Name: ______________________________

Rensselaer Polytechnic Institute
Department of Electrical, Computer, and Systems Engineering
ECSE-2500: Engineering Probability, Spring 2023

Exam 3. Closed book, one two-sided page of notes.
May 4, 2023, 12:00-1:30 PM

Show all work for full credit.

- A calculator is allowed but you can only use (and will only need) non-programmed, simple functions (multiplications, exponentials, logarithms, factorials).
- Reduce expressions involving factorials as much as possible (but do not attempt to evaluate expressions involving huge numbers).
- Evaluate combinatorials when they are small and easy to compute.
- Use the provided Q and significance tables on the last page to estimate integrals involving the Gaussian distribution.
- When in doubt, show more work!
- Have a great break!

1		**20**
2		**18**
3		**12**
4		**15**
5		**20**
6		**15**
Total		**100**

1. (20 points.) A professor feeds his probability class candies at the final exam, which have the following calories:

Candy	Calories	Candy	Calories
Hershey's Milk	42	Special Dark	39
Mr. Goodbar	50	Krackel	46
Kit Kat	42	Reese's Cup	45

We assume that a student is equally likely to choose any of the 6 possible candies.

(a) (4 points.) Let X be the number of calories in a student-chosen candy. Compute $E(X)$.

$$\frac{1}{6}(42+50+42+39+46+45) = \frac{1}{6}(264) = \boxed{44}$$

(b) (4 points.) Compute $\text{Var}(X)$.

$$\sigma^2 = \frac{1}{6}(2^2+6^2+2^2+5^2+2^2+1^2)$$
$$= \frac{37}{3} = 12.33$$

(c) (4 points.) Now suppose the student eats 56 randomly chosen candies during the final exam. Use the central limit theorem to estimate the probability that he exceeds his daily recommended allowance of 2500 calories in this candy binge.

$$\mu = 44, \quad \sigma = \sqrt{12.33} = 3.512$$

$$P(S_{56} > 2500) = P\left(\frac{S_{56} - 56\cdot 44}{\sqrt{56}\cdot\sqrt{12.33}} > \frac{2500 - \overset{2464}{\overbrace{56\cdot 44}}}{\sqrt{2072/3}}\right)$$

$$= P(Z > 1.37) \approx Q(1.37)$$
$$\approx \boxed{0.08}$$

(d) (4 points.) Suppose instead that the number of candies N the student eats is a geometric random variable with $p = \frac{1}{3}$. Using the law of iterated expectation, compute $E\left(\sum_{i=1}^{N} X_i\right)$, where X_i is the number of calories in the i^{th} candy.

$$E\left(\sum_{i=1}^{N} X_i\right) = E\left(E\left(\sum_{i=1}^{N} X_i \mid N\right)\right)$$

$$= E(N E(X_i))$$

$$= E(N)E(X_i)$$

$$= \underbrace{\frac{1}{\frac{1}{3}}}_{\left(N \text{ GEOMETRIC } p=\frac{1}{3}\right)} \cdot \underbrace{44}_{\text{PART (a)}} = \boxed{132}$$

(e) (4 points.) Finally, suppose instead that the professor offers only one type of candy: the probBar, whose caloric content is a random variable with unknown mean μ calories and standard deviation **5** calories.[1] We collect 100 probBars and find that their sample mean is 62 calories. Compute the 95% confidence interval for the mean μ based on this sample.

$$P(M_n - c < \mu < M_n + c) = 0.95$$

$$\text{so} \quad P(M_n - \mu > c) = 0.025$$

$$P\left(\frac{100(M_{100} - \mu)}{\sqrt{100} \cdot 5} > \frac{100c}{\sqrt{100} \cdot 5}\right) = 0.025$$

$$\downarrow \qquad\qquad \downarrow$$

$$Z \qquad\qquad z_{0.025} = 1.96 \quad \text{FROM TABLE}$$

$$\text{So} \quad 1.96 = 2c \quad \Rightarrow \quad c = 0.98$$

CONFIDENCE INTERVAL IS $62 \pm 0.98 = [61.02, 62.98]$

[1] Like this exam, the probBar is hard to swallow, but ultimately good for you.

2. (18 points.) A bag contains 16 tiles. The tiles are labeled D, O, U, B, L, E, S, and ! and there are 2 indistinguishable copies of each tile. A monkey pulls 4 tiles out of the bag at random, one at a time, without replacement.

D	O	U	B	L	E	S	!
D	O	U	B	L	E	S	!

(a) (6 points.) What is the probability that none of the drawn tiles are vowels (i.e., not A, E, I, O, or U)?

$$\left(\frac{\cancel{10}}{\cancel{16}}\right)\left(\frac{\cancel{9}}{\cancel{15}}\right)\left(\frac{\cancel{8}}{\cancel{14}}\right)\left(\frac{\cancel{7}}{13}\right) = \frac{3}{26} = 0.115$$

$$\left(\text{ALSO} = \frac{\binom{10}{4}}{\binom{16}{4}}\right)$$

(b) (6 points.) What is the probability that there are no doubles in the drawn tiles — that is, no two tiles in the draw have the same symbol?

$$\frac{\cancel{14}}{\cancel{15}} \cdot \frac{\cancel{12}}{\cancel{14}} \cdot \frac{\cancel{10}}{13} = \frac{8}{13} = 0.615$$

$$\text{ALSO} = \left(\binom{8}{4} \cdot 2^4\right) \Big/ \binom{16}{4}$$

(c) (6 points.) How many unique draws are there? That is, how many distinguishable sets of 4 tiles are there when order doesn't matter?

NO DOUBLES: $\binom{8}{4} = 70$

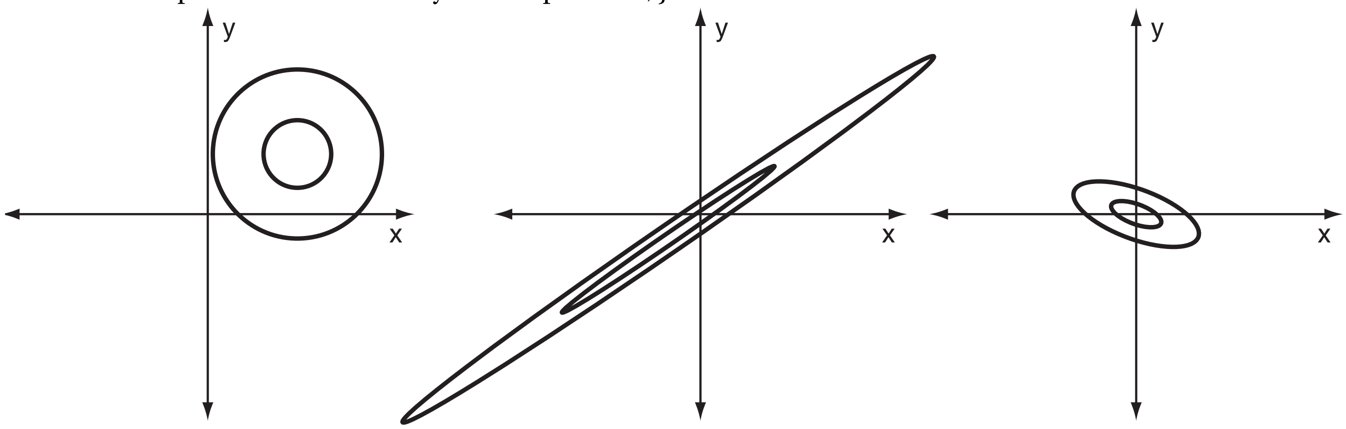
ONE DOUBLE: $\binom{8}{3} \cdot 3 = 168 = 8 \cdot \binom{7}{2}$

CHOOSE 3, DOUBLE 1 — CHOOSE 1 TO DOUBLE, CHOOSE 2 MORE

TWO DOUBLES: $\binom{8}{2} = 28$

TOTAL = $\boxed{266}$

3. (12 points.) The below figures illustrate the joint PDFs $f_{XY}(x,y)$ for several two-dimensional Gaussian distributions. The ellipses are isocontours (i.e., every point has the same PDF value) centered at the mean of the distribution. The inner ellipse indicates the region containing 25% of the probability and the outer ellipse indicates the region containing 75% of the probability. All the axes are drawn at the same scale. No work/explanation is necessary for this problem, just an answer of the form `A B C`.



Distribution A Distribution B Distribution C

(a) (3 points.) Sort the distributions from the lowest to highest value of μ_X.

B C A

$\mu_X < 0$ $\mu_X = 0$ $\mu_X > 0$

(b) (3 points.) Sort the distributions from the lowest to highest value of σ_Y.

C A B

SMALL σ_Y MEDIUM σ_Y BIG σ_Y

(c) (3 points.) Sort the distributions from the lowest to highest value of the correlation coefficient ρ.

C A B

$\rho < 0$ $\rho = 0$ $\rho > 0$

(d) (3 points.) Sort the distributions from the lowest to highest value of the *absolute value* of the correlation coefficient ρ.

A C B

$\rho = 0$ $|\rho|$ SMALL $|\rho|$ LARGE

4. (20 points) Consider the function

$$g(x,y) = \begin{cases} cx(1-x)y & x \in [0,1], y \in [0,1] \\ 0 & \text{otherwise} \end{cases}$$

(a) (4 points) Determine the value of c so that $g(x,y)$ is a valid joint PDF $f_{XY}(x,y)$ for two random variables X and Y.

Let's integrate this function and see what we get:

$$\begin{aligned} \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x,y)\,dx\,dy &= k\int_0^1\int_0^1 x(1-x)y\,dx\,dy \\ &= k\left(\frac{1}{2}x^2 - \frac{1}{3}x^3\Big]_{x=0}^{x=1}\right)\left(\frac{1}{2}y^2\Big]_{y=0}^{y=1}\right) \\ &= k(\frac{1}{6})(\frac{1}{2}) \\ &= \frac{k}{12} \end{aligned}$$

Since for a valid PDF we need this integral to equal 1, this means that $k = 12$.

(b) (4 points) Compute the joint CDF $F_{XY}(x,y)$ corresponding to the PDF in part (a). You only need to compute the CDF for values (x,y) such that $x \in [0,1], y \in [0,1]$.

The joint CDF in the "interesting" range $x \in [0,1], y \in [0,1]$ is computed as

$$\begin{aligned} F_{XY}(x,y)\,dx\,dy &= \int_0^x\int_0^y f_{XY}(x,y)\,dx\,dy \\ &= 12\left(\frac{1}{2}x^2 - \frac{1}{3}x^3\right)\left(\frac{1}{2}y^2\right) \\ &= (3x^2 - 2x^3)y^2 \end{aligned}$$

(c) (4 points) Compute the marginal PDF $f_X(x)$. Remember to specify the PDF at all values of x.

To compute the marginals, we integrate out the variable we don't care about. For $x \in [0,1]$ we have

$$\begin{aligned} f_X(x) &= \int_0^1 12x(1-x)y\,dy \\ &= (12x(1-x))\left(\frac{1}{2}y^2\Big]_{y=0}^{y=1}\right) \\ &= (12x(1-x))(\frac{1}{2}) \\ &= 6x(1-x) \quad x \in [0,1], 0 \text{ otherwise} \end{aligned}$$

(d) (4 points) Are X and Y independent? Justify your answer.

Yes, X and Y are independent since we can see that

$$\begin{aligned} f_{XY}(x,y) &= 12x(1-x)y \\ &= (6x(1-x))(2y) \\ &= f_X(x)\,f_Y(y) \end{aligned}$$

(e) (4 points) Compute $P(Y \le \sqrt{X})$.

$$\begin{aligned} P(Y < \sqrt{X}) &= \int_0^1\int_0^{\sqrt{x}} 12x(1-x)y\,dy\,dx \\ &= \int_0^1 (12x(1-x))\left(\frac{1}{2}y^2\Big]_{y=0}^{y=\sqrt{x}}\right)dx \\ &= \int_0^1 (12x(1-x))\left(\frac{1}{2}x\right)dx \\ &= \int_0^1 6x^2(1-x)\,dx \\ &= 2x^3 - \frac{6}{4}x^4\Big]_{x=0}^{x=1} \\ &= \frac{1}{2} \end{aligned}$$

More fun on the next page ⟶

5. (15 points) Rent The Runway (RTR) offers a service for customers to rent clothes through the mail. We call each package of clothes a Shipment. Suppose RTR's internal shipping and handling cost to process the exchange of Shipment i is X_i, a uniform random variable on [0.30, 1.20] dollars. Let Y_i be the (continuous) number of days a customer holds on to Shipment i, modeled as an exponential random variable with mean 5 days. For the purposes of this problem, assume the exchange of shipments is instantaneous, and that all the $\{X_i\}$ and $\{Y_i\}$ are mutually independent. Let N be the number of shipments a customer rents over the course of 30 days, and let Z be the internal monthly cost to RTR for shipping and handling these N shipments. Note that N is a discrete random variable (we can't ship half a package)!

   (a) (5 points) Determine the conditional expected value $E(Z \mid N)$.

$$\begin{aligned} E(Z \mid N) &= E(X_1 + X_2 + \cdots + X_N) \\ &= E(X_1) + E(X_2) + \cdots + E(X_N) \\ &= NE(X_i) \\ &= N \cdot \frac{1}{2}(0.3 + 1.2) \\ &= 0.75N \end{aligned}$$

   The second step follows from the fact that the X_i's are all independent and identically distributed, and the third step follows from the information that this distribution is uniform on $[0.3, 1.2]$ dollars.

   (b) (5 points) Determine the expected value of Z. Remember the connection between the exponential and Poisson random variables!

   By the law of iterated expectations,

$$E(Z) = E(E(Z \mid N)) = E(0.75N) = 0.75E(N)$$

   What is $E(N)$? It is the expected number of orders in a span of 30 days, where the interval between arrivals is exponentially distributed with mean 5 days (i.e., $\lambda = \frac{1}{5}$). Thus the expected number of arrivals per day is $\frac{1}{5}$ and we expect $30 \cdot \frac{1}{5} = 6$ arrivals in 30 days. This means that N is a Poisson random variable with $\alpha = 6$, which means that $E(N) = \alpha = 6$. Then

$$E(Z) = \$0.75 \cdot 6 = \$4.50$$

   (c) (5 points) How much should RTR charge for its monthly service to make an average of 5 dollars net profit per month from each customer?

   Since it costs RTR \$4.50 on average for the raw monthly shipping and handling, they should charge \$9.50 per month to make a \$5 profit.

6. (15 points.) Let X be a random variable representing the length of an episode of Andor, in minutes. Initially we don't know anything about the PDF of X, other than that it is symmetric about its mean.

   (a) (5 points.) Suppose we learn that $E(X) = 50$. Compute the Markov bound on $P(X \geq 55)$.

   (a) (5 points.) The Markov bound says that $P(X \geq 55) \leq \frac{50}{55} = 0.91$.

   (b) (5 points.) The Chebyshev bound says that $P(|X-50| \geq 5) \leq \left(\frac{4}{5}\right)^2 = 0.64$, which, assuming the distribution is symmetric, would bound $P(X \geq 55) \leq 0.32$ (a lot less than the previous estimate).

   (c) (5 points.) If the random variable is in fact Gaussian, we can compute from the Q table that $P(X \geq 55) = Q(\frac{5}{4}) = Q(1.25) = 0.106$. So while both bounds are satisfied, neither is very tight. (The bounds work better when we ask for values really far from the mean.)

   (b) (5 points.) We now further learn that $\text{Var}(X) = 16$. Compute the Chebyshev bound on $P(|X-50| \geq 5)$, and then divide this by two to get an estimate of $P(X \geq 55)$ since we know the distribution is symmetric.

   (c) (5 points.) Finally, we learn the complete distribution of X: it's actually a Gaussian with the mean and variance given above. Compute the actual value of $P(X \geq 55)$. How does it compare to the bounds above?

Tables and page for extra work

z	Q(z)	z	Q(z)
0.0	5.000e-01	3.0	1.350e-03
0.1	4.602e-01	3.1	9.677e-04
0.2	4.207e-01	3.2	6.872e-04
0.3	3.821e-01	3.3	4.835e-04
0.4	3.446e-01	3.4	3.370e-04
0.5	3.085e-01	3.5	2.327e-04
0.6	2.743e-01	3.6	1.591e-04
0.7	2.420e-01	3.7	1.078e-04
0.8	2.119e-01	3.8	7.237e-05
0.9	1.841e-01	3.9	4.812e-05
1.0	1.587e-01	4.0	3.17e-05
1.1	1.357e-01	4.5	3.40e-06
1.2	1.151e-01	5.0	2.87e-07
1.3	9.680e-02	5.5	1.90e-08
1.4	8.076e-02	6.0	9.87e-10
1.5	6.681e-02	6.5	4.02e-11
1.6	5.480e-02	7.0	1.28e-12
1.7	4.457e-02	7.5	3.19e-14
1.8	3.593e-02	8.0	6.22e-16
1.9	2.872e-02	8.5	9.48e-18
2.0	2.275e-02	9.0	1.13e-19
2.1	1.786e-02	9.5	1.05e-21
2.2	1.390e-02	10.0	7.62e-24
2.3	1.072e-02		
2.4	8.198e-03		
2.5	6.210e-03		
2.6	4.661e-03		
2.7	3.467e-03		
2.8	2.555e-03		
2.9	1.866e-03		

Critical values for standard Gaussian random variable

α	z_α
0.1000	1.2816
0.0500	1.6449
0.0250	1.9600
0.0100	2.3263
0.0050	2.5758
0.0025	2.8070
0.0010	3.0903
0.0005	3.2906
0.0001	3.7191

Have a great break!