especially an exponential(β).

You should recognize the $\exp(\beta)$ pdf $f_Y(y) = \frac{1}{\beta} e^{-y/\beta}$, $y \geq 0$

and DF $F_Y(y) = 1 - e^{-y/\beta}$, $y \geq 0$.

16] Often the sum of ind. RV's is a brand name RV. Use the back of the text to recognize mgf's.

ex] If $Y_1, \ldots, Y_n$ are ind gamma(α_i, $\underset{\text{same } \beta}{\beta}$), then $\sum_{i=1}^{n} Y_i$ is gamma $\left(\alpha = \sum_{i=1}^{n} \alpha_i, \beta\right)$.

ex] If $Y_1, \ldots, Y_n$ are ind normal μ_i, σ_i^2, then $\sum_{i=1}^{n} Y_i$ is normal $\mu = \sum_{i=1}^{n} \mu_i$, $\sigma^2 = \sum_{i=1}^{n} \sigma_i^2$.

ex] $Y_1, \ldots, Y_n$ ind binomial n_i, $\underset{\text{same } p}{p}$
$\sum_{i=1}^{n} Y_i$ is binomial $\sum_{i=1}^{n} n_i$, p

ex] $Y_1, \ldots, Y_n$ ind poisson λ_i, then $\sum_{i=1}^{n} Y_i$ is Poisson $\sum_{i=1}^{n} \lambda_i$.

ex] $Y_1, \ldots, Y_n$ ind $\exp(\lambda)$, then $\sum_{i=1}^{n} Y_i$ is gamma(n, λ).

ex] If $Y_1, \dots, Y_n$ are ind $\chi^2_{n_i}$,

then $\sum_{i=1}^{n} Y_i$ is $\chi^2_{\sum_{i=1}^{n} n_i}$

proof of $Y_1, \dots, Y_n$ iid $\exp(\lambda)$ example:

$m_{Y_i}(t) = (1-\lambda t)^{-1} = \frac{1}{(1-\lambda t)}$, $U = \sum_{i=1}^{n} Y_i$

$m_U(t) = \prod_{i=1}^{n} m_{Y_i}(t) = \left(\frac{1}{1-\lambda t}\right)^n = (1-\lambda t)^{-n}$

which is the mgf of a gamma$(\alpha=n, \beta=\lambda)$ RV.

so U is gamma$(\alpha=n, \beta=\lambda)$

ch7] D p 330 Let $Y_1, \dots, Y_n$ be iid RV's.

A random sample $y_1, \dots, y_n$ are the observed values of the RV's.

2] * p 347 A statistic is a function of the observed RVs in a sample and known constants.

ex] $\overline{Y} = \frac{1}{n}\sum_{i=1}^{n} Y_i$ is a RV,

the statistic $\overline{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$.

ex] $\overline{y} - \mu$ is not a statistic if μ is not known.

3] Let $Y_1, \ldots, Y_n$ be a random sample of size n from a normal dist with mean μ and variance σ^2.

Then $\overline{Y}$ is normal with mean $\mu_{\overline{Y}} = \mu$ and variance $\sigma^2_{\overline{Y}} = \frac{\sigma^2}{n}$.

variable	population	mean	variance
Y	pop of Y	μ	σ^2
$\overline{Y}$	sampling distribution of $\overline{Y}$	$\mu_{\overline{Y}}$	$\sigma^2_{\overline{Y}}$

4] The sampling distribution of a statistic is the distribution of all the values of the statistic in all possible samples of the same size n = population of the statistic.

ex] pop net worths of 3 people
Bill Gates 90 Billion MJ 0.8 B Y 0.0 B
Taylor Swift

Find the sampling distribution of $\overline{Y}$ for samples of size 2

		$\overline{Y}$	
BG	MJ	90.8/2	= 45.4
BG	Y	90/2	= 45
MJ	Y	0.8/2	= 0.4

$\overline{Y}$	45.4	45	0.4
$P(\overline{Y})$	1/3	$\frac{1}{3}$	$\frac{1}{3}$

	0.8	90	0
Y	MJ	BG	Y
$P(Y)$	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{3}$

5) Common problem] Forwards calculations with $\overline{Y}$: use $z = z(\overline{Y}) = \dfrac{\overline{Y}_{value} - \mu_{\overline{Y}}}{\sigma_{\overline{Y}}}$

and proceed as before.

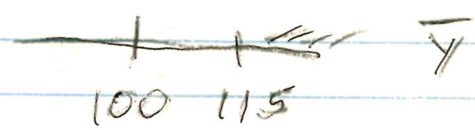
ex] $n = 100$ Y_i = ith IQ score

$\approx$ Normal with $\mu = 100$, $\sigma = 15$

a) Find $P(\overline{Y} > 115)$

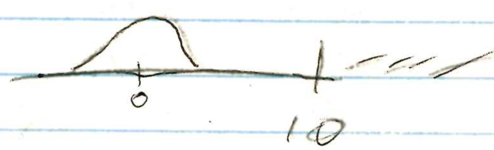
step 0] $\mu_{\overline{Y}} = \mu = 100$, $\sigma_{\overline{Y}} = \dfrac{\sigma}{\sqrt{n}} = \dfrac{15}{\sqrt{100}} = 1.5$

step i] $\overline{Y}$ picture



step ii) $z(\overline{Y}) = \dfrac{\overline{Y}value - \mu_{\overline{Y}}}{\sigma_{\overline{Y}}} = \dfrac{115 - 100}{1.5} = 10$

step iii) z picture

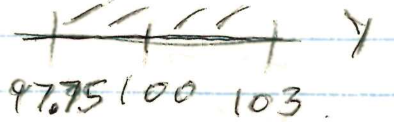


step iv) $P(z > 10) = \boxed{0.0}$

b) Find $P(97.75 < \overline{Y} < 103)$

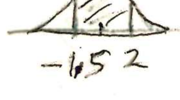
$\mu_{\overline{Y}} = 100$ $\sigma_{\overline{Y}} = 1.5$

step i]



ii) $z = \dfrac{97.75 - 100}{1.5} = -1.5$, $z = \dfrac{103 - 100}{1.5} = 2$

iii)



iv) $P(-1.5 < Z < 2)$

$= P(0 < Z < 1.5) + P(0 < Z < 2)$

$= \frac{1}{2} - P(Z > 1.5) + \frac{1}{2} - P(Z > 2)$

$= 1 - .0668 - .0228 = 0.9104.$

6] Common mistake: use $Z = \dfrac{\bar{Y}\text{value} - \mu}{\sigma}$

instead of $\dfrac{\bar{Y}\text{ value} - \mu_{\bar{Y}}}{\sigma_{\bar{Y}}} = \dfrac{\bar{Y}\text{ value} - \mu}{\sigma/\sqrt{n}}$.

p357

7) Know The sample variance

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

used to estimate σ^2.

8] Know p357 Let $Y_1, \ldots, Y_n$ be

a random sample from a normal distribution with mean μ and variance σ^2.

Then $\bar{Y}$ and S^2 are independent

and $\dfrac{(n-1)S^2}{\sigma^2} = \dfrac{1}{\sigma^2} \sum (Y_i - \bar{Y})^2$ is

χ^2_{n-1}.

9) * p360 Let Z be a standard normal RV and let W be a χ^2_r RV. If Z and W

are independent, then

$T = \dfrac{Z}{\sqrt{W/\nu}}$ has a t distribution

with ν degrees of freedom, written $T \sim t_\nu$.

10] It is crucial that Z and W are independent: note that

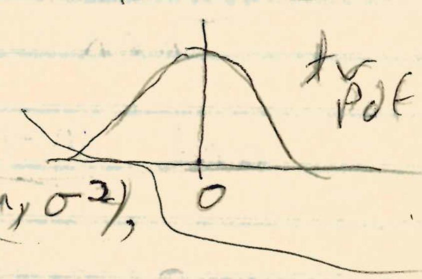
$W = Z^2$ is χ^2_1. If $T = \dfrac{Z}{\sqrt{Z^2/1}}$,

then $P(T = 1) = \frac{1}{2} = P(Z > 0)$

$P(T = -1) = \frac{1}{2} = P(Z < 0)$.

Here T is not t_1 because $W = Z^2$ and Z are dependent.

11] A t_ν pdf looks a lot like a $N(0,1)$ pdf. Both are symmetric about 0.



* p360 If $Y_1, \ldots, Y_n$ are iid $N(\mu, \sigma^2)$,

12] then $\sqrt{n}\left(\dfrac{\bar{Y} - \mu}{S}\right) \sim t_{n-1}$.

Proof Let $Z = \sqrt{n}\left(\dfrac{\bar{Y} - \mu}{\sigma}\right)$. Then Z is $N(0,1)$. Let $W = \dfrac{(n-1)S^2}{\sigma^2} \sim \chi^2_{n-1}$.

Since $\bar{Y}$ and S are independent, so are Z and W.

Hence $\dfrac{Z}{\sqrt{W/(n-1)}} \sim t_{n-1}$.

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

$= \dfrac{\sqrt{n}\,(\bar{Y}-\mu)}{S}$. $\boxed{\checkmark}$

13] *p 362 Let $W_1 \sim \chi^2_{v_1}$ and $W_2 \sim \chi^2_{v_2}$.

If W_1 and W_2 are independent, then

$\dfrac{W_1/v_1}{W_2/v_2} \sim F_{v_1, v_2}$, an F

distribution with v_1 numerator and v_2 denominator degrees of freedom.

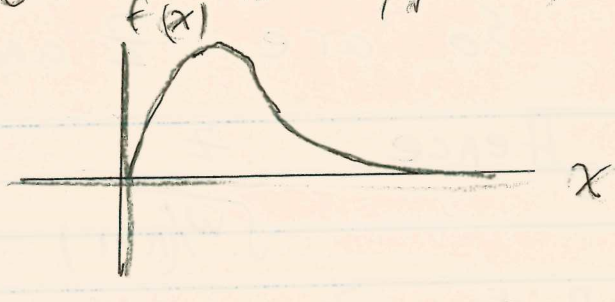
14] It is crucial that W_1 and W_2 be independent.

Let $Z \sim N(0,1)$, $W_1 = Z^2/1$, $W_2 = Z^2/1$

then $P\left(\dfrac{W_1/1}{W_2/1} = 1\right) = 1$

since $W_1 = W_2$ (so W_1 and W_2 are dependent)

15] An F_{r_1, r_2} pdf is typically right skewed.



See Fig 7.4 on p 363.

16]* p362 Suppose there are 2 independent random samples

$X_1, \ldots, X_{n_1}$ iid $N(\mu_1, \sigma_1^2)$

$Y_1, \ldots, Y_{n_2}$ iid $N(\mu_2, \sigma_2^2)$.

Let S_i^2 be the sample variance from sample i, $i = 1, 2$.

Then $\dfrac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2} \sim F_{n_1-1, n_2-1}$.

proof Let $W_1 = \dfrac{(n_1-1) S_1^2}{\sigma_1^2}$ and $W_2 = \dfrac{(n_2-1) S_2^2}{\sigma_2^2}$.

Since the X's and Y's are independent, so are W_1 and W_2. $W_1 \sim \chi^2_{n_1-1}$ $W_2 \sim \chi^2_{n_2-1}$ by point 8]. So $\dfrac{W_1/(n_1-1)}{W_2/(n_2-1)} \sim F_{n_1-1, n_2-1}$

by point 13]. But $\dfrac{W_1/(n_1-1)}{W_2/(n_2-1)} = \dfrac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2}$.

Skip examples 7.3, 7.4, 7.5, 7.6 and 7.7 for now.

These examples will make more sense after confidence intervals and hypothesis testing.

§7.3

17] Know Central Limit Theorem (CLT) Let $Y_1, Y_2, \ldots, Y_n$ be iid RV's with $EY_i = \mu$ and $V(Y_i) = \sigma^2$.
p372

Let $U_n = \sqrt{n}\left(\frac{\bar{Y}-\mu}{\sigma}\right) = \frac{\sum_{i=1}^{n} Y_i - n\mu}{\sigma\sqrt{n}}$.

$\frac{\sqrt{n}}{\sigma} = \frac{1}{\sigma/\sqrt{n}}$

Then the distribution of U_n converges to the standard normal distribution as $n \to \infty$.

18] * Note that U_n is the z score for $\bar{Y}$: $U_n = \frac{\bar{Y} - \mu_{\bar{Y}}}{\sigma_{\bar{Y}}}$.

If the CLT holds, then forwards calculations for $\bar{Y}$ can be done as long as

i) the sample $Y_1, \ldots, Y_n$ is iid

ii) $E Y_i = \mu$, $V Y_i = \sigma^2$

iii) n is large enough.

46.5

19] KNOW How large should n be to use the CLT?

i) $n \geq 1$ for Y_i normal

ii) $n \geq 5$ for Y_i close to normal

iii) If Y has a highly skewed underlying population, don't

use the normal approximation if $n \leq 29$.

iv) If $n \geq 100$, usually the CLT holds in this class.

20] Measurements from identical expt. are iid, but

answers from magazine and web surveys are not iid: Ann Landers voluntary response sample

answers from samples of convenience (eg class or fliers left at student center) are not iid.

ex] Time.com asked who the most important person of the 20th century was.

≈ 2000

Ronnie O Brien, an Irish soccer player, was the leader until a software crash. The magazine narrowed the field to 100 candidates and Elvis won.

American Idol: candidate with fewest votes is eliminated

21] Common E3 problem

Suppose $n = 10$ and $Y_1, \ldots, Y_n$ are a random sample from a population with mean $E(Y) = \mu = 100$ and standard deviation $\sigma = 15$.

a) If Y comes from a highly skewed population find $P(\bar{Y} > 115)$ if possible.

b) If Y comes from a normal population, find $P(\bar{Y} > 115)$ if possible.

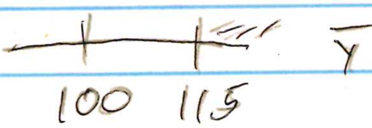
Soln

a) CLT does not apply not possible

b) CLT does apply: forwards calculation

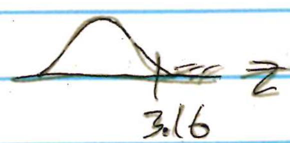
step 0) $\mu_{\bar{Y}} = \mu = 100 \quad \sigma_{\bar{Y}} = \frac{\sigma}{\sqrt{n}} = \frac{15}{\sqrt{10}} = 4.74$

step i)



ii) $z = \frac{115 - 100}{4.74} = 3.16$

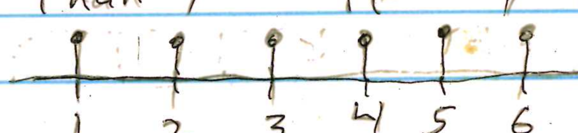
iii)



iv) table $P(z > 3.16) = 0.00$

ex] CLT says that $\bar{Y} = \frac{1}{n} \sum Y_i$ is closer to normal than Y if Y is not normal

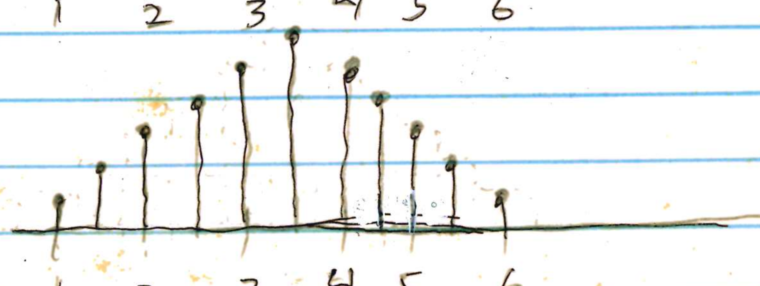
die



← heights are $\frac{1}{6}$

$n = 2$ $\quad \bar{y}$

2	3	4	5	6	7	8	9	10	11	12
1	$\frac{3}{2}$	2	$\frac{5}{2}$	3	$\frac{7}{2}$	4	$\frac{9}{2}$	5	$\frac{11}{2}$	6
$\frac{1}{36}$	$\frac{2}{36}$	$\frac{3}{36}$	$\frac{4}{36}$	$\frac{5}{36}$	$\frac{6}{36}$	$\frac{5}{36}$	$\frac{4}{36}$	$\frac{3}{36}$	$\frac{2}{36}$	$\frac{1}{36}$

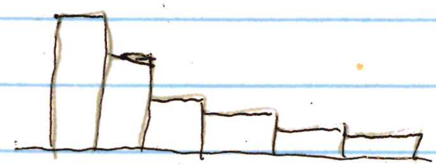


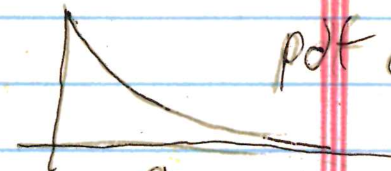
ex] p370 − 371 Y_i iid $Exp(1)$

Get 1000 samples of size n

compute $\bar{Y}_{1,n}, \ldots, \bar{Y}_{1000,n}$ and make a histogram

a) if $n=1$

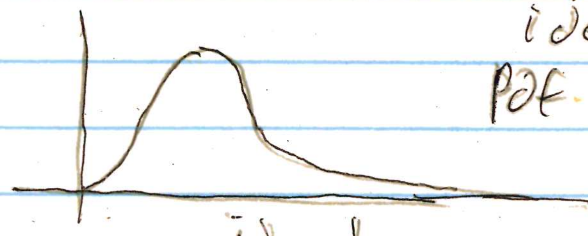




pdf of $\bar{Y}_1$

ideal

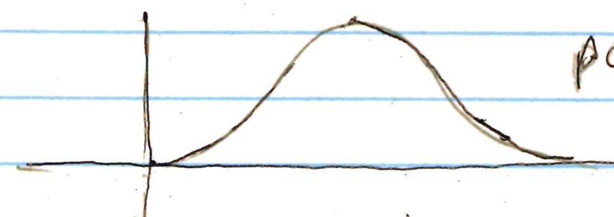
b) $n=5$



ideal pdf $\bar{Y}_5$

ideal

c) $n=25$



pdf of $\bar{Y}_{25}$

much more normal

Read examples 7.8 and 7.9.

22] Another name for the sample mean is the sample <u>average</u> or

just average. Average is also used for mean or expected value.

↑ end exam3 material   begin exam4 material

§7.5 23] The probability function of a discrete RV can be represented by a probability histogram.

ex]. die

$\frac{1}{6}$

1 2 3 4 5 6

$p(y=1) = \underbrace{\left(\frac{3}{2}-\frac{1}{2}\right)\frac{1}{6}}_{\text{area from histogram}} = \frac{1}{6}$