}{2})\right]$$
$$= \left[\Gamma(2) - (\Gamma(\tfrac{3}{2}))^2\right] = (2-1)! - (\tfrac{1}{2}\sqrt{\pi})^2 = 1 - \tfrac{1}{4}\pi$$

By the CLT, $Y_n = \bar{X} = \frac{\sum_{i=1}^{35} X_i}{35} \xrightarrow{d} Y \sim N(\mu_{X_i}, \frac{\sigma^2_{X_i}}{n})$
$\hookrightarrow \sim N(\underbrace{\tfrac{1}{2}\sqrt{\pi}}_{\mu_Y}, \underbrace{\tfrac{(1-\frac{1}{4}\pi)}{35}}_{\sigma^2_Y})$.

So ~~$P(a < \bar{X} < b)$~~

$$0.95 = P(a < \bar{X} < b)$$
$$\approx P\left(\frac{a - \frac{1}{2}\sqrt{\pi}}{\sqrt{\frac{(1-\frac{\pi}{4})}{35}}} < Z < \frac{b - \frac{1}{2}\sqrt{\pi}}{\sqrt{\frac{(1-\pi/4)}{35}}}\right)$$
$$= P\left(\frac{a - 0.886}{0.078} < Z < \frac{b - 0.886}{0.078}\right)$$
$$= P\left(Z < \frac{b-0.886}{0.078}\right) - P\left(Z < \frac{a-0.886}{0.078}\right)$$

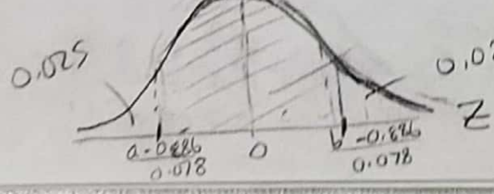
For 95th percentile, $z_\gamma = 1.645$, but how do we solve here since we have no way to just say $1.645 = \frac{b-0.886}{0.078}$, solve for b and likewise for a?



$0.05/2 = 0.025$

symmetric; ~~$a = -b$~~

not really...? No.



~~But P(Z~~

→

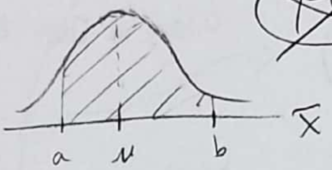
So $P\left(Z < \frac{b-0.886}{0.078}\right) = 1 - 0.025 = 0.975$,

$P\left(Z < \frac{a-0.886}{0.078}\right) = 0.025$ ← too small; but ~~[illegible]~~ we can get b.

· can we assume the precise locations of b and a...?

~~-why can't b be further to the right~~

- couldn't we have



, for instance? ~~Ask~~

No because then the area under the curve wouldn't be symmetric. Must have symmetry cuz of shape of Normal dist.

According to Table 3, the 97.5th percentile $z_{.975} = 1.960$.

So $\frac{b-0.886}{0.078} = 1.960 \Rightarrow b - 0.886 = 1.960(0.078...)$

$\Rightarrow b = 1.960(0.078...) + 0.886...$

$\Rightarrow b = 1.039702322$

$\Rightarrow \boxed{b \approx 1.04}$

~~- must a = -b?~~

So $0.95 = P\left(Z < \frac{b-0.886}{0.078}\right) - P\left(Z < \frac{a-0.886}{0.078}\right)$

$\approx P\left(Z < \frac{1.04-0.886}{0.078}\right) - P\left(Z < \frac{a-0.886}{0.078}\right)$

$= P(Z < 1.960) - P\left(Z < \frac{a-0.886}{0.078}\right)$

$= P(Z < 1.960) - P(Z < -1.960)$ by symmetry ~~(a = -b, I guess)~~

$\Rightarrow -1.960 = \frac{a-0.886}{0.078} \Rightarrow a - 0.886 = -1.960(0.078)$

$\Rightarrow a = -1.960(0.078..) + 0.886...$

$\Rightarrow a = -0.15347... + 0.886...$

$\Rightarrow a = 0.732751528...$

$\Rightarrow \boxed{a \approx 0.73}$

note: $\mu_{\bar{X}} = E(\bar{X})$

$= E\left(\frac{1}{35}\sum_{i=1}^{35} X_i\right)$

$= \frac{1}{35}\cdot\sum_{i=1}^{35} E(X_i)$

$= \frac{1}{35}\sum_{i=1}^{35}\left(\frac{1}{2}\sqrt{\pi}\right)$

$= \frac{1}{35}\left(35\cdot\frac{1}{2}\sqrt{\pi}\right)$

$= \frac{1}{2}\sqrt{\pi} \approx 0.886$,

$|0.886 - 0.73| = 0.153...$

$|0.886 - 1.04| = 0.153...$

So $P(a < \bar{X} < b) \approx P(Z < 1.960) - P\left(Z < \frac{0.73-0.886}{0.078}\right)$

$= P(Z < 1.960) - P(Z < -1.960)$

$= P(Z < 1.960) - (1 - P(Z < 1.960))$ by symm

$= 0.975 - (1 - 0.975)$

$= 0.975 - 0.025$

$= \boxed{0.95}$

~~Ask~~ how ~~compute~~

8, 9, 13 remaining from Ch. 7 sugg. exercises.

Ch. 8:

1) let X denote the weight in pounds of a bag of feed, where $X \sim N(101, 4)$. what is prob. that 20 bags will weigh at least a ton? (2000 lbs)

· Let $X_1, X_2, \ldots, X_{20}$ be a random sample of the weights of 20 of the above-mentioned bags of feed; then need find $P(\sum_{i=1}^{20} X_i \geq 2000 \text{ lbs})$.

$\frac{2000}{20} = 100$

ask — is it ok to treat this as just $20X$?

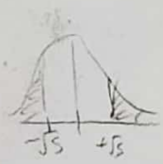
↳ Then have $P(20X \geq 2000)$
$= P(X \geq 100 \text{ lbs})$
$= 1 - P(X < 100)$
$= 1 - P(Z < \frac{100-101}{\sqrt{4}})$
$= 1 - P(Z < \frac{-1}{2})$
$= 1 - P(Z < -0.50)$
by symm $= 1 - P(Z > 0.50)$
$= \cancel{1} - (\cancel{1} - P(Z < 0.50))$
$= P(Z < 0.50)$
$\neq 0.6915$
= wrong.

Why is this approach flawed?

· First, $\sum_{i=1}^{20} X_i \sim N(\sum_{i=1}^{20} (1)\mu_i, \sum_{i=1}^{20} (1)^2 \sigma_i^2)$
$\sim N(20\mu, 20\sigma^2)$ ~~$\sim N(20(101)$~~
$\sim N(20(101), 20(4))$
$\sim N(2020, 80)$

so $P(\sum_{i=1}^{20} X_i \geq 2000) = 1 - P(\sum_{i=1}^{20} X_i < 2000)$
$= 1 - P(Z < \frac{2000-2020}{\sqrt{80}})$
$= 1 - P(Z < \frac{-20}{4\sqrt{5}})$
$= 1 - P(Z < \frac{-5}{\sqrt{5}})$
$= 1 - P(Z < -\sqrt{5}) = 1 - P(Z > \sqrt{5})$ by symm.
$= \cancel{1} - (\cancel{1} - P(Z < \sqrt{5})) = P(Z < \sqrt{5}) = P(Z < 2.24)$
$= \boxed{0.9875 = \text{right}}$

$\sqrt{80}$: $\sqrt{8}\ \sqrt{10}$; $\sqrt{2}\ \sqrt{4}\ \sqrt{2}\ \sqrt{5}$; $\sqrt{2}\ \sqrt{2}$



$\sqrt{80} = 2 \cdot 2 \cdot \sqrt{5}$
$= 4\sqrt{5}$

$\frac{5}{\sqrt{5}} \cdot \frac{\sqrt{5}}{\sqrt{5}} = \frac{5\sqrt{5}}{5} = \sqrt{5}$

ask

15c)

$$\frac{X_1 - X_2}{\sigma S_Z \sqrt{2}} \quad , \quad X_i \sim N(\mu, \sigma^2), \; i = 1, 2, \ldots, m \quad , \quad Z_i \sim N(0,1) \text{ with } i = 1, 2, \ldots, k$$

$X_1 - X_2 \sim N(0, 2\sigma^2)$ Lin combination

$$\frac{X_1 - X_2 - (\overset{\mu}{0})}{\sqrt{2}\,\sigma} \sim N(0,1), \quad \text{Recall: } \frac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1)$$

need

$$\frac{(k-1)S_Z^2}{\sigma_Z^2} \sim \chi^2(k-1)$$

$\sigma_Z^2 = 1$

$\Rightarrow (k-1)S_Z^2 \sim \chi^2(k-1)$

$\Rightarrow$ T-dist.

need a sqr root:

$$\frac{\frac{X_1 - X_2}{\sqrt{2}\,\sigma}}{\sqrt{\frac{\cancel{(k-1)}S_Z^2}{\cancel{(k-1)}}}} \sim t(k-1)$$

$$= \boxed{\frac{X_1 - X_2}{\sqrt{2}\,\sigma\, S_Z} \sim t(k-1)}$$

χ^2 dist

- get by squaring a Z dist.

T dist

- get by a $Z / \sqrt{\chi^2 / \nu}$

F dist

- get by χ^2 / χ^2

$$Z_i \sim \frac{Y_i - \mu_2}{\sigma_2} \sim N(0,1)$$

$$\sum_{i=1}^{m} Z_i \sim N(0, m)$$

$$Z = \frac{1}{\sqrt{m}} \sum_{i=1}^{m} Z_i \sim N(0,1)$$

$$Z_i' = \frac{X_i - \mu_1}{\sigma_1} \sim N(0,1)$$

$$(Z_i')^2 = \left(\frac{X_i - \mu_1}{\sigma_1}\right)^2 \sim \chi^2(1)$$

$$\sum_{i=1}^{m} (Z_i')^2 \sim \chi^2(m).$$

15) Sps that $X_i \sim N(\mu, \sigma^2)$, $i = 1, 2, \ldots, n$ and $Z_i \sim N(0,1)$, $i = 1, 2, \ldots, k$, and all variables are indep. State distribution of the following variables if possible.

a) $X_1 - X_2$.

- lin comb. of two normal r.v.s;

let $Y = X_1 - X_2$, then $Y \sim N((1)\mu + (-1)\mu,\ (1)^2\sigma^2 + (-1)^2\sigma^2)$

$\sim N(0, 2\sigma^2)$

b) —

c) $\dfrac{X_1 - X_2}{\sigma S_Z \sqrt{2}}$

- numerator $X_1 - X_2 \sim N(0, 2\sigma^2)$
- $S_Z = \sqrt{S_Z^2} = \sqrt{\frac{1}{k-1}\left(\sum_{i=1}^{k}(Z_i - \bar{Z}^2)\right)}$

- note that $X_1 - X_2 \sim N(0, 2\sigma^2)$, if converted to, say, Z ~~...~~ standard normal, then we have $Z = \dfrac{X_1 - X_2 - (0)}{\sqrt{2\sigma^2}} = \dfrac{X_1 - X_2}{\sigma\sqrt{2}} \sim N(0,1)$

- so have a ~~standard normal~~ ~~...~~ over a sample variance form (think χ^2),

$\Rightarrow$ get into the form $T = \dfrac{Z}{\sqrt{U/v}}$.

- In general, $\dfrac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1)$. Here, for S_Z^2,

we have $\dfrac{(k-1)S_Z^2}{\sigma_Z^2} \sim \chi^2(k-1)$, where $\dfrac{(k-1)S_Z^2}{\underset{=1}{\sigma_Z^2}} = (k-1)S_Z^2 \sim \chi^2(k-1)$

call this U.

So we have $\dfrac{(Z =)\ \frac{X_1 - X_2}{\sigma\sqrt{2}}}{S_Z}$; need sqrt on bottom...

need to make $\dfrac{1}{S_Z}$ look like $\dfrac{1}{\frac{(k-1)S_Z^2}{k-1}} = \dfrac{1}{U/_{k-1}}$

$\dfrac{1}{S_Z} = \dfrac{1}{\frac{(k-1)S_Z}{(k-1)}} = \dfrac{1}{\quad}$

$S_Z = \sqrt{S_Z^2} = \sqrt{\dfrac{(k-1)S_Z^2}{(k-1)}} = \sqrt{\dfrac{U}{k-1}}$, so

$$\frac{X_1 - X_2}{\sigma\sqrt{2}\, S_Z} = \frac{\frac{X_1 - X_2}{\sigma\sqrt{2}}}{\sqrt{\frac{(k-1)S_Z^2}{(k-1)}}} = \frac{Z}{\sqrt{\frac{U}{k-1}}} \sim t(v-1)$$

(d) Z_1^2

$Z_1^2 \sim \chi^2(1)$

(e) $\dfrac{\sqrt{n}(\bar{X} - \mu)}{\sigma S_z} \leftarrow = \dfrac{(\bar{X} - \mu)}{\sigma \cdot S_z/\sqrt{n}}$ ← actually maybe not quite right cuz S_z is for Z_i, not for the ~~X~~ X_i's.

S_z

18) $V_1 \sim \chi^2(5)$, $V_2 \sim \chi^2(9)$ > indep.

Find b s.t. $P\left(\frac{V_1}{V_1+V_2} < b\right) = 0.90$

↑ to say this is like an F-dist, first need that $V_1 \perp (V_1+V_2)$. Well they def. are dependent! So F out of question.

positive cuz χ^2

$$\frac{V_1}{V_1+V_2}$$

positive cuz χ^2 → can flip, allowing for simplification/progress.

$$P\left(\frac{V_1}{V_1+V_2} < b\right) = 0.90$$

$$\Rightarrow P\left(\frac{V_1+V_2}{V_1} > \frac{1}{b}\right) = 0.90$$

NOW can get a F-dist, cuz V_2 and V_1 indep.

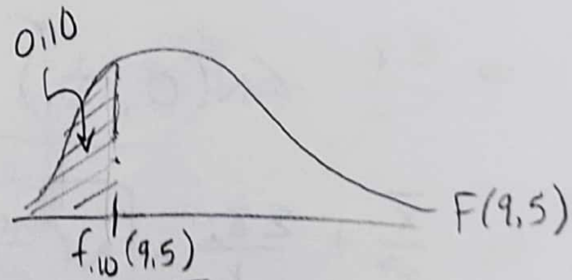
$$\Rightarrow P\left(1+\frac{V_2}{V_1} > \frac{1}{b}\right) = 0.90 \Rightarrow P\left(\frac{V_2}{V_1} > \frac{1}{b}-1\right) = 0.90.$$

Now
$$\frac{\frac{V_2}{9}}{\frac{V_1}{5}} \sim F(9,5)$$

$$= \frac{5}{9}\cdot\frac{V_2}{V_1}$$

So $P\left(\underbrace{\frac{5}{9}\cdot\frac{V_2}{V_1}}_{\sim F(9,5)} > \frac{5}{9}\left(\frac{1}{b}-1\right)\right) = 0.90$

$1-0.90 =$



$$\Rightarrow P\left(F(9,5) < \underbrace{\frac{5}{9}\left(\frac{1}{b}-1\right)}_{= f_{.10}(9,5) = \frac{1}{f_{.90}(5,9)}}\right) = 0.10$$

need flip

So $\frac{1}{f_{.90}(5,9)} = \frac{5}{9}\left(\frac{1}{b}-1\right) \Rightarrow \frac{1}{2.61} = \frac{5}{9}\left(\frac{1}{b}-1\right) \Rightarrow \ldots \Rightarrow \boxed{b = 0.5918.}$

3) $X_1, \dots, X_n \overset{iid}{\sim} N(\mu, \sigma^2)$

$U = \Sigma X_i$

$W = \Sigma X_i^2$

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

$\Rightarrow aX \sim N(a\mu, a^2\sigma^2)$

$a = \frac{1}{\sigma^2}$

(a) Find a statistic, a fcn of U and W, s.t. ~~is~~ unbiased w.r.t. the param. $\theta = 2\mu - 5\sigma^2$.

$\rightarrow$ I.e., want $E(\;) = \theta$

Recall: $E(\bar{X}) = \mu$ in the Normal case. Also,

$E(S^2) = \sigma^2$ " " " " .

$\rightarrow$ Find T such that $E(T) = \theta$.

- Let $\boxed{T = 2\bar{X} - 5S^2}$

$= 2\frac{U}{n} - 5\frac{(W - U^2/n)}{n-1}$

computational formula for S^2, put in terms of ~~Σ~~ W and U.

15n) $\bar{X} \sim N(\mu, \frac{\sigma^2}{n})$

$\Rightarrow \frac{\bar{X}}{\sigma^2} \sim N(\frac{\mu}{\sigma^2}, \frac{1}{n\sigma^2})$

$E(\frac{\bar{X}}{\sigma^2}) = \frac{1}{\sigma^2}E(\bar{X}) = \frac{\mu}{\sigma^2}$,

$Var(\frac{\bar{X}}{\sigma^2}) = (\frac{1}{\sigma^2})^2 Var(\bar{X})$

$= (\frac{1}{\sigma^4}) \cdot \frac{\sigma^2}{n} = \frac{1}{\sigma^2 n}$

$\sum_{i=1}^{k} Z_i \sim N(0, k)$

$Var(\frac{\Sigma Z_i}{k}) = \frac{1}{k^2} Var(\Sigma Z_i) = \frac{1}{k^2}(k) = \frac{1}{k}$

$\Rightarrow \frac{\Sigma Z_i}{k} \sim N(0, \frac{1}{k})$

So $\underbrace{\frac{\bar{X}}{\sigma^2} + \frac{\Sigma Z_i}{k}}_{\text{Lin comb. of Normal}} \sim \boxed{N(\frac{\mu}{\sigma^2}, \frac{1}{n\sigma^2} + \frac{1}{k})}$

Lin comb. of Normal =

this only works w/ Normal

χ^2 coefficients must be +1 only, no others allowed for a lin. combo. of them

#4) Want to show Y_1 & Y_2 indep.

To do so, try to show $M_{Y_1+Y_2}(t) = M_{Y_1}(t) \cdot M_{Y_2}(t)$.

$$M_{Y_1+Y_2}(t) = E(e^{t(Y_1+Y_2)})$$

$$= E(e^{t(X_1+\cancel{X_2}+X_1-\cancel{X_2})})$$

$= E(e^{t}$ $\quad$ X_i's go away. issue.

No do joint MGF. version.

Show $M_{Y_1 Y_2}(t_1, t_2) = M_{Y_1}(t_1) \cdot M_{Y_2}(t_2)$

$$= E(e^{t_1 Y_1 + t_2 Y_2}) = E(e^{t_1(X_1+X_2) + t_2(X_1-X_2)})$$

$$= \ldots$$

22) Find $E(X^n)$, given that $X \sim BETA(p, q)$.

$$E(X^n) = \int_0^1 x^n \cdot \frac{\Gamma(p+q)}{\Gamma(p)\cdot\Gamma(q)} \cdot x^{p-1} \cdot (1-x)^{q-1} \, dx$$

make look like pdf...

17c)

2a) S denotes the diameter of a shaft and B the diameter of a bearing, where S and B are indep. with $S \sim N(1, 0.0004)$ and $B \sim N(1.01, 0.0009)$. If a shaft & a bearing are selected @ random, what is the prob. that the shaft diameter will exceed the bearing diameter?

- Find $P(S > B)$.

$$P(S > B) = P(\underbrace{S - B}_{} > 0)$$

lin comb. of N R.V.s ; call this Y, then $Y \sim N(\mu_Y, \sigma_Y^2)$

where

$$\mu_Y = (1)\mu_S + (-1)\mu_B = 1 - 1.01 = -0.01 \text{, units}$$

$$\sigma_Y^2 = (1)^2\sigma_S^2 + (-1)^2\sigma_B^2 = 0.0004 + 0.0009 = 0.0013 \text{ units}^2$$

so $P(S > B) = P(S - B > 0)$

$$= P(Y > 0) = 1 - P(Y \le 0)$$

$$= 1 - P\left(Z \le \frac{0 - (-0.01)}{\sqrt{(0.0013)}}\right) = 1 - P\left(Z \le \frac{0.01}{0.036055\ldots}\right)$$

$$= 1 - P(Z \le 0.28) \text{ (rounded)}$$

$$= 1 - 0.6103 = \boxed{0.3897}$$ ✓

3ab)