J. Sanchez
UCLA Department of Statistics

EVERYTHING IN THIS BOX MUST APPEAR ON A FRONT, PAGE 1, OF YOUR SUBMITTED EXAM. YOU MAY WRITE EVERYTHING BY HAND ON PAGE 1 OF YOUR EXAM. ONLY WHAT IS IN THIS BOX MUST BE ON YOUR PAGE 1 OF YOUR EXAM.

LAST NAME: –Dean––––––––– **FIRST NAME:** –Nicholas––––– **ID:** 105312823

Write a statement like this below and add your full signature (in English) below it. The statement must be exactly like this, with your name on it and signature. Do not cut the phrases or simplify.

I, ––(your name here)Nicholas Dean–––––––––sign to confirm that this exam reflects my work and only my work, that I have not consulted with anyone or anything except the class material posted in CCLE, the textbook, and Cognella active learning and that I have taken the time specified in the instructions or very close to that time to complete the exam from the moment that I first looked at it until it was in Gradescope. I also confirm that I have adhered to Section 102.01 or 102.02 of the UCLA Student Conduct Code and that I have and will not share this exam with anything or anyone.

YOUR SIGNATURE (in English):– Nicholas Dean––––––––

INSTRUCTIONS

(points deducted for not following these instructions and those posted in the midterm folder)

(1) The exam must be submitted to Gradescope link for the exam **before 8 PM on 3/13/2021 Los Angeles Time.**
(2) **This is a two hour exam but you have three hours from the moment you click on the pdf file or the gradescope submission link for the first time, until it is showing submitted in gradescope. The time at which you access gradescope, the pdf or either or both will be logged in by CCLE. If you click in any of the two, you must start and do the exam right away. "Accidental clicks" are clicks, and the clock starts for you with that click, if that happens to you.** The three hours includes the time it takes you to scan and submit the exam. You can choose any three hours between 8:00 AM on 3/12/2021 and 8 PM on 3/13/2021 to do your exam, but you must submit before the deadline. No excuses will be accepted because you wait until the last minute to look at, download and (or) submit your exam. Start early and submit early to prevent problems.
(3) You must work on your own. No group work allowed, no consulting with anyone or anything allowed. No sharing of information allowed. You may use all the material in our CCLE course web site, and the required textbook, including Cognella active learning materials that come with the textbook. You may not talk to other students in the class until the exam window ends at 8 PM on 3/13/2021, even if you all finished the exam. The exam may not be shared with anyone or anything outside UCLA, during or after the exam time window. See note on conduct at the bottom of the instructions.
(4) Do not contact the TA or anybody else regarding class material or anything regarding the exam and the course during the time window allowed for the exam. You may contact only me, Dr. Sanchez, within non-sleep time, but first check the Q& A file and the instructions given on these pages. If the answer to what you asked is there you will not receive an answer from Dr. Sanchez by email.
(5) There is a file called Q&A, where I will be posting anything worthwhile that may benefit the whole class, such as a typo found. You will look at that file before you email me (Dr. Sanchez). Read the instructions of the exam well to avoid wasting time asking questions that are answered in the instructions.
(6) The exam has two parts.
   - Part I (25 points, 1 point for each question) is 25 multiple choice questions. *Answer on the table given on page 3 or a table exactly like that written by hand. Required.* You must write the letter of the chosen answer or answers (A, B, C,...etc). Read MC instructions on page 3 to avoid losing points for just not reading the instructions.

- Part II (20 points) problems where you must show work. Work is 80% or more of your grade, as in homework. Make sure you read the instructions carefully.

(7) **Uploading to (Gradescope)**: You will scan and upload to the gradescope link in the final exam folder a pdf file called lastname-uclaID-final.pdf, containing, in this order, and telling gradescope the page where things are (scan your work using the adobe scan app downloaded to your phone,as in homework):

- (i)The front page (you may either print the one on page 1 of the exam or make your own on your notebook. The page must contain ONLY everything that appears in the framed (boxed) part of the top of page of the exam (name, ID, exact statement saying what the paragraph says and full signature). This will be page 1 of your pdf file submission. There will be points deduction if this page is not submitted. Do not put answers for questions or anything else on this page.
- (ii) the page with ONLY the table containing the multiple choice answer(s) and the letter chosen. Work is not required. You must use the table on page 3, printed, or make one of your own exactly like that one, drawn by hand. This will be page 2 of your pdf submission. Warning: if your table is not like that one on page 3 (i.e., with 5 columns) points will be deducted from your exam.
- The remaining pages of your submitted exam will contain the work and answers for the questions that require work, in the order given. But you do not need to copy the question. You must indicate to gradescope the page number where the question appears. It is not necessary to have each question in a separate page but points will be deducted if you do not indicate the page number where the answer is when you submit to gradescope. Students using the pdf file posted to enter their answers must use only the space provided. **Read the note regarding the calculations for the normal in the instructions for the work questions**

(8) Only gradescope is allowed as a form of submission. Allow yourself enough time to be able to upload on time. I will not accept emails with files or any other form of submission. No late exams will be allowed.

(9) Any indication that you did not follow these instructions or Sections 102.01 or 102.02 of the UCLA Student Code of Conduct

`https://www.deanofstudents.ucla.edu/portals/16/documents/uclacodeofconduct_rev030416.pdf`

or indication that you shared information or work with others will result in an F in the class and this exam, and a virtual visit to the Dean of Students office.

(10) Indication that this exam has been posted or shared with anyone during or after the exam window ends will result in a visit to the Dean of Students Office. The exam can never be shared with anyone or anything outside CCLE, Stat 100A, Sanchez.

Part I. MULTIPLE CHOICE ANSWERS. WRITE YOUR ANSWER FOR QUESTIONS 1-25 ON A TABLE EXACTLY LIKE THE ONE GIVEN BELOW, WITH 5 COLUMNS AND 5 QUESTIONS PER COLUMN. YOU MAY WRITE A TABLE LIKE THIS BY HAND ON YOUR PAPER OR SUBMIT THIS PAGE. SHOWING WORK IS NOT REQUIRED IN THE MULTIPLE-CHOICE-QUESTIONS SECTION OF THE EXAM.

ONLY THIS TABLE MUST BE ON PAGE 2 OF YOUR EXAM

Question(Q)	Answer	Q	Answer	Q	Answer	Q	Answer	Q	Answer
Q 1	A	Q 6	D	Q 11	C	Q 16	C	Q 21	A
Q 2	A	Q 7	B	Q 12	D	Q 17	A	Q 22	C
Q 3	B	Q 8	A	Q 13	C	Q 18	C	Q 23	B
Q 4	D	Q 9	A	Q 14	C	Q 19	B	Q24	C
Q 5	B	Q 10	D	Q 15	D	Q 20	E	Q25	C

Table 1: There is a box for each question. Write your answer for each question inside the box where the question number is. The table you submit must be exactly like this, with 5 columns and 5 questions per column. Other formats will be deducted points. Only the table must be on this page.

PART I. Multiple choice questions. *Showing work is not required in the multiple choice questions. Write all your answers to the multiple choice questions in a table exactly identical (5 columns) to the one on page 3. Other table formats or orientation of the table will receive point deductions.* **There is only one answer to each question. If your calculator does not give more than 2 or 3 decimals, and the answer is numeric, choose the closest answer.**

Question 1. Which of the following defines a moment generating function for a random variable X?

(a) $M_X(t) = \sum_x e^{tx} P(X = x)$

(b) $M_X(t) = e^{tx} \int_x f(x)dx$

(c) $M_X(t) = \int_x f(e^{tx})dx$

(d) $M_X(t) = P(e^{tx})$

(e) None of the above.

Question 2. The voltage of a circuit is a random variable denoted by X. A particular technique for measuring the voltage gives readings which are normally distributed with expected value $\mu_X = 115$ volts and standard deviation $\sigma_X = 5$ volts. The average of four readings of the voltage of the circuit has

(a) smaller probability of differing from μ_X by at least 3 volts than an individual reading.

(b) larger probability of differing from μ_X by at least 3 volts than an individual reading.

(c) ~~the same probability of differing from μ_X by at least 3 volts than an individual reading.~~

(d) less accuracy as a measure of the voltage of the circuit than an individual reading.

(e) None of the above.

Question 3. Consider a set of independent and identically distributed random variables $X_1, X_2, \ldots\ldots, X_n$, where $\mu_{X_i} = \mu, \sigma^2_{X_i} = \sigma^2, \quad i = 1, 2, , \ldots, n$. Determine what the following expression is equal to.

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

$$E\left[\sum_i X_i^2 + n(\bar{X})^2 - 2n(\bar{X})^2\right].$$

$n\sigma^2 + \cancel{n\mu^2} - \sigma^2 - \cancel{n\mu^2}$

$n\sigma^2 - \sigma^2$

$\sigma^2(n-1)$

(a) $\sigma^2 + \mu^2$

(b) $(n-1)\sigma^2$

(c) $n\mu$

(d) $\frac{n-1}{n}\sigma^2$

(e) None of the above

$E[\sum_i X_i^2 - n(\bar{X})^2]$

$E[\sum_i X_i^2] - n E[\bar{X}^2]$

$E[\sum_i X_i^2] - n(\frac{\sigma^2}{n} + \mu^2)$

$E[X_1^2 + X_2^2 + \ldots + X_n^2] - \sigma^2 - n\mu^2$

$n(\sigma^2 + \mu^2) \quad -\sigma^2 - n\mu^2$

Question 4. A sample space contains the following outcomes:

$$S = \{1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16\}$$

and we define the following events

$A^C = \{2, 5, 6, 7, 8, 11, 12, 13, 14, 15, 16\}$

$$A = \{\cancel{1}, \cancel{3}, \cancel{9}, \cancel{10}, \cancel{4}\} \quad B = \{\cancel{3}, \cancel{2}, \cancel{8}, \cancel{5}, \cancel{9}\} \quad C = \{15, 13, 16\},$$

then the event

$B^C = \{1, 4, 6, 7, 10, 11, 12, 13, 14, 15, 16\}$

$$(A \cap B^c) \cup (A^c \cap B)$$

is equal to

$(\cancel{1}, \cancel{10}, \cancel{4}) \cup (\cancel{2}, \cancel{8}, \cancel{5})$

$(1, 2, 5, 8, 10, 4)$

(a) $\{1, 4, 10\}$

(b) $\{6, 7, 11, 12, 13, 14, 15, 16\}$

(c) the empty set, the events are mutually exclusive

(d) $\{1, 2, 4, 5, 8, 10\}$

(e) $\{0, 2, 1\}$

Question 5. Let

$$f(x, y) = 15e^{-(3x+5y)}, \qquad x \geq 0, y \geq 0$$

Which of the following statements is correct?

(a) $P(X > 3, Y > 3) = F_x(3)F_y(5)$, where F denotes cumulative distribution

(b) $P(X < 2 \mid Y = 4) = 0.9975$

~~(c) $Cov(X, Y) = 1.34$~~

~~(d) $f(Y \mid X = 3) = \frac{2x}{y^2}, \quad y > 3$~~

~~(e) $f(y)$ is uniform~~

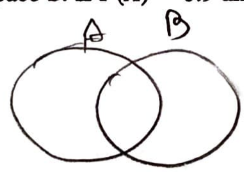
$f(y) = 5e^{-5y} \quad y \geq 0$ $\qquad f(y|x) = \frac{15e^{-3x}e^{-5y}}{3e^{-3x}}$

$f(x) = 3e^{-3x} \quad x \geq 0$ $\qquad f(y|x) = 5e^{-5y}$

independent

$P(X<2 \mid Y=4) = P(X<2) = \int_0^2 3e^{-3x}\,dx$

Question 6. Let A and B be two events in a sample space S. If $P(A) = 0.9$ and $P(B) = 0.8$, which of the following is true?



~~(a) $P(A \cap B) < P(A) + P(B) - 1$~~

~~(b) $P(A \cap B) \leq 0.17$~~

(c) $P(A \cap B) = P(A)P(B)$.

(d) $P(A \cap B) \geq 0.17$

~~(e) $P(A \cap B) = 0.17$~~

Question 7. Suppose the joint density function of two random variables, X, Y, is as follows:

$$f(x,y) = \frac{xy}{2}, \quad 0 \le x \le y \le 2.$$

Let $g(Y)$ be equal to the general expression of $\underline{E(X|Y)}$ as a function of Y. The expected value of g(Y) equals

(a) 8/5

(b) 16/15

(c) $E(Y \mid X)$

(d) 0.5414

(e) None of the above

$E(g(Y)) = \iint E(X|Y) \frac{xy}{2}\, dx\, dy = 16/15$

$f(y) = \frac{yx^2}{4}\Big|_{x=0}^{x=y} = \frac{y^3}{4} \quad 0 \le y < 2$

$f(x|y) = \frac{xy}{2} \div \frac{y^3}{4} = \frac{2x}{y^2}$

$E(X|Y) = \int_{x=0}^{y} x \cdot \left(\frac{2x}{y^2}\right) dx = \frac{2x^3}{3y^2}\Big|_{x=0}^{y} = \frac{2y}{3}$

Question 8. The roll of a fair six-sided die is equivalent to drawing a number at random from an urn containing numbers 1 to 6 with replacement. The roll of a fair six-sided die is also equivalent to drawing a random number from a probability mass function:

x= roll number	1	2	3	4	5	6
P(X=x)	1/6	1/6	1/6	1/6	1/6	1/6

The sum of the roll of two fair six-sided dice has the following distribution

(a) Normal with mean 7 and standard deviation 2.415229

~~(b) Uniform with mean 7 and standard deviation 2.415229~~

~~(c) Binomial(n=1, p=1/36)~~

~~(d) Symmetric but not normal~~

(e) None of the above.

$(2-7)^2(1/36) + (3-7)(2/36) + (4-7)^2(3/36) + (5-7)^2(4/36) + (6-7)^2(5/36) + (7-7)^2(6/36) + (8-7)^2(5/36) + (9-7)^2(4/36) + (10-7)^2(3/36) + (11-7)^2(2/36)(12-7)^2(1/36) = 5.83 \quad \sqrt{5.83} = 2.41523$

Question 9. Let $X_1, X_2, ..., X_n$ be a set of independent and identically distributed random variables. Which of the following is true?

$E(n\bar{X}) = E(S) \quad S = X_1 + X_2 + \dots + X_n$

(a) $E(n\bar{X}) = n\mu_X$

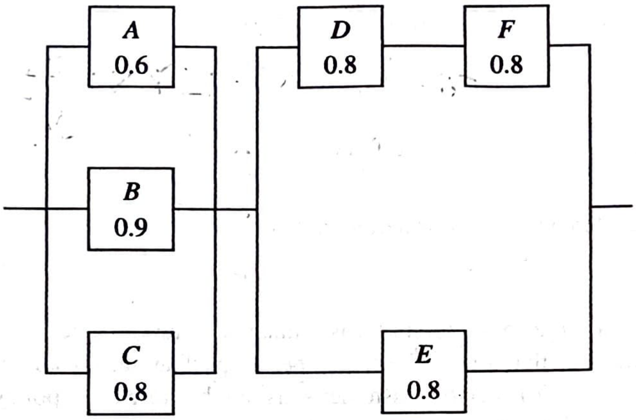
~~(b) $E(n\bar{X}) = (n)(n)\bar{X}$~~

~~(c) $E(n\bar{X}) = \mu_X$~~

~~(d) $E(n\bar{X}) = n\bar{X}$~~

(e) None of the above

Question 10. Consider the circuit shown below showing components A,B, C, D, E and the probability that each of the components work. The components function independently. Calculate the probability that at least one of the components work.



(a) 0.000064

(b) 0.920576

(c) 0.27648

(d) 0.999936

(e) None of the above

Question 11. The cumulative distribution of the largest of the two numbers that you get when two fair six-sided dice are rolled is given below

x= largest number	1	2	3	4	5	6
$P(X = x)$	1/36	4/36	9/36	16/36	25/36	36/36

Suppose that they will give us 1000 times the value we get for the largest number plus a constant amount of $10 for just playing the game. How much should we expect to make and how much would be the standard deviation of the amount we could make, respectively?

pmf

1	2	3	4	5	6
1/36	3/36	5/36	7/36	9/36	11/36

(a) 1404.066, 1482

(b) 4231, 1971400

(c) 4482, 1404.066

(d) 31567, 131.4

(e) None of the above

$E(X) = 1(1/36) + 2(3/36) + 3(5/36) + 4(7/36) + 5(9/36) + 6(11/36) = 4.472$

$E(Y) = 1000(E(X)) + 10 = 4482.222$

Question 12. The number of years a radio functions is exponentially distributed with parameter $\lambda = 1/8$. If Medha buys a used radio, a radio that has already been used for 10 years, what is the probability that this radio will be working after an additional 8 years?

$f(x) = \lambda e^{-\lambda x} = \frac{1}{8} e^{-\frac{1}{8}x}$

(a) 0.991

(b) 0.27189

$P(A|B) = \dfrac{1-\int_0^{18} \frac{1}{8} e^{-1/8x}\,dx}{1-\int_0^{10} \frac{1}{8} e^{-1/8x}\,dx} = .3678794$

(c) 0.6321206

(d) 0.3678794

A= works after 18 years
B= works after 10 years

(e) Can not be calculated with the information given

Question 13. 8% of a large lot of tomatoes is damaged. The tomatoes are sold to customers in packages of 4 randomly chosen tomatoes that are not inspected before packaging. Customers complain if at least one tomato is damaged in a package of 4. To keep the customer satisfied the store has a policy of replacing any damaged tomatoes in a package and giving the customer a coupon for future purchases. The cost of this program has, through time, been found to be $C = 0.5X^2$, where X is the number of tomatoes damaged in the package of 4. Find the expected cost of the program per package.

$X \sim Binom(n=4, p=.08)$

(a) 0.8016

$P(X=1) = .249$

(b) 0.0992

$P(X=2) = .0325$

(c) 0.1984

$P(X=3) = .001884$

(d) 0.048

$P(X=4) = .00004096$

(e) None of the above

$E(X^2) = 1(.249) + 4(.0325) + 9(.001884) + 16(.00004096)$
$= 0.3968 \quad E(C) = .5E(X^2) = .1984$

Question 14. When there are no epidemics, normal body temperature in degrees Celsius is 37 degrees with standard deviation 0.4. With COVID 19, a body temperature of 38 degrees Celsius will cause concern and suspicion of being infected by COVID-19. If everyone in a population known to not have been affected by the epidemic is measured their boy temperature, what percent of the population will be considered suspect of having been infected by COVID-19, approximately

(a) 6%

(b) 50%

(c) 0.6%

(d) 0%

(e) 30%

Question 15. You are asked to do prediction of the amount of repair time it would take to get a machine in the factory restarted and working as usual after breaking down. The prediction is needed to budget money for Enriqueta, the engineer employed by the company to do the restart. The factory has three types of machines: 20% of type A, 35% of type B, and 45% of type C. Based on past historical records, the factory knows that a machine's repair time after breaking down, X (in hours), varies by type of machine, and that time includes figuring if that is the broken machine, which only Enriqueta can figure out with engineering instruments made to detect where defects come from in a complex machine system. To do the prediction, you make use of your knowledge of the moment generating functions of X, for each type of machine:

$M_A(t) = (1 - 0.5t)^{-1}$
$M_B(t) = (1 - 0.75t)^{-1}$
$M_C(t) = (1 - 0.9t)^{-1}$

Based on this information, what would be the probability that it takes less than 1.5 hours to restart a broken machine?

(a) 0.1423188

(b) 0.352229

(c) 0.00114

(d) 0.8576812

(e) None of the above

Question 16. Suppose, in a certain region, the annual rainfall (in inches) is a normally distributed random variable with parameters $\mu = 40$ and $\sigma = 4$. Starting with this year, what is the probability that it will take over 5 years before a year occurs having a rainfall over 45 inches?

$X \sim Norm(\mu = 40, \sigma = 4)$

$P(X > 45) = .1066$

$(1 - .1066)^5$

(a) 0.10565

(b) 0.03229

(c) 0.57219

(d) 0.00819

(e) None of the above.

Question 17. We are interested in random variable X, the number of miles run per week by an individual training for the LA marathon. Consider n individuals, training for the LA marathon and let $X_1, X_2, ..., X_n$ denote their number of miles run per week. If these random variables are independent and identically distributed, which of the following statements is NOT true?

(a) The larger the n, the more the distribution of $\frac{\sum_{i=1}^{n} X_i}{n}$ resembles the distribution of any of the X_i.

~~(b) The expected value of~~ $\cancel{\frac{\sum_{i=1}^{n} X_i}{n}}$ ~~will be the same whether n is 100 or n is 5.~~

~~(c) The larger n is, the more the distribution of $\frac{\sum_{i=1}^{n} X_i}{n}$ will resemble a Gaussian distribution.~~

~~(d) An increase in n will decrease the standard deviation of $\frac{\sum_{i=1}^{n} X_i}{n}$~~

(e) None of the above

Question 18. The average annual income per capita in Madagascar is about \$400 per year. Find an upper bound for the percentage of income-earning individuals with incomes over \$1000.

(a) at least 60%

(b) at least 40%

(c) at most 40%

(d) at most 95%

(e) None of the above.

$P(X \geq a) \leq \frac{E(X)}{a}$

$P(X \geq 1000) \leq \frac{400}{1000}$

$P(X \geq 1000) \leq .40$

Question 19. During the COVID-19 pandemic, a lot of the shopping has been done online. People have used companies such as FEDEX and UPS to receive their merchandise. Suppose that the average delivery time for FEDEX is 150 hours and the standard deviation is 3 hours, and for UPS the average delivery time is 160 hours with standard deviation of 20 hours. If a person needs to have a package delivered in less than 115 hours, which service should be chosen, FEDEX or UPS? Assume that you are making decisions rationally, i.e., using probabilistic decision making.

(a) FEDEX

(b) UPS

(c) Either of them

P(FEDEX) = 0

P(UPS) = .0122

Question 20. The variance of random variable X is 10, the variance of random variable Y is 4, and the variance of the following random variable, $W = 5X + 3Y$, is 421. The correlation between variables X and Y must be approximately

(a) 0

(b) 4.5

(c) 0.2396

(d) 0.7115

(e) None of the above

$421 = Var(W) = Var(5X+3Y) = 5^2 Var(X) + 3^2 Var(Y) + 2(5 \cdot 3) Cov(X,Y)$

$421 = 25(10) + 9(4) + 30\, Cov(X,Y)$

$4.5 = Cov(X,Y)$

$Cov(X,Y) = \sigma_{XY}$

$\rho_{XY} = \frac{\sigma_{XY}}{\sigma_X \sigma_Y} = \frac{4.5}{10 \cdot 4} = 0.1125 = \rho_{XY}$

Question 21. Consider the joint density function

$$f(x, y) = 6x^2y, \quad 0 \le x \le 1, \quad 0 \le y \le 1$$

For this density function, $E(XY)$ (the expectation of the product of X and Y) equals

(a) E(X)E(Y)

(b) 0

(c) Cov(X, Y)

(d) Var(X+Y)

(e) All of the above

$f(x) = \int_0^1 6x^2y\,dy = 3x^2y^2 \Big|_{y=0}^{1} = 3x^2$

$f(y) = \int_0^1 6x^2y\,dx = 2x^3y \Big|_{x=0}^{1} = 2y$

$f(x)f(y) = 3x^2 \cdot 2y = 6x^2y = f(x,y)$

x and y are independent

Question 22. The cumulative distribution function of a random variable, Y, is

$$F(y) = \frac{1}{2} + \frac{1}{2}y + \beta\left(\frac{y^2-1}{4}\right), \quad -1 < y < 1,$$

where β is a parameter. The expected value of Y is

(a) $(1/2)\beta^2$

(b) $y + \beta$

(c) $\beta/3$

(d) 3β

(e) $4\beta + 1/2$

$\frac{dF(y)}{dy} = f(y) = \frac{1}{2} + \beta\left(\frac{2y}{4}\right) = \frac{1}{2} + \frac{\beta y}{2}$

$E(Y) = \int_{-1}^{1} y\left(\frac{1}{2} + \beta\frac{y}{2}\right) dy = \frac{y^2}{4} + \frac{\beta y^3}{6} \Big|_{-1}^{1}$

$= \frac{1}{4} + \frac{\beta}{6} - \left(\frac{1}{4} - \frac{\beta}{6}\right) = \frac{\beta}{3}$

Question 23. The following table was given to us to illustrate information available about HIV status of a population and results of a test that screens for HIV. The table contains joint probabilities. H in the table means the person has HIV and + is the event that a person tests positive for HIV.

		—Test Result—		
		Positive +	Negative –	
Disease Status	H	0.3	0.1	.4
	H^c	0.14	0.46	.6
		.44	.56	

Calculate the absolute difference between the two following quantities: (i) the probability that a randomly chosen person in this population has the disease and (ii) the probability that a person from this population that tested negative for HIV actually had the disease.

$P(\text{"(ii)"}) = \frac{P(H \cap \text{Neg.})}{P(\text{Neg})} = .17857$

(a) 0.6818

(b) 0.2214286

(c) 0.003

(d) 0.1785714

(e) 0.891

Question 24. In a city, 20% of the taxi cabs are yellow, 70 percent are blue and 10 percent are orange. We know that 30% of yellow taxicabs overcharge customers, 10% of the blue overcharge customers and 80% of the orange overcharge customers. A customer complained that in a late night ride, the taxicab overcharged, but the color of the taxicab was forgotten due to being kind of sleepy at that time of the night. Which cab color is more likely to have produced the overcharge?

(a) a blue cab

(b) a yellow cab

(c) an orange cab

(d) can not be determined

(e) they are all equally likely to have overcharged.

P(overcharge | yellow) = .06
P(overcharge | blue) = .07
P(overcharge | orange) = .08

Question 25. Let X be a discrete random variable with the following moment generating function.

$$M_X(t) = \frac{2e^{-t} + 3e^{2t} + e^{4t}}{6}$$

Compute $P(X \geq 2)$

(a) $\frac{1}{3}$

(b) $\frac{1}{2}$

(c) $\frac{2}{3}$

(d) 0.424

(e) 0.85

$M_X(t) = \frac{1}{3}e^{-t} + \frac{1}{2}e^{2t} + \frac{1}{6}e^{4t}$

we know that $M_X(t) = \sum_x e^{tx} p(x)$

we have a very similar form in the MGF. above, meaning $S = \{-1, 2, 4\}$ and the coefficients are the probabilities

x	-1	2	4
$P(X=x)$	1/3	1/2	1/6

GO TO NEXT PAGE FOR WORK QUESTIONS IN PART II

J. Sanchez
UCLA Department of Statistics

PART II. SHOW WORK FOR THE FOLLOWING QUESTIONS

For this part of the exam, you must show work to obtain full credit. The grading rubric will be as in the homework: 80 or 90% of the grade comes from work, including proper definition of your random variable(s), events, notation, assumptions and final answer. When the results are numeric, please, do not leave your final result as a fraction. Calculate the value of the fraction, providing at least three decimals (or more, if the first three decimals are 0).

MAKE SURE TO READ THIS: ⇒ For questions where the app for the Gaussian density function is used, you must provide either a screenshot or a drawing of everything appearing in the image written by hand. Image or drawing must appear in the space where you are answering the questions and text must appear with it indicating what you are calculating. The Z score must also be calculated in all problems involving the Gaussian density, and visible when reading your answer. We will not count as credit images or drawings that are provided in pages separate from the answer.

Question 26. (2 points) Consider two random variables W, T, which are linear functions of X and Y as follows:

$$W = aX + bY, \qquad T = cX + dY,$$

where X, Y have joint density $f(x, y)$.

We are interested in the covariance between W and T, so a student defined

$$Cov(W, T) = E[(W - \mu_W)(T - \mu_T)] = E[(aX + bY - E(aX + bY))(cX + dY) - E(cX + dY))],$$

and then the student wrote the continuation steps as follows, leaving all the steps incomplete. We leave blank space at each line for you to complete the derivations. Each line starts with an equal sign. Complete all the steps.

$$= E[(a(X - \mu_X) + b(Y - \mu_Y))(c(X - \mu_X) + d(Y - \mu_Y) \quad)]$$

distribute the above equation

$$= E[a(X - \mu_X)c(X - \mu_x) + b(Y - \mu_Y)(c(X - \mu_X) + a(X - \mu_X)d(Y - \mu_Y) + b(Y - \mu_Y)d(Y - \mu_Y) \quad)]$$

$$= E[a(X - \mu_X)c(X - \mu_x)] + E[b(Y - \mu_Y)c(X - \mu_X)] + E[a(X - \mu_X)d(Y - \mu_Y)] + E[b(Y - \mu_Y)d(Y - \mu_Y)]$$

$$= \quad ac\ Var(X) + bc\ Cov(X,Y) + ad\ Cov(X,Y) + \quad bd\ Var(Y)$$

Question 27. The contest "The will of chance" is a daily TV contest that consists of asking 10 questions to the contestant of the day. In each contest, the contestant is asked 10 questions drawn at random from a randomly chosen test bank of the many test banks secretly kept by the contest organizers. These test banks are such that 30% of them contain 15 questions, 5 of which have already been asked in past contests and 70% of them contain 20 questions, 7 of which have been asked in past contests. A new contestant, who has been a follower of the contest for many years, is going to be participating in tonight's contest. Calculate the probability that the contestant will be asked 3 questions that were asked in past contests.

Let X = the # of questions asked to the contestant that appeared in past contests from the type of test bank that has 15 questions

X is a hypergeometric random variable because there are a fixed number of trials (n = 10), each trial has 2 outcomes: question from past contest or not from past contest, and the trials are not independent and p changes because the size of our population is very small (15 questions)

$M = 15 \quad n = 10$
$M_s = 5 \quad k = 3$

$$P(X=k) = \frac{\binom{M_s}{k}\binom{M-M_s}{n-k}}{\binom{M}{n}} \Rightarrow P(X=3) = \frac{\binom{5}{3}\binom{10}{7}}{\binom{15}{10}} = 0.3996004$$

Let Y = the # of questions asked to the contestant that appeared in past contests from the type of test bank that has 20 questions

Y is a hypergeometric random variable because there are a fixed number of trials (n = 10), each trial has 2 outcomes: question from past contest or not from past contest, and the trials are not independent and p changes because the size of our population is very small (20 questions)

$M = 20 \quad n = 10$
$M_s = 7 \quad k = 3$

$$P(X=k) = \frac{\binom{M_s}{k}\binom{M-M_s}{n-k}}{\binom{M}{n}} \Rightarrow P(X=3) = \frac{\binom{7}{3}\binom{13}{7}}{\binom{20}{10}} = 0.3250774$$

to find the overall probability, we must apply weights to the earlier probabilities to show the higher frequency of certain types of test banks that are more likely to be drawn

$$(0.3)(0.3996004) + (0.7)(0.3250774) = \boxed{0.3474343}$$

Question 28. Let X and Y denote random variables denoting the proportion of contaminants of type A and the proportion of contaminants of type B present in a landfill, respectively. A student calculated the conditional density function

$$f(y|x) = \frac{2(x+y)}{3x^2}, \quad 0 \le y \le x,$$

and the marginal density function

$$f(x) = 3x^2, \quad 0 \le x \le 1$$

(a) Calculate the join density of X and Y, and indicate the domain.

$$f(y|x) = \frac{f(x,y)}{f(x)}$$

$$f(y|x) \cdot f(x) = f(x,y)$$

$$f(x,y) = \frac{2(x+y)}{3x^2} \cdot 3x^2$$

$$\boxed{f(x,y) = 2(x+y) \quad 0 \le y \le x \le 1}$$

I got the domain $0 \le y \le x \le 1$ because I knew x had to be ≤ 1 from f(x), but it had to be $\ge y$ from f(y|x), so that domain satisfies both conditions.

(b) Calculate the Var(3X+Y).

$$Var(a_1X_1 + a_2X_2) = a_1^2 Var(X_1) + a_2^2 Var(X_2) + 2(a_1a_2)Cov(X_1, X_2)$$

$$Var(a_1X_1 + a_2X_2) = \int_{X_2}\int_{X_1} (a_1x_1 + a_2x_2 - E(a_1x_1 + a_2x_2))^2 f(x_1, x_2)\, dx_1 dx_2$$

$$Var(3X+Y) = \int_0^1 \int_0^x (3X+Y - E(3X+Y))^2 (2x+2y)\, dy\, dx$$

$$Var(3X+Y) = \int_0^1 \int_0^x (3X+Y-(3E[X]+E[Y]))^2 (2x+2y)\, dy\, dx$$

$$E[X] = \int_0^1 x \cdot 3x^2 dx$$

$$E[X] = 0.75$$

$$f(y) = \int_y^1 2x+2y\, dx = x^2 + 2xy \Big|_{x=y}^{1}$$

$$1 + 2y - (y^2 + 2y^2)$$

$$f(y) = -3y^2 + 2y + 1$$

$$0 \le y \le 1$$

$$E[Y] = \int_0^1 y(-3y^2 + 2y + 1)\, dy$$

$$E[Y] = 0.4166667$$

$$Var(3X+Y) = \int_0^1 \int_0^x (3X+Y-(2.25+.41667))^2 (2x+2y)\, dy\, dx$$

*continued on next page *

(c) Would you invoke the Central Limit theorem to calculate $P(3X + Y > 0.8)$? Why or why not?

I would not invoke the Central Limit theorem (CLT) to calculate $P(3X+Y > 0.8)$ because this would be an incorrect application of the CLT. The CLT says that when many independent random variables are added, their normalized sum tends towards a normal distribution. X and Y are dependent as $f(x)f(y) \neq f(x,y)$ and thus the CLT cannot be used for this reason. Additionally, it is recommended that the number of independent variables summed for the CLT is ≥ 30 for the sum to tend towards a normal distribution and in this case, we are only summing 2 variables, which is far below that requirement.

J. Sanchez
UCLA Department of Statistics

for questions involving the applet, I have included a larger screenshot on the following page so it is more easily readible

Question 29. A biologist in a field trip at location A examines frogs for a genetic trait. Previous research suggests that this trait is found in 1 out of every 5 frogs. It is also known that the average weight of a frog is approximately 0.8 ounces with standard deviation 0.15 ounces (although it is possible that the weigh depends on the sex of the frog and there is a different average for each sex, so overall weight for the whole frog population is not normal).

(a) The biologist plans to examine 150 randomly chosen frogs at location A. Calculate the probability that the biologist will find more than 16 frogs with the trait. Does this evidence suggest that perhaps the "1 out 5 frogs have the trait" assumption is not correct? Explain.

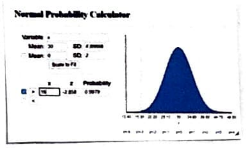
Let X = the number of frogs with the genetic trait

$n = 150 \quad p = 1/5 = 0.2 \quad 1-p = 0.8$

$np = (150)(0.2) = 30 \geq 10$

$n(1-p) = (150)(0.8) = 120 \geq 10$

therefore, can approximate using the normal distribution



$\mu_X = np = (150)(0.2) = 30$

$\sigma^2 = np(1-p) = (150)(.2)(.8) = 24$

$\sigma = \sqrt{\sigma^2} = \sqrt{24} = 4.89898$

⇒ use the app

$z = \frac{16-30}{\sqrt{24}}$

$z = -2.8577$

$\boxed{P(Z > -2.8577) = 0.9979}$

The evidence does not suggest anything against the assumption, but it could be that more than 1 out of 5 frogs have the trait.

(b) Calculate the probability that the average weight of the 150 frogs examined is larger than 0.9 ounces.

Y = the weight of a given frog

$\bar{Y} = \frac{\sum_{i=1}^{150} Y_i}{150}$ = the average weight of 150 frogs

Because of the large sample size, we can use the CLT to say that the $\bar{Y} \sim Normal(\mu_{\bar{Y}} = 0.8, \sigma_{\bar{Y}} = 0.012247)$

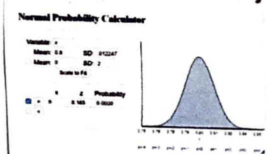
$E(\bar{Y}) = E(Y) = \mu_Y = 0.8$ oz.

$SD(\bar{Y}) = \frac{\sigma_Y}{\sqrt{n}} = \frac{0.15}{\sqrt{150}} = 0.012247$

⇒ use the app

$z = \frac{0.9-0.8}{0.012247}$

$z = 8.165 \Rightarrow \boxed{P(Z > 8.165) \approx 0}$



(c) Next week, the biologist plans to do research at locations B, C, D, to examine the frogs in that location for the genetic trait. Past research suggests that in all those locations 1 out of 5 frogs have the genetic trait. The geneticists plans to select the frogs at random, selecting 5 frogs in location B, 3 in C and 4 in D. Calculate the probability that the biologist will find more than 3 frogs in total with the trait in the three locations combined. You must explain (showing work) why you choose the probability model that you use to do this calculation.

$5+3+4 = 12$ frogs selected

X = the number of frogs in total that have the genetic trait

p = probability of having the trait $= 1/5 = 0.2$

$1-p = 0.8$

X is a binomial distribution because there are a fixed number of trials, each trial has two outcomes: frog with trait or without trait, and the trials are independent and p stays the same from trial to trial because our population is very large.

$P(X > 3) = 1 - P(X \leq 3)$

$= 1 - [P(X=0) + P(X=1) + P(X=2) + P(X=3)]$

$= 1 - [\binom{12}{0}(0.2)^0(0.8)^{12} + \binom{12}{1}(0.2)^1(0.8)^{11} + \binom{12}{2}(0.2)^2(0.8)^{10} + \binom{12}{3}(0.2)^3(0.8)^9]$

$= 1 - [.0687194 + .2061584 + .2834678 + .236223]$

$\boxed{= 0.205431}$

I chose the binomial distribution because it most accurately represented our problem, and the probability calculation with the binomial distribution was fairly straightforward here.

Nicholas Dean, 705-312-823

Stats 100A Final Exam

28.) b.) (continued)

$$Var(3X+Y) = \int_0^1 \int_0^x \left(3x+y-\frac{8}{3}\right)^2 (2x+2y)\,dy\,dx$$

$$= \int_0^1 \left(33x^4 - 32x^3 + \frac{x^4}{2} + \frac{14x^4}{3} - \frac{32x^3}{9} - \frac{64x^3}{3} + \frac{64x^2}{9} + \frac{128x^2}{9}\right) dx$$

$$= \left(\frac{33x^5}{5} - 8x^4 + \frac{x^5}{10} + \frac{14x^5}{15} - \frac{8x^4}{9} - \frac{16x^4}{3} + \frac{64x^3}{27} + \frac{128x^3}{27}\right\rvert_{x=0}^{1}$$

$$= \frac{33}{5} - 8 + \frac{1}{10} + \frac{14}{15} - \frac{8}{9} - \frac{16}{3} + \frac{64}{27} + \frac{128}{27}$$

$$= -8 + \frac{67}{10} - \frac{8}{9} - \frac{66}{15} + \frac{192}{27}$$

$$= -\frac{720}{90} - \frac{80}{90} + \frac{603}{90} - \frac{396}{90} + \frac{192}{27}$$

$$= \frac{-593}{90} + \frac{192}{27}$$

$$= \frac{-593}{90} + \frac{640}{90}$$

$$= \frac{47}{90}$$

$$\boxed{Var(3X+Y) = \frac{47}{90} = 0.522222}$$

Normal Probability Calculator

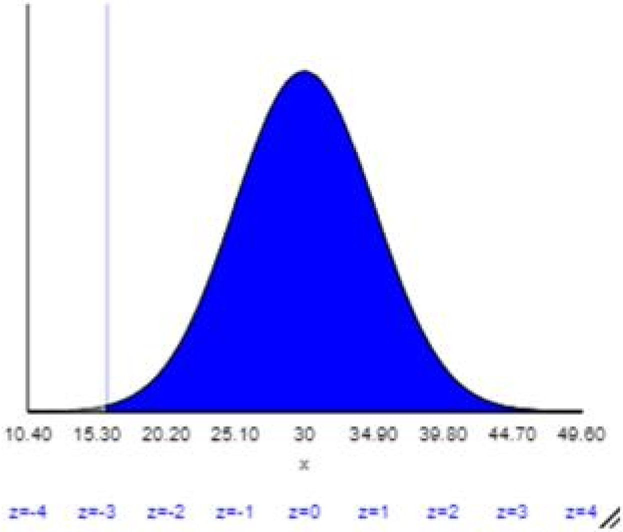
Variable: x

Mean: 30 SD: 4.89898

Mean: 0 SD: 2

Scale to Fit

	x	z	Probability
>	16	-2.858	0.9979
<			



Normal Probability Calculator

Variable: x

Mean: 0.8 SD: .012247

Mean: 0 SD: 2

Scale to Fit

	x	z	Probability
>	.9	8.165	0.0000
<