

Key

kindly report any errors so they may be corrected sap

THE EXERCISES BELOW ARE A PART OF YOUR PREPARATIONS FOR EXAM 1.

DO NOT TURN THEM IN.

I WILL GO OVER SOME OF THEIR SOLUTIONS IN CLASS THIS COMING WEEK OF 9-24-07. The exam is Friday 9-28-07.

THE NUMBER OF PROBLEMS BELOW IS LARGE SO BE SELECTIVE AND PREPARE THE ONES YOU NEED SOME STRENGTHENING ON.

I EXPECT THAT YOUR EXAM WILL HAVE AROUND 10 PROBLEMS, SOME WITH PARTS. YOU WILL NOT USE CALCULATORS.

IN MANY CASES I WILL ASK THAT YOU SET UP THE SOLUTION BUT DO NOT EVALUATE IT. FOR INSTANCE, YOU MAY BE ASKED TO GIVE THE SUM OR INTEGRAL DEFINING A MEAN BUT NOT EVALUATE IT. IN OTHER CASES THAT MAY BE SIMPLE TO DO AND I MAY ASK FOR IT TO BE DONE.

0. Red and Green dice are thrown. Determine:

a. $E(1/R) = (1/6)(1/1) + (1/6)(1/2) + \dots + (1/6)(1/6)$ (mean of reciprocals of R)

b. $ER = \frac{1}{6}1 + \frac{1}{6}2 + \dots + \frac{1}{6}6 = 3.5$

c. $ER^2 = \frac{1}{6}1^2 + \frac{1}{6}2^2 + \dots + \frac{1}{6}36$

d. $Var R = ER^2 - (ER)^2 = (c) - (b)^2$

e. $sd R = \sqrt{(d)}$

f. $E(R + G) = ER + EG = 7$

g. $Var(R + G) = (\text{are R and G independent?}) \stackrel{INDEP}{=} Var R + Var G = 2(d)$
YES

h. $sd(R + G) = \sqrt{(g)}$

i. $E(7R - 2G + 14) = 7ER - 2EG + 14 = 5(3.5) + 14$ DOES NOT REQUIRE INDEP

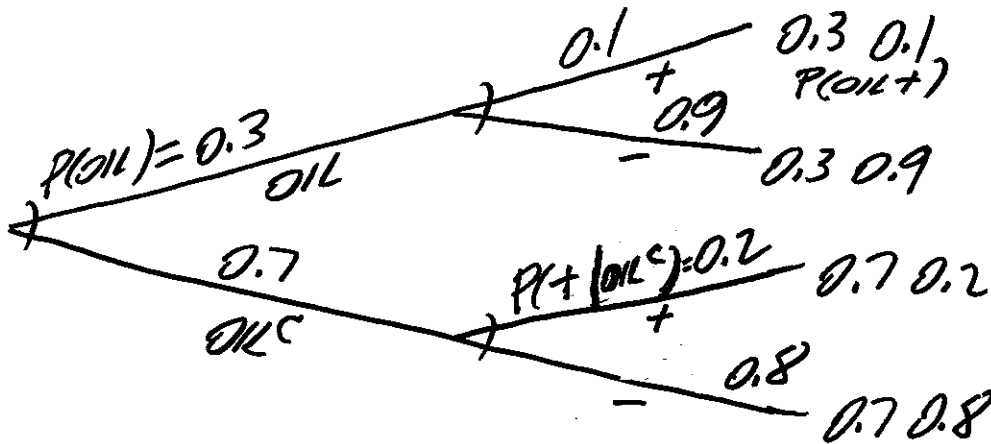
j. $Var(7R - 2G + 14) \stackrel{INDEP}{=} 7^2 Var R + 2^2 Var G$

k. $sd(7R - 2G + 14) = \sqrt{(j)}$

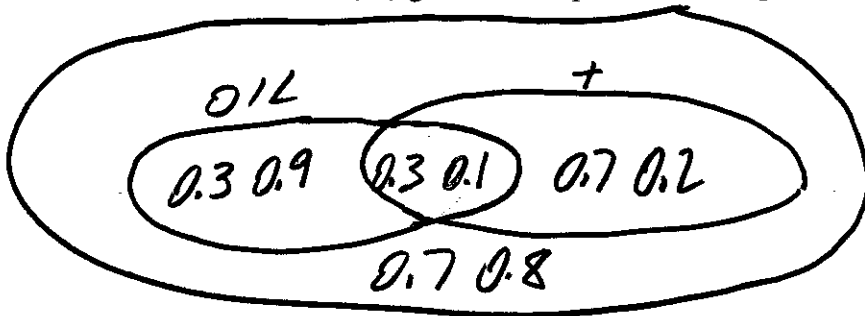
"A
FOR
TEST"

1. OIL. Suppose $P(\text{OIL}) = 0.3$, $P(+ | \text{OIL}) = 0.1$, $P(+ | \text{OIL}^c) = 0.2$.

a. Give a complete Tree Diagram for the given information. This means that two unconditional probabilities (e.g. .3) and four conditional probabilities (e.g. 0.1) appear as well as four intersection probabilities (e.g. 0.3 0.1). One label of each type must appear. For example, labels $P(\text{OIL})$, $P(- | \text{OIL})$, $P(+ | \text{OIL}^c)$, and $P(\text{OIL}+)$ might appear.



b. From tree (1a) give the complete Venn diagram for events "OIL" and "+".



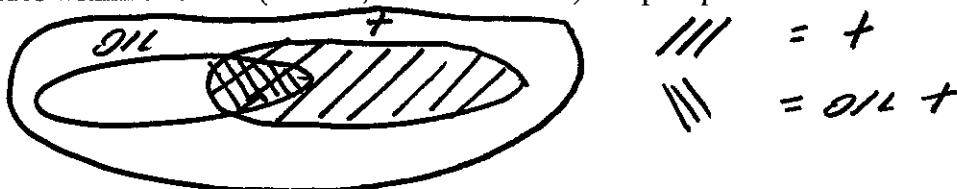
c. From (1a) determine Bayes' probability $P(\text{OIL} | +)$. Display the definition of this probability in terms of $P(\text{OIL}+)$ and $P(+)$ and show how you evaluate these two from the tree. Show how the law of total probability comes into play in evaluating $P(+)$.

$$P(\text{OIL} | +) \stackrel{\text{DEF}}{=} \frac{P(\text{OIL}+)}{P(+)} \leftarrow \text{PR OF BOTH}$$

$$= \frac{0.3 \cdot 0.1}{0.3 \cdot 0.1 + 0.7 \cdot 0.2}$$

\uparrow PR OF CONDITIONING EVENT
 \uparrow \uparrow
 $P(\text{OIL}+) + P(\text{OIL}^c+)$
 TOTAL PR = $P(+)$

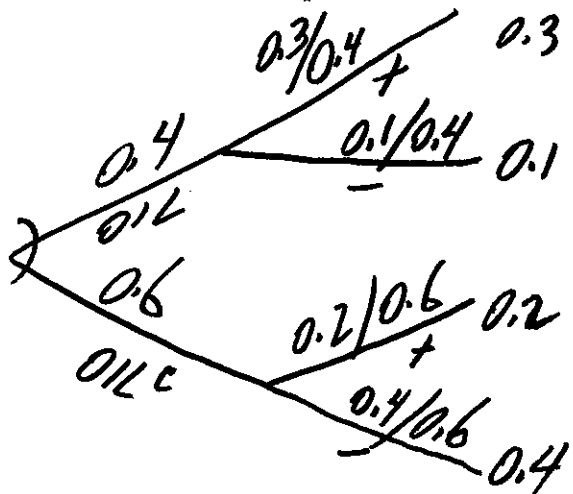
d. In the Venn diagram (1b) shade the sample space that comes into play when we obtain the conditional probability $P(\text{OIL} | +)$. Cross hatch the part of the event "OIL" that resides within this new (smaller, and conditional) sample space.



e. Should $P(\text{OIL} | -)$ be smaller or larger than $P(\text{OIL} | +)$ and why? Confirm your reasoning by calculating $P(\text{OIL} | -)$.

$P(\text{OIL} | -)$ SHOULD BE SMALLER THAN $P(\text{OIL} | +)$ SINCE WE WANT "+" TO BE INDICATIVE OF THE PRESENCE OF OIL.

2. Another OIL problem has the following probabilities for the end branches of its tree:



*$\text{OIL} = \text{OIL}^+ + \text{OIL}^- = 0.3 + 0.1 = 0.4$
 $+/\text{OIL} = \text{OIL}^+ / \text{OIL} = 0.3/0.4$*

OIL+ 0.3

OIL- 0.1

OIL^{c+} 0.2

OIL^{c-} *$1 - 0.3 - 0.1 - 0.2 = 0.4$*

Fill out the last line above and a **complete** tree diagram for this information. For example, $P(\text{OIL}) = P(\text{OIL}^+) + P(\text{OIL}^-)$ etc. So you can get the tree from the Venn.

3. Refer to (2). Suppose it costs 40 to test, 150 to drill, and the gross return from OIL is 900 (oil must be there and we must drill otherwise the gross return is zero). Determine $E X$, $\text{Var } X$, $\text{SD } X$, where random variable (r.v.) X denotes the NET return from the policy

*OIL+ 0.3
 OIL- 0.1
 OIL^{c+} 0.2
 OIL^{c-} 0.4*

"drill, do not test." Be sure to calculate values x and $p(x)$ for the two contingencies. Show all terms contributing to each value x .

	$p(x)$	x
OIL	0.4	$-150 + 900$
OIL ^c	0.6	$-150 + 0$

$$EX = 0.4(750) + 0.6(-150) \quad \left. \begin{array}{l} \\ \\ \end{array} \right\} \Rightarrow \text{Var } X = E(X^2) - (EX)^2$$

$$E(X^2) = 0.4(750)^2 + 0.6(150)^2 \quad \left. \begin{array}{l} \\ \\ \end{array} \right\} \Rightarrow \text{SD } X = \sqrt{\text{Var } X}$$

4. Refer to (2). Suppose it costs 40 to test, 150 to drill, and the gross return from OIL is 900 (oil must be there and we must drill otherwise the gross return is zero). Determine $E Y$, $\text{Var } Y$, $\text{SD } Y$, where random variable (r.v.) Y denotes the NET return from the policy "test, but drill only if the test is +." Be sure to calculate values y and $p(y)$ for the four contingencies. Show all terms contributing to each value y . You need not give the tree.

*CORRECTED
TYPO
(Y NOT X)* →

	$p(y)$	y
OIL+	0.3	$-40 - 150 + 900 = 710$
OIL-	0.1	$-40 - 0 + 0 = -40$
OIL ^c +	0.2	$-40 - 150 + 0 = -190$
OIL ^c -	$\frac{0.4}{1}$	$-40 - 0 + 0 = -40$

Some people think that larger expectation is more desired. Which policy, X or Y , has the larger expectation?

$$EY = 710(0.3) - 40(0.1) - 190(0.2) - 40(0.4)$$

5. Refer to (4). Suppose testing costs rise to 70. Then $E Y$ is no longer valid. Let W denote the net return from the policy "test, but drill only if the test is +" in the situation of the greater testing cost. Write an equation expressing W in terms of Y . Use that equation to evaluate $E W$ from $E Y$ (without re-figuring everything as was done in (4)).

$$W = Y - 30 \leftarrow \text{ADD'L COST TO TEST (IN ALL CASES)}$$

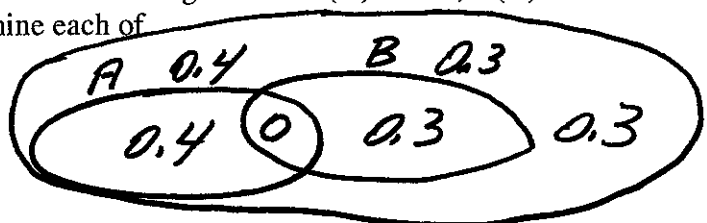
$$EW = EY - 30$$

$$\text{Var } W = \text{Var}(Y - 30) = \text{Var } Y$$

$$\text{SD } W = \text{SD } Y$$

6. Events A , B are disjoint. Prepare a Venn diagram for $P(A) = 0.4$, $P(B) = 0.3$. Identify the correct formulas and determine each of

$$P(AB) = P(\phi) = 0$$



$$P(A \text{ intersection } B) = P(AB) = P(\emptyset) = 0 \text{ (DISJOINT)}$$

$$P(A \text{ and } B) = P(\emptyset) = 0$$

$$P(A \text{ union } B) = P(A) + P(B) - P(AB) = P(A) + P(B) = 0.4 + 0.3$$

$$P(A \text{ or } B) = \text{SAME AS ABOVE} \nearrow$$

$$P(A | B) = P(AB) / P(B) = 0 / P(B) = 0$$

$$P(B | A) = P(AB) / P(A) = 0 / P(A) = 0$$

7. Events A, B are **independent**. Prepare a Venn diagram for $P(A) = 0.4$, $P(B) = 0.3$. Identify the correct formulas and determine each of

$$P(AB) = \overset{\text{GEN'L}}{P(A)P(B|A)} = \overset{\text{INDEP}}{P(A)P(B)} = 0.4 \cdot 0.3 = 0.12$$

$$P(A \text{ union } B) = P(A) + P(B) - P(AB) = 0.4 + 0.3 - 0.12$$

$$P(A | B) = \overset{\text{INDEP}}{P(A)} = 0.4$$

$$P(B | A) = \overset{\text{INDEP}}{P(B)} = 0.3$$

$$P(B | \text{not } A) = \overset{\text{INDEP}}{P(B)} = 0.3$$

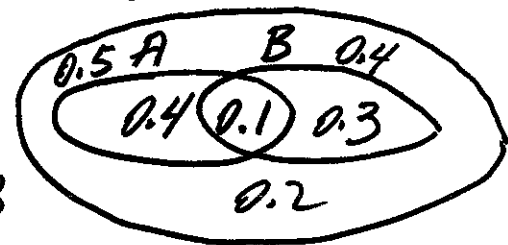
$$P(B | \text{not } B) = \overset{\text{INDEP}}{P(B^c)} / P(B^c) = 0 / P(B^c) = 0$$

8. For events A, B we are given $P(A) = 0.5$, $P(B | A) = 0.2$, $P(B) = 0.4$. Prepare a Venn diagram. These three pieces of information are enough to do it!

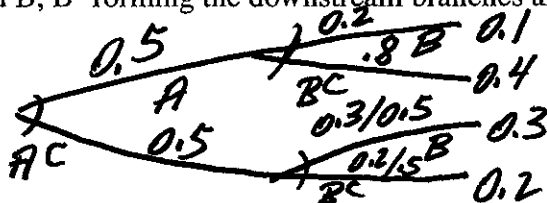
$$P(AB) = P(A)P(B|A) = 0.5(0.2) = 0.1$$

$$P(AB^c) = P(A) - P(AB) = 0.5 - 0.1 = 0.4$$

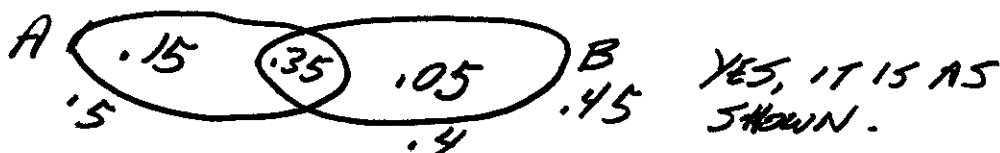
$$P(A^cB) = P(B) - P(AB) = 0.4 - 0.1 = 0.3$$



9. Refer to (8). Prepare a tree diagram with A, A^c forming the initial (left) branching and B, B^c forming the downstream branches after each of A, A^c .



10. For events A, B we are given $P(A) = 0.5$, $P(B | A) = 0.7$, $P(B) = 0.4$. Is this possible, why? IF SO, $P(AB) = P(A)P(B|A) = 0.5 \cdot 0.7 = 0.35$



11. A box contains colored balls [R R R R G G G Y Y Y Y]. Balls will be withdrawn **without** replacement and with equal probability for those remaining.

a. $P(R_1) = \frac{4}{12}$

b. $P(R_2) =$ (naive answer, no calculation) SAME AS $P(R_1) = \frac{4}{12}$

c. Calculate $P(R_2)$ using total probability and multiplication rules.

$$P(R_2) = P(R_1 R_2) + P(R_1^c R_2) \text{ TOTAL PR}$$

$$= P(R_1) P(R_2 | R_1) + P(R_1^c) P(R_2 | R_1^c)$$

$$= \frac{4}{12} \frac{3}{11} + \frac{8}{12} \frac{4}{11} = \frac{4}{12} \text{ (k. it!)}$$

d. $P(R_2 | R_1) =$ (naive answer, no calculation)

GIVEN R1
DRAW 2 IS
FROM

[R R R G G G Y Y Y Y]

e. $P(R_1 | R_2) =$ (naive answer, no calculation)

SAME AS $P(R_2 | R_1)$ ORDER OF THE
DEAL DOES NOT
MATTER
 $= \frac{3}{11}$

$P(R_2 | R_1) =$ ASSUMPTION ON $\frac{3}{11}$
OF SAME PR & W/
EQUAL PR FROM
THOSE REMAINING

f. If many people independently reproduce this ball selection experiment and 400 of them have R2, around how many people participated in the experiment?

$\frac{400}{\# \text{ PARTICIPATING}} \approx P(R_2) = \frac{4}{12}$ SO # PARTICIPATING
 $\approx (400) \frac{12}{4} = 1200$

g. Refer to (e). If many people independently reproduce this ball selection experiment and 400 of them have R2, around how many of those 400 have R1?

$$P(R_1 R_2) / P(R_2) \approx \frac{\# \text{ WITH } R_1 R_2}{\# \text{ WITH } R_2}$$

h. $P(R_1 G_2 R_3 Y_4 G_5) =$

$$P(R_1) P(G_2 | R_1) P(R_3 | R_1 G_2) P(Y_4 | R_1 G_2 R_3) P(G_5 | R_1 G_2 R_3 Y_4)$$

$$\frac{4}{12} \frac{3}{11} \frac{3}{10} \frac{5}{9} \frac{2}{8}$$

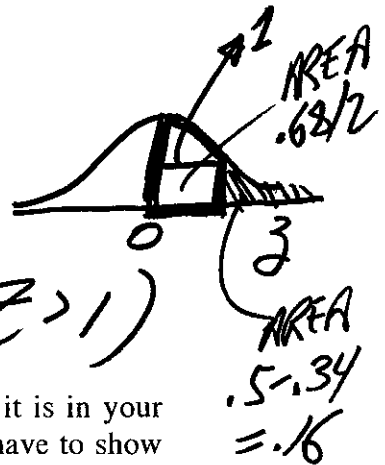
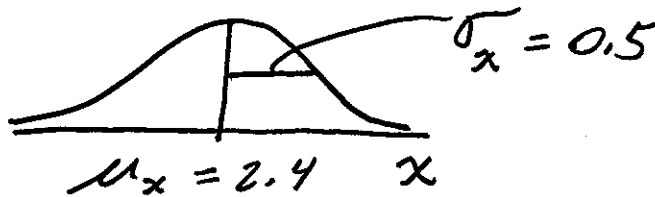
i. $P(R_1 G_2 R_3 Y_4 G_5 G_6 G_7) = 0$

↑ ↑ ↑ ↑

NOT POSSIBLE
IF G2 G5 G6
SINCE ONLY
3 G EXIST.

12. A process producing transmissions is thought to be under statistical control for score x = overall shift lag (tenths of seconds) with $\mu = 2.4$ and $\sigma = 0.5$.

a. Sketch the distribution of x scores for the population of such transmissions. Be sure to label the mean and sd of your curve as recognizable elements in your sketch.

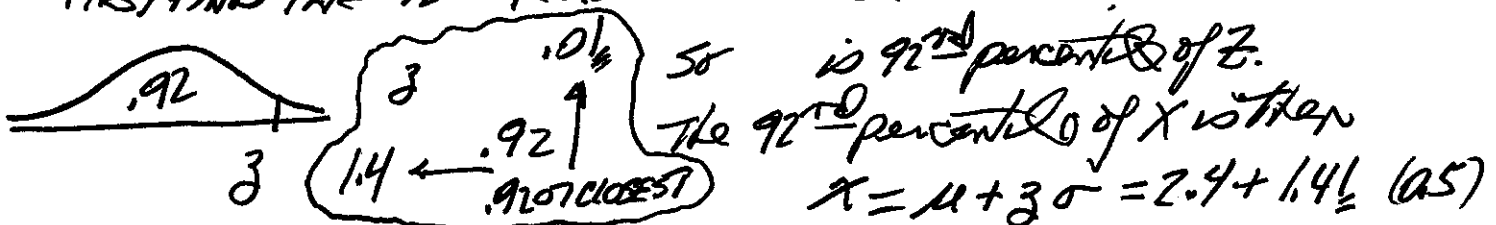


b. What percentage of transmissions have overall shift lag $x > 2.9$?

$$P(X > 2.9) = P\left(\frac{X - \mu_x}{\sigma_x} > \frac{2.9 - 2.4}{0.5}\right) = P(Z > 1) = .16$$

c. Give the 92nd percentile of overall shift lag x . Illustrate where it is in your sketch and what the 92 represents in the sketch. On an exam you will have to show how you do it.

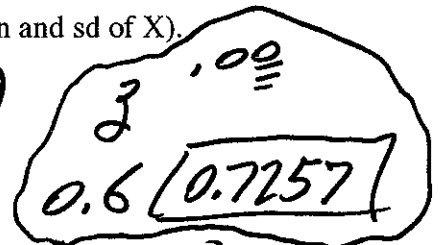
FIRST FIND THE 92nd PERCENTILE OF Z.



13. Refer to (12). It is decided to check up on production by pulling $n = 10$ transmissions and examining them to determine their shift lag scores X_1, \dots, X_{10} . IF THE PROCESS IS TRULY IN STATISTICAL CONTROL AS JUST DESCRIBED then these scores represent independent samples having the specified distribution NO MATTER HOW THE TRANSMISSIONS ARE SELECTED SO LONG AS SELECTION IS NOT KEYED TO THEIR x -SCORES. That is part of what it means to be in control.

a. Evaluate $P(X_1 < 2.70)$ (you need to standardize by the mean and sd of X).

$$= P\left(Z < \frac{2.7 - 2.4}{0.5}\right) = P(Z < .6) = 0.7257$$



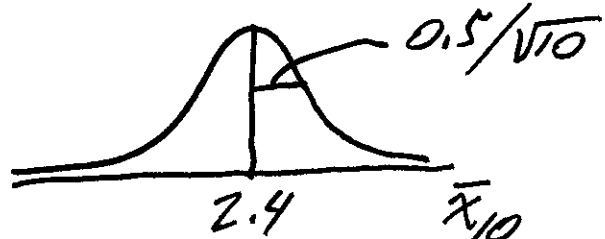
b. Evaluate $P(X_1 < 2.70 \text{ and } X_2 > 2.4)$.

$$X_1 \text{ INDEP OF } X_2. \rightarrow P(X_1 < 2.7) P(X_2 > 2.4) = P\left(Z < \frac{2.7 - 2.4}{0.5}\right) \left(1 - P\left(Z < \frac{2.4 - 2.4}{0.5}\right)\right)$$

c. Sketch EXACT distribution of the sample mean \bar{x} for $n = 10$ (remember, the sample mean \bar{x} of independent normally distributed r.v. is itself

$$E \bar{X}_{10} = \mu_{\bar{X}_{10}} = \mu_x = 2.4$$

$$\sigma_{\bar{X}_{10}} = \frac{\sigma_x}{\sqrt{10}} = \frac{0.5}{\sqrt{10}}$$



normally distributed). Be sure to indicate the mean and sd of \bar{x} BAR as recognizable elements in your sketch.

d. Evaluate $P(\bar{x}\text{BAR} < 2.70)$ (you need to standardize by the mean and sd of $\bar{x}\text{BAR}$).

$$P(\bar{X}_{10} < 2.7) = P\left(Z < \frac{2.7 - \mu_{\bar{X}_{10}}}{\sigma_{\bar{X}_{10}}}\right)$$

$$= P\left(Z < \frac{2.7 - 2.4}{0.5/\sqrt{10}}\right)$$

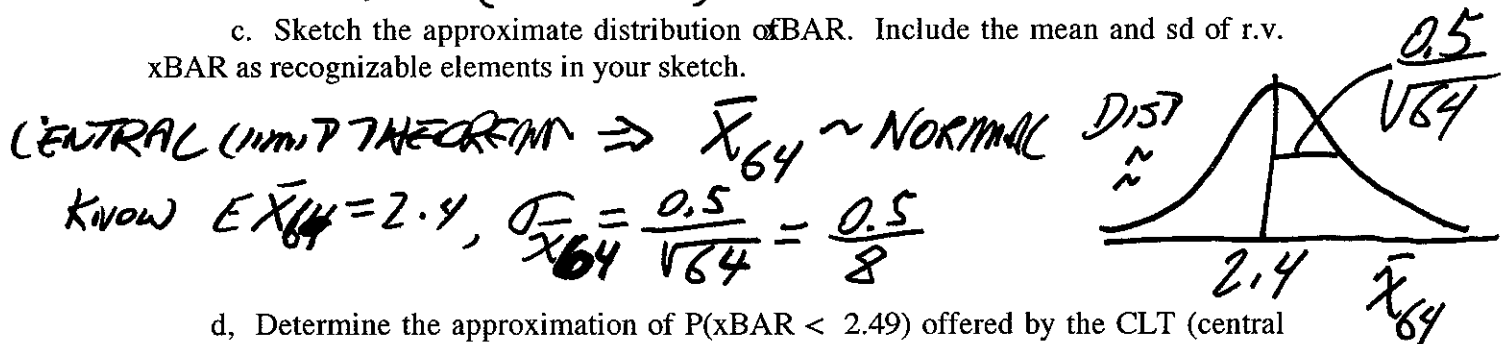
14. Refer to (12). The process is **not known to be under statistical control**. To check the process we select a **random sample** of $n = 64$ transmissions. As before, we are operating under the assumption that the population mean x -score is 2.4 and the population sd is 0.5.

a. We have no way of evaluating $P(X_1 < 2.70)$. Why? **DO NOT KNOW THE DISTRIBUTION OF X_1 (IT IS NOT, IN GENERAL, DETERMINED BY ITS MEAN 2.4 AND S.D 0.5).**

b. If $P(X_1 < 2.70)$ and $P(X_2 > 2.4)$ are known, what is $P(X_1 < 2.70 \text{ and } X_2 > 2.4)$?

RANDOM (WITHA-REP) SAMPLE $\Rightarrow X_1, X_2$ INDEP.
 $\Rightarrow P(X_1 < 2.70) P(X_2 > 2.4)$ **BOTH OF WHICH ARE ASSUMED KNOWN.**

c. Sketch the approximate distribution of $\bar{x}\text{BAR}$. Include the mean and sd of r.v. $\bar{x}\text{BAR}$ as recognizable elements in your sketch.



d. Determine the approximation of $P(\bar{x}\text{BAR} < 2.49)$ offered by the CLT (central limit theorem).

$$P(\bar{X}_{64} < 2.49) \approx P\left(Z < \frac{2.49 - 2.4}{0.5/\sqrt{64}}\right)$$

e. If the process truly has mean 2.4 and sd 0.5 what would you think if the sample mean \bar{x}_{64} (of 64 independently sampled transmission lag times) is found to be 2.68? ← SEEMS LARGE!!

$$P(\bar{X}_{64} > 2.68) \approx P(Z > \frac{2.68 - 2.4}{0.5/\sqrt{64}}) = P(Z > \frac{.28(8)}{0.5})$$

$$= P(Z > .28(16)) \sim 0$$

15. Refer to (14). The population of transmissions has size $N = 600$ (made this week perhaps). Suppose we sample $n = 64$ transmissions WITHOUT replacement and with equal probability on those remaining.

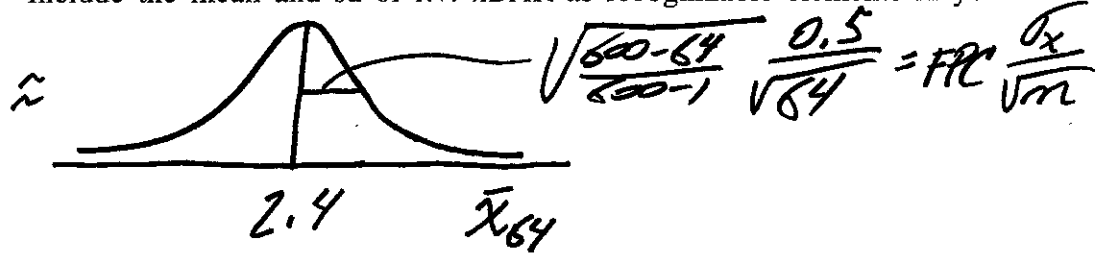
a. Calculate the FPC (finite population correction) root $((N-n)/(N-1))$.

$$FPC = \sqrt{\frac{600-64}{600-1}}$$

b. If $P(X_1 < 2.70)$ and $P(X_2 > 2.4)$ were known, could you determine $P(X_1 < 2.70 \text{ and } X_2 > 2.4)$? NO.

$P(X_2 > 2.4 | X_1 < 2.7)$
CANNOT BE DETERMINED EXACTLY

c. Sketch the approximate distribution \bar{x}_{64} for this **without-replacement** sampling. Include the mean and sd of r.v. \bar{x}_{64} as recognizable elements in your sketch.



d. Determine the approximation of $P(\bar{x}_{64} < 2.49)$ offered by the CLT (central limit theorem). Is it much changed from 14.d.?

$$\approx P(Z < \frac{2.49 - 2.4}{\sqrt{\frac{600-64}{600-1}} \frac{0.5}{\sqrt{64}}})$$

e. If the process truly has mean 2.4 and sd 0.5 what would you think if the sample mean \bar{x}_{64} (of 64 independently sampled transmission lag times) is found to be 2.68? ← AGAIN, SEEMS LARGE

$$P(\bar{X}_{64} > 2.68) \approx P(Z > \textcircled{1}) \sim 0$$

16. Consider $f(x) = 5/(x^6)$ for $x > 1$, and zero otherwise.

a. Verify that $f(x)$ is indeed a non-negative function having integral one over the entire line and thus a probability density.

$$\int_1^{\infty} f(x) dx = \int_1^{\infty} \frac{5}{x^6} dx = \left[-\frac{1}{x^5} \right]_1^{\infty} = 1$$

↑ NON-NEG

b. Determine the cumulative distribution function $F(x) = \text{integral of } f(x) \text{ from } 1 \text{ to } x$, for values $x > 1$.

$$\int_1^x f(u) du = -\frac{1}{2.5} \Big|_1^x = 1 - \frac{1}{x^{2.5}} \text{ for } x > 1.$$

c. Using (b) for convenience, determine $P(X > 3)$.

$$= 1 - F(3) = 1 - \left(1 - \frac{1}{3^{2.5}}\right) = \frac{1}{3^{2.5}}$$

d. Using (b) for convenience, determine $P(2 < X < 5.1)$.

$$= F(5.1) - F(2) = \left(1 - \frac{1}{5.1^{2.5}}\right) - \left(1 - \frac{1}{2^{2.5}}\right)$$

e. Calculate $E X$, $\text{Var } X$, and $\text{sd } X$.

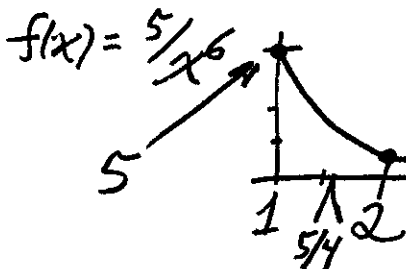
$$E X = \int_1^{\infty} x \cdot f(x) dx = \int_1^{\infty} x \cdot \frac{5}{x^6} dx = \int_1^{\infty} \frac{5}{x^5} dx = \left. -\frac{5}{4x^4} \right|_1^{\infty} = \frac{5}{4}$$

$$E X^2 = \int_1^{\infty} x^2 \cdot \frac{5}{x^6} dx = \int_1^{\infty} \frac{5}{x^4} dx = \left. -\frac{5}{3x^3} \right|_1^{\infty} = \frac{5}{3}$$

$$\text{Var } X = E X^2 - (E X)^2 = \frac{5}{3} - \left(\frac{5}{4}\right)^2 = \frac{5}{48}$$

$$\sigma_x = \sqrt{\text{Var } X} = \sqrt{5/48}$$

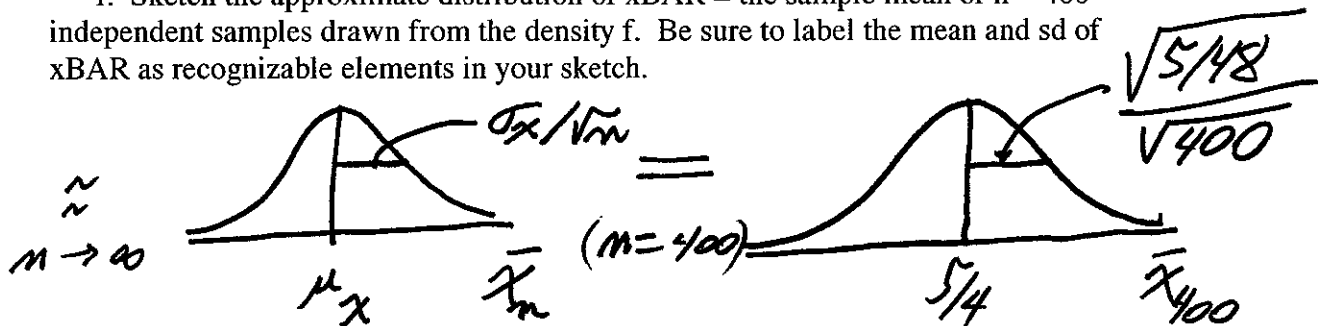
f. Sketch f over $x > 1$. Does $E X$ calculated for (e) seem correct by eye?



NOT EASY TO CONFIRM BY EYE BUT LOOKS "REASONABLE"

$$f(1) = 5 \quad f(2) = \frac{5}{2^6} = \frac{5}{64}$$

f. Sketch the approximate distribution of \bar{x} (the sample mean of $n = 400$ independent samples drawn from the density f). Be sure to label the mean and sd of \bar{x} as recognizable elements in your sketch.



17. A space object dies by stray radiation and has a lifetime that is exponentially distributed.

a. What is the probability that the object lives longer than its expected lifetime? The answer is not one-half and does not depend upon the expected lifetime.

FOR EXPONENTIAL, $P(X > x) = e^{-\lambda x}$ for $\lambda = 1/EX$.

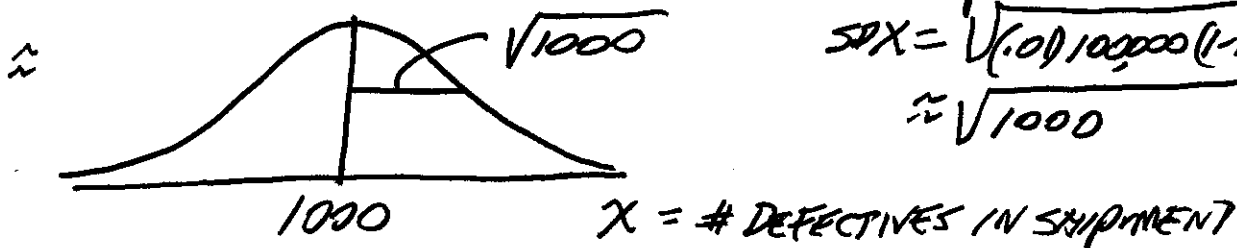
So $P(X > \frac{1}{\lambda}) = P(X > \frac{1}{\lambda}) = e^{-\lambda \cdot \frac{1}{\lambda}} = e^{-1}$, FOR EVERY λ .

b. If we have 800 similar objects in space, around how many of them will be alive at twice their expected individual lifetimes?

$800 P(X > 2EX) = 800 P(X > \frac{2}{\lambda}) = e^{-\lambda \cdot \frac{2}{\lambda}} 800 = 800e^{-2}$

18. We've a large shipment of $n = 100,000$ items. Each, independently, has probability $p = 0.01$ of being defective.

a. Let X denote the number defective in the entire shipment. The distribution of X is binomial; with $n = 100,000$ and $p = 0.01$. It is not tabled. You can however sketch its normal approximation using $EX = np$ and $sd X = \sqrt{np(1-p)}$. Do so, including the mean and sd as recognizable elements in your sketch.



b. Since n is large and p is small there is another discrete distribution (famously applicable to counts of rare events) approximating the true distribution (a) but having only a single parameter. It is the Poisson, and its parameter is the mean np of the binomial. The sd of the Poisson is the square root of its mean. The Poisson having a given mean μ is approximated well by the normal if $\mu > 3$. Sketch the approximate normal for this case, comparing with (a).

POISSON w/ $\lambda = np = 100,000 (.01) = 1000$

$\lambda > 3$ so WE OFFER THEN NORMAL APPROX OF POISSON \approx

