E1: Midterm Exam

CEE 260/MIE 273: Probability & Statistics in Civil Engineering

October 21, 2021

time limit: Twenty-Four Hours

Name

Please print your name clearly in the box below.

Turn to the next page to read the instructions.

Instructions

This exam contains **22 pages** (including the front and back pages) and **11 problems, 106 points** (with 8 points extra credit). You have **24 hours** to complete it. You may print out the PDF, complete it and upload as a PDF on Moodle, or *neatly* answer the questions on blank pages of paper, scan and upload.

This is an **open resource examination**. You are expected to complete the exam individually. Asking anyone (colleague, friend, tutor, etc) questions about the exam is *not allowed*. If any questions arise during the exam, direct them to me (via email).

The following rules apply:

- **Organize your work**, in a reasonably neat and coherent way.Work scattered all over the page without a clear ordering will receive very little credit.
- **Show ALL your work where appropriate**. The work you show will be evaluated as well as your final answer. Thus, provide ample justification for each step you take. Indicate when you have used a probability table or MATLAB to obtain a result. In the case of MATLAB, briefly include the function or statement you used to arrive at your result. In the long response questions, simply putting down an answer without showing your steps will not merit full credit. **EXCEPTION:** For short response or "True/False" questions, *no explanations are required*. However, the more work you show, the greater your chance of receiving partial credit if your final answer is incorrect.
- If you need more space, use the blank pages at the end, and clearly indicate when and where you have done this.
- Questions are roughly in order of the lectures, so later questions may not necessarily be harder. If you are stuck on a problem, it may be better to skip it and get to it later.
- Manage your time wisely.

Problem 1 *True/False questions (16 points)*

Respond "T" (*True*) or "F" (*False*) to the following statements. Use the boxes provided. Each response is worth 1 point. Note that a statement can only be regarded as true in this framework if it always holds in all circumstances. If a statement does not hold under a given condition not already explicitly excluded, then it should be regarded as false.

(i) F — The probability of the union of collectively exhaustive events is 0. This is equal to 1.

(ii) F — If two events A and B are mutually exclusive, then $P(AB) = P(A) + P(B)$. The union, not the intersection, is the sum.

(iii) F — For a given online account, your password must have seven unique small letters (this means you cannot repeat a letter in your password). Assuming you are the first user to create an account, the number of possibilities for your password is 26^7. Since the letters cannot repeat, the options reduce by 1 for each subsequent position. Thus, $n = 26 \times 25 \times \cdots 21 \times 20 = \frac{26!}{(26-7)!} = \frac{26!}{19!}$.

(iv) T — The area under the curve of a PDF is equal to 1.

(v) F — You begin your turn in a board game by rolling a die. The number you roll is a continuous random variable. The outcomes are countably finite; hence discrete.

(vi) T — In a left-skewed distribution, the mean always lies to the left of the median.

(vii) T — $P(AB) = 1 - (P(\overline{A}) + P(\overline{B}))$ for any two events A and B. According to De Morgan's Rule, $P(\overline{A}) + P(\overline{B}) = P(\overline{A}) \cup P(\overline{B}) = P(\overline{AB})$. Therefore, the RHS is $1 - P(\overline{AB}) = P(AB)$.

(viii) F — If A and B are statistically independent events, then $P(B|A)$ is always equal to $P(A|B)$. S.I. implies that $P(B|A) = P(B)$ and $P(A|B) = P(A)$. So, this statement only holds if $P(A) = P(B)$.

(ix) T The union of collectively exhaustive events yields the entirety of the corresponding sample space.

(x) F The minimum value of any cumulative distribution function can be either 0 or -1. The lowest value of the CDF is always 0.

(xi) F Under certain conditions, a binomial distribution with parameters (n, p) can be approximated by a normal distribution with $\mu = n(1-p)$. $\mu = np$.

(xii) F The probability of the number of heads that occur in 50 tosses of a coin (which can yield either a head or a tail) cannot be appropriately modeled by a binomial distribution. Coin toss outcomes are Bernoulli sequences; and hence can be modeled by binomial distribution.

(xiii) T Given a normal distribution with parameters μ and σ, the variance of the distribution is σ^2.

(xiv) F The standard normal variate Z has a mean of 1. The mean is 0, not 1

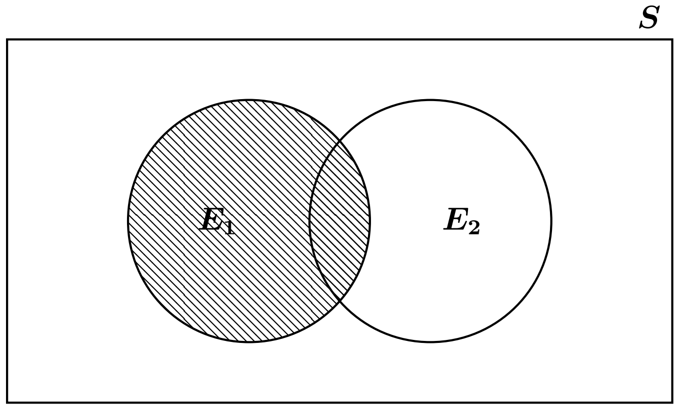
(xv) F The probability of the elapsed time between events that constitute a Poisson process can be modeled using a Poisson distribution. Elapsed time in Poisson process is modeled by exponential the distribution.

(xvi) T If a variable X is lognormally distributed with parameters μ and σ, then the median of X is given by $\exp(\mu)$.

Problem 2 *Venn diagrams (7 points)*

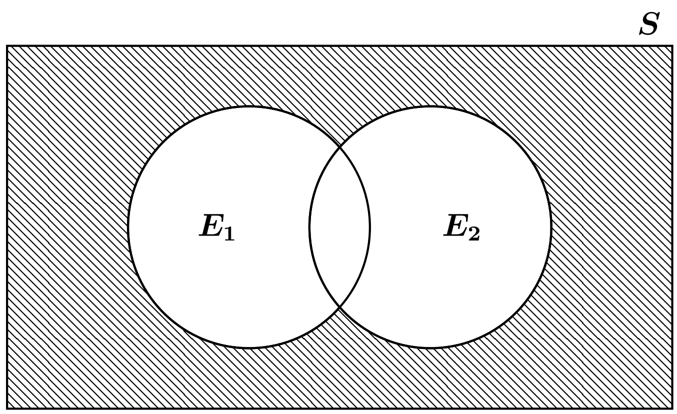
Write the combination of events (using set notation) depicted in each of the figures below.

(i) *[1]*



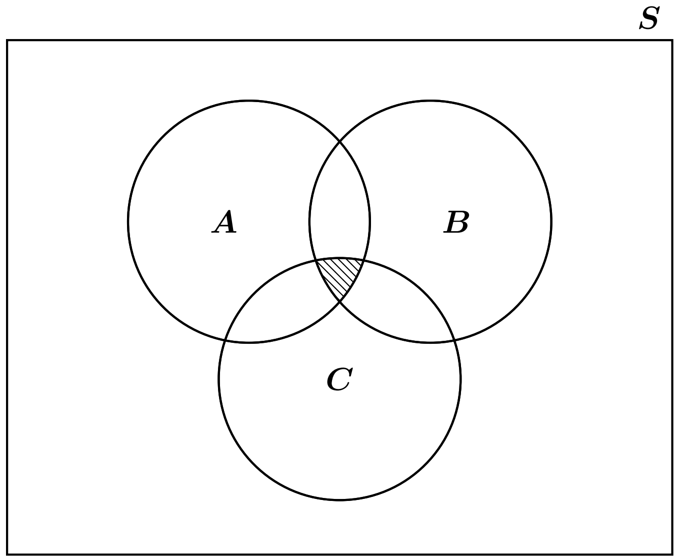
Answer: E_1

(ii) *[2]*



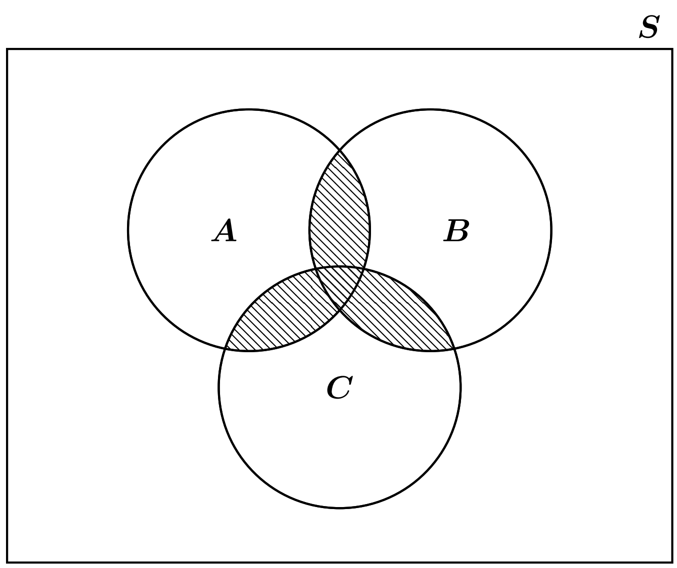
Answer: $\overline{E_1 \cup E_2}$ OR $\overline{E_1} \cap \overline{E_2}$

[2] **(iii)**



Answer: $\boldsymbol{ABC}$ OR $\boldsymbol{A \cap B \cap C}$

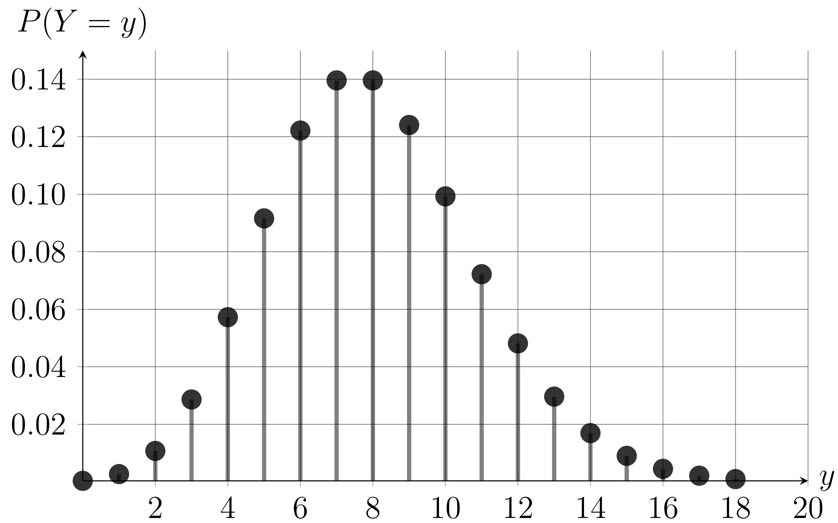
[2] **(iv)**



Answer: $\boldsymbol{AB \cup AC \cup BC}$ OR $\boldsymbol{A(B \cup C) \cup BC}$

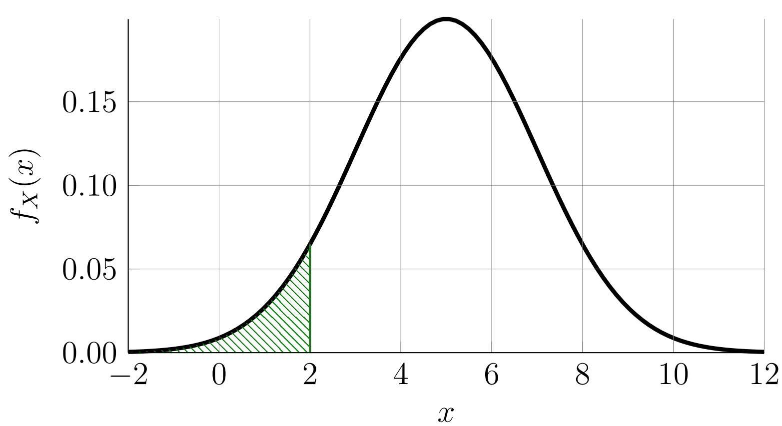
Problem 3 *Short answer questions (16 points)*

(i) The PMF of a random variable Y is given in the figure below. Use the figure to estimate the probability $P(7 \leq Y < 9)$. *[2]*

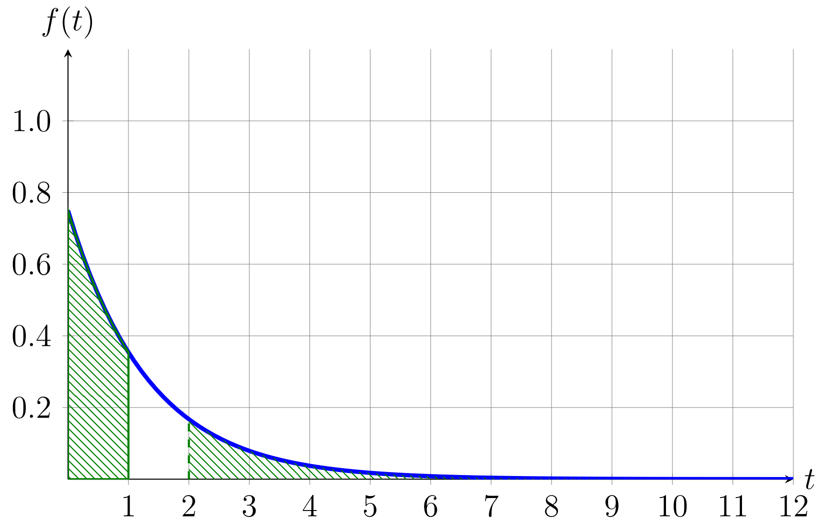


Answer: $P((7 \leq Y < 9) = P(7) + P(8) \approx 0.14 + 0.14 = 0.28$

(ii) Shade the area under the curve that gives you the probability $P(X \leq 2)$. *[1]*



[2] **(iii)** Shade the area under the curve that gives you the probability $P(X \leq 1 \cup X > 2)$.



[1] **(iv)** The figure shows the graph of the CDF of an exponential random variate. What is the median of this distribution?

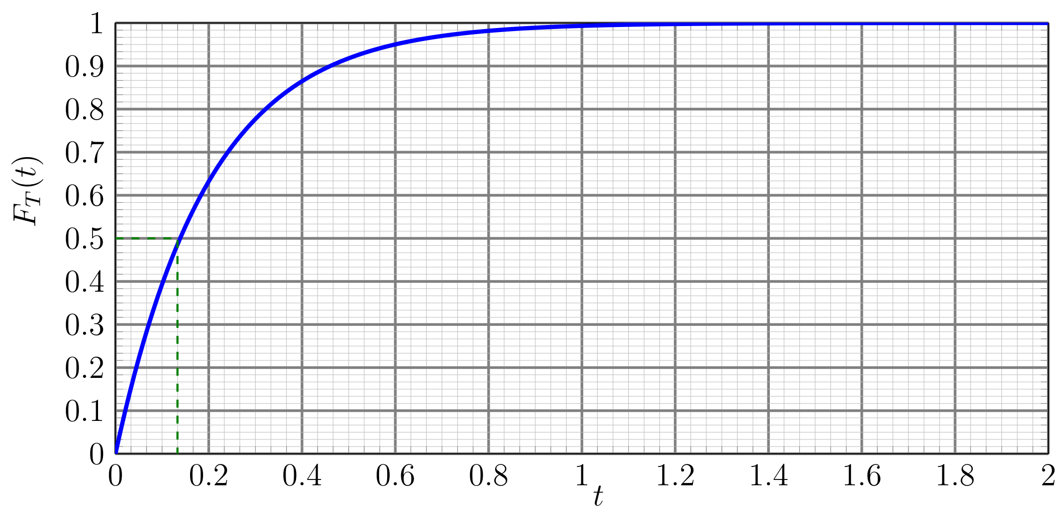
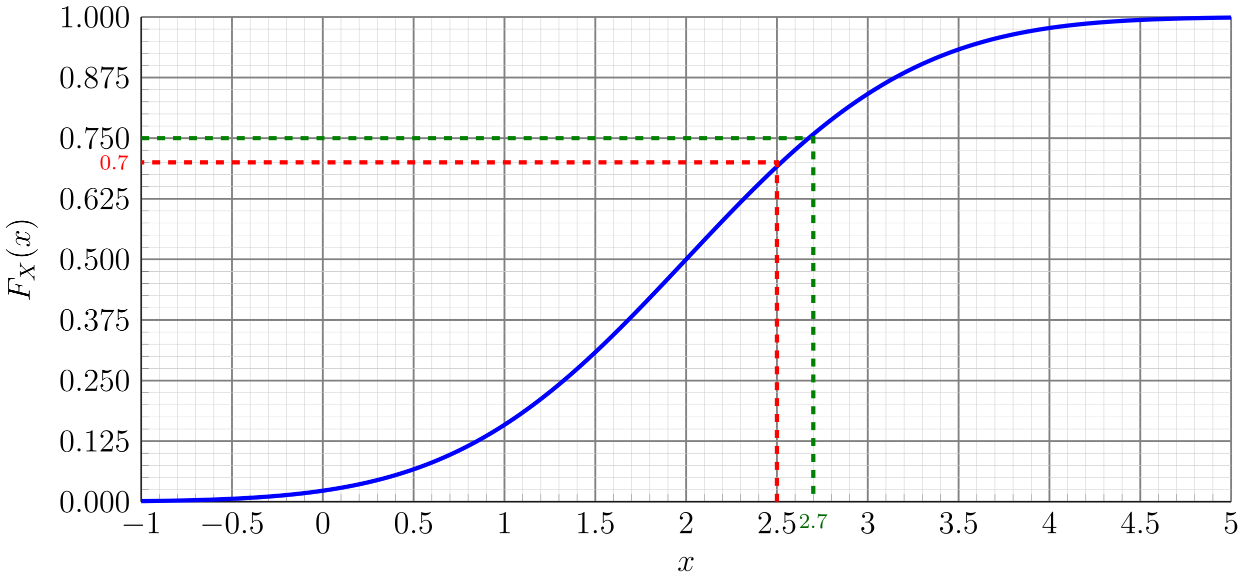


Figure 1: CDF of an exponential random variate

Answer: Median t_m is given by $F_T(t_m) = 0.5$. Thus, $t_m \approx 0.133$

(v) Below is the CDF of a given normal distribution. Use the figure to answer the following 4 questions (a – d).



(a) What is the median of this distribution? *[1]*

Answer: $x_m = 2$ (where $F_X(x) = 0.5$)

(b) What is the mean of this distribution? *[1]*

Answer: $\mu_X = 2$ (in normal distr., mean = median)

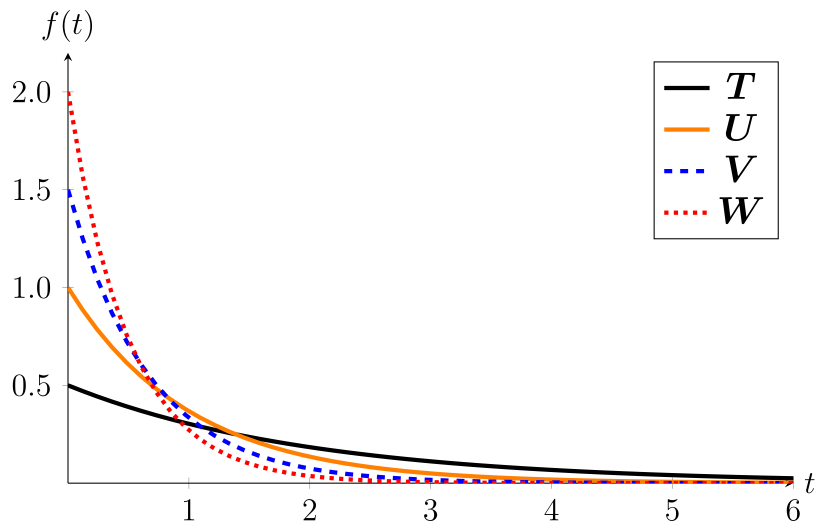
(c) Estimate the third quartile. *[1]*

Answer: $Q3 \approx 2.7$ (given as $F_X^{-1}(0.75)$)

(d) Estimate the probability $P(X > 2.5)$. *[1]*

Answer: $P(X > 2.5) \approx 1 - 0.7 = 0.3$

[1] **(vi)** Consider the PDFs of the epxonential random variates T, U and V (measured in hours) shown in the figure below. Which of them has the greatest mean?

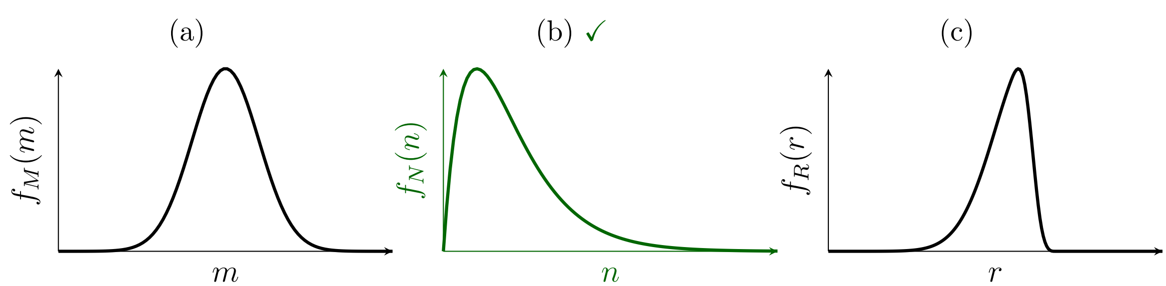


Answer: T has the greatest mean: $\lambda_T = 0.5 \implies \mu_T = \frac{1}{\lambda} = 2$

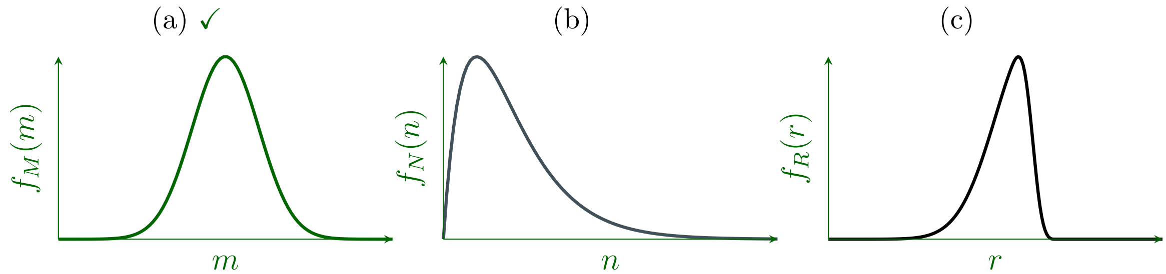
[1] **(vii)** In the figure above, which random variable has standard deviation of 2 hours?

Answer: T: $\sigma_T = \mu_T = 2$.

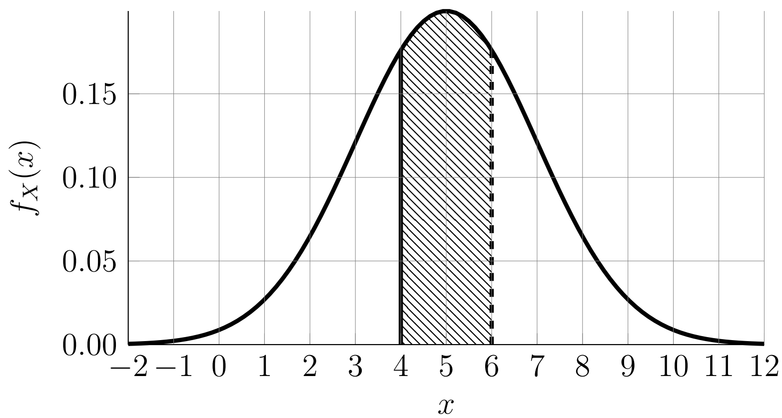
[1] **(viii)** In which of the distributions does the mode appear to be less than the mean? Circle (a), (b) or (c).



(ix) In which of the distributions does the mean appear equal to the median? Circle (a), (b) or (c). *[1]*



(x) $X \sim N(\mu = 5, \sigma = 2)$. Find the probability indicated by the shaded portion of the PDF below. *[2]*



Answer:

$$P(4 \le X < 6) = \Phi\left(\frac{6-5}{2}\right) - \Phi\left(\frac{4-5}{2}\right)$$

$= $ `normcdf(.5) - normcdf(-.5)`

$= .383$

Problem 4 *Counting (6 points)*

[3] **(a)** License plates in a certain state consist of 4 digits followed by 3 letters. How many different license plates can be manufactured?

Let n be the number of possible license plates. There are 10 digits available for each of the first four characters, and then 26 letters available for each of the next 3 characters. Thus:

$$\begin{aligned} n &= 10 \times 10 \times 10 \times 10 \times 26 \times 26 \times 26 \\ &= 10^4 \times 26^3 \\ &= \boxed{175,760,000} \approx \boxed{1.76 \times 10^8} \end{aligned}$$

[3] **(b)** How many distinct rearrangements of the letters in the word *PARALLELOGRAM* are possible?

Let the number of distinct rearrangements be n. We consider that the original spelling is excluded from n.

We also note that the following letters are repeated in the word: $A(3)$, $R(2)$ and $L(3)$, while the word itself has 13 letters.

Thus:

$$\begin{aligned} n &= \frac{13!}{3!2!3!} = \frac{13 \cdot 12 \cdot 11 \cdot \cdots \cdot 3 \cdot 2 \cdot 1}{6 \cdot 2 \cdot 6} \\ &= 13 \cdot 11 \cdot 10 \cdot 9 \cdot \cdots \cdot 5 \cdot 4 \\ &= \boxed{86,486,400} \approx \boxed{8.6 \times 10^7} \end{aligned}$$

Problem 5 *Bayes' and total probability (9 points)*

Given that $P(A) = 0.06$[1], $P(B) = 0.3$ and $P(C) = 0.1$ represent the production of machines in a factory. The conditional probabilities of defective items are $P(D|A) = 0.02$, $P(D|B) = 0.03$ and $P(D|C) = 0.04$.

(a) Find the probability $P(D)$. We apply the theorem of total probability. *[3]*

Solution with $P(A) = 0.06$:

$$\begin{aligned} P(D) &= P(D|A)P(A) + P(D|B)P(B) + P(D|C)P(C) \\ &= 0.02(0.06) + 0.03(0.3) + 0.04(0.1) = \boxed{0.0142} \end{aligned}$$

Solution with $P(A) = 0.6$:

$$P(D) = 0.02(0.6) + 0.03(0.3) + 0.04(0.1) = \boxed{0.025}$$

(b) Find the probability that an item was produced by machine A, given that it is defective. *[3]*

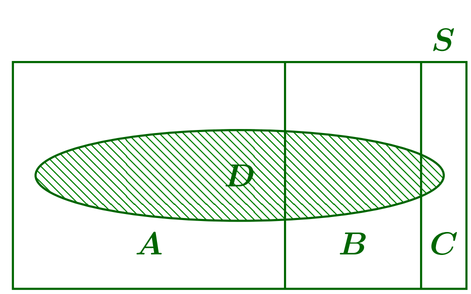
Solution with $P(A) = 0.06$:

$$\text{Bayes' Theorem: } P(A|D) = \frac{P(D|A)P(A)}{P(D)} = \frac{0.02(0.06)}{0.0142} = \boxed{0.0845}$$

Solution with $P(A) = 0.6$:

$$\text{Bayes' Theorem: } P(A|D) = \frac{P(D|A)P(A)}{P(D)} = \frac{0.02(0.6)}{0.025} = \boxed{0.48}$$

(c) Draw a Venn diagram depicting the interaction among the events A, B, C and D in sample space S. *[3]*



[1] This was a typo. It should have been rendered as $P(A) = 0.6$. Solutions for both cases are given.

Problem 6 *Binomial distribution (9 points; 5 points EC)*

Seventy percent of all vehicles examined at a certain emissions inspection station pass the inspection. Assuming that successive vehicles pass or fail independently of one another, find the following:

[2] (a) The variance of the binomial distribution governing the probability of the next four vehicles passing inspection.

$$\sigma^2 = np(1-p) = 4(0.7)(1-0.7) = 4(0.7)(0.3) = 0.84$$

[3] (b) The probability that all of the next four vehicles inspected pass.

The probability that a vehicle will pass is $p = 0.7$. Thus, the probability that all four successive vehicles inspected will pass is:

$$P(\text{next four vehicles pass}) = (0.7)^4 = 0.2401$$

More formally, we can model the number of vehicles passing as a binomial random variable X with $n = 4$. So we can see that in this case, we just found:

$$P(X = 4) \equiv \texttt{binopdf}(4, 4, .7) = \binom{4}{4} 0.7^4 0.3^{4-4} = 1(0.7^4)(1) = 0.2401$$

[4] (c) The probability that at least one of the next four vehicles inspected fails.

The event that at least one of the next 4 vehicles fails inspection is the complement of the event that none of the vehicles fail inspection. Thus:

$$\begin{aligned} P(\text{at least one of the next four vehicles fails}) &= 1 - P(\text{next four vehicles pass}) \\ &= 1 - 0.2401 = \boxed{0.7599} \end{aligned}$$

OR, using the binomial distribution, we can find $P(X \leq 3)$:

$$\begin{aligned} P(X \leq 3) &= \sum_{k=0}^{3} \binom{4}{k} 0.7^k 0.3^{4-k} \equiv \texttt{binocdf(3, 4, 0.7)} \\ &= 1 - P(X = 4) \\ &= 1 - 0.2401 = \boxed{0.7599} \end{aligned}$$

*You can also check a binomial CDF table (with $n = 4; p = 0.7; x = 3$) to see that you get the same answer.

(d) (Extra Credit) Use the normal distribution to estimate the probability that 40 of the next 200 vehicles inspected will pass inspection. (Show all your work to earn all the extra points.) *[5]*

The associated normal distribution has mean $\mu = np = 200(0.7) = 140$ and a variance of $\sigma^2 = np(1-p) = 140(0.3) = 42$.

One way to interpret this question, given that the normal distribution is continuous is to find the probability that X is between a narrow range centered on 40, i.e. $P(39.5 \leq X < 40.5)$:

$$\begin{aligned} P(39.5 \leq X < 40.5) &= \texttt{normcdf}(40.5, 140, \texttt{sqrt}(42)) - \texttt{normcdf}(39.5, 140, \texttt{sqrt}(42)) \\ &\approx \boxed{0} \end{aligned}$$

This tells us that this particular event is improbable.

An alternative way to interpret this question is to simply find $P(X \leq 40)$ and see what this gives:

$$\begin{aligned} P(X \leq 40) &= \texttt{normcdf}(40, 140, \texttt{sqrt}(42)) \\ &\approx \boxed{0} \end{aligned}$$

The result is the same (an extremely small number that is essentially zero).

Thus, the desired probability is 0 (an improbable event).

Problem 7 *Poisson distribution (8 points)*

Occasionally an airline will lose a bag. Suppose a small airline has found it can reasonably model the number of bags lost each weekday using a Poisson model with a variance of 5.

[2] **(a)** What is the mean of the Poisson distribution?

Let X be the number of bags lost each weekday by the airline. The mean of X is given by the rate parameter λ, which is the same as the variance in a Poisson distribution. Thus:

$$E(X) = \lambda = 5 \text{ bags}$$

[2] **(b)** What is the probability that the airline will lose 3 bags next Friday?

$$\begin{aligned} P(X=3) &= \frac{\lambda^x}{x!}e^{-\lambda} \equiv \texttt{poisspdf}(3,5) \\ &= \frac{5^3}{3!}e^{-5} \\ &= \boxed{0.1404} \end{aligned}$$

[4] **(c)** Find the probability that the airline will lose no more than 2 bags over two consecutive days.

In a two-day interval, $\lambda_2 = 2(5) = 10$. Thus,

$$\begin{aligned} P(X \leq 2) &\equiv \texttt{poisscdf}(2,10) \\ &= \sum_{k=0}^{2} \frac{\lambda^k}{k!}e^{-\lambda} \\ &= e^{-10} + 10e^{-10} + \frac{10^2}{2}e^{-10} \\ &= e^{-10}(1+10+50) \\ &= \boxed{0.0028} \end{aligned}$$

Problem 8 *Normal distribution (6 points)*

The mean daily high temperature in June in LA is 77°F with a standard deviation of 5°F. Suppose that the temperatures in June closely follow a normal distribution.

(a) What is the probability that the high temperature on a random day in June in LA is greater than 80°F? *[3]*

$\mu = 77$ and $\sigma = 5$. Thus:

$$\begin{aligned} P(X > 80) &\equiv \texttt{normcdf(80, 77, 5, 'upper')} \\ &= 1 - P(X \leq 80) \\ &= 1 - \Phi\left(\frac{80 - 77}{5}\right) \\ &= 1 - \Phi(0.6) \equiv \texttt{1 - normcdf(0.6)} \\ &= \boxed{0.2743} \end{aligned}$$

(b) What is the 30th percentile of daily high temperatures in June in LA? *[3]*

We use the inverse CDF of 0.3 to find P_{30} (30th percentile):

$$\begin{aligned} \frac{P_{30} - 77}{5} &= \Phi^{-1}(0.3) \equiv \texttt{norminv}(0.3) = -.5244 \\ \implies P_{30} &= 77 + 5(-.5244) \\ &\approx \boxed{74.38°\text{F}} \end{aligned}$$

You can also solve directly without standardizing as follows:

$$P_{30} \equiv \texttt{norminv}(0.3, 77, 5) = 74.378 \approx \boxed{74.38°\text{F}}$$

Problem 9 *Logormal distribution (10 points)*

The lifetime of a drill (number of holes that a drill machines before it breaks) is lognormally distributed with a $\mu = 4.5$ and $\sigma = 0.8$.

[4] (a) Find the mean and standard deviation of lifetime.

Let X be the drill lifetime. Thus:

$$\begin{aligned} E(X) &= \exp\left(\mu + \frac{1}{2}\sigma^2\right) = \exp\left(4.5 + \frac{1}{2}\left(0.8^2\right)\right) \\ &\approx \boxed{124 \text{ holes}} \\ SD(X) &= \sqrt{Var(X)} = \sqrt{E(X)^2\left(\exp(\sigma^2) - 1\right)} \\ &= 124\sqrt{\mathrm{e}^{0.8^2} - 1} \\ &\approx \boxed{117.4 \text{ holes}} \end{aligned}$$

[3] (b) What is the probability that the lifetime is at most 100?

$$\begin{aligned} P(X \leq 100) &\equiv \texttt{normcdf}(\texttt{log}(100), 4.5, 0.8) \equiv \texttt{normcdf}((\texttt{log}(100) - 4.5)/0.8) \\ &\equiv \texttt{logncdf}(100, 4.5, 0.8) \\ &= \Phi\left(\frac{\ln 100 - 4.5}{0.8}\right) \\ &= \Phi(0.1315) \equiv \texttt{normcdf}(0.1315) = \boxed{0.552} \end{aligned}$$

[3] (c) What is the probability that the lifetime is greater than 200?

$$\begin{aligned} P(X > 200) &\equiv \texttt{normcdf(log(200), 4.5, 0.8, 'upper')} \\ &\equiv \texttt{logncdf(200, 4.5, 0.8, 'upper')} \\ &= 1 - P(X \leq 200) \\ &= 1 - \Phi\left(\frac{\ln 200 - 4.5}{0.8}\right) \\ &= 1 - \Phi(0.9979) \equiv \texttt{1 - normcdf(0.9979)} \\ &= \boxed{0.159} \end{aligned}$$

Problem 10 *Exponential distribution (8 points)*

The delay time T of a train is exponentially distributed with $\lambda = 3$ (mean rate of occurrence per hour).

(a) What is the mean of T? *[1]*

The mean is given by

$$E(T) = \frac{1}{\lambda} = \frac{1}{3} \text{ hr}$$

(b) What is the variance of T? *[1]*

The variance is given by

$$Var(T) = \frac{1}{\lambda^2} = \frac{1}{3^2} = \frac{1}{9} \text{ hr}^2$$

(c) What is the probability that a train is delayed by no more than 1 hour? *[3]*

$$\begin{aligned} P(T \leq 1) &= 1 - e^{-3(1)} \equiv \texttt{expcdf}(1, 1/3) \\ &= \boxed{0.9502} \end{aligned}$$

(d) Given that a family member has already waited for 2 hours, what is the probability that a certain flight will be further delayed by over an hour? *[3]*

Use the memoryless property:

$$\begin{aligned} P(T > 1 + 2 | T > 2) &= P(T > 1) \equiv \texttt{expcdf(1, 1/3, 'upper')} \\ &= e^{-3(1)} \\ &\approx \boxed{0.0498} \end{aligned}$$

Problem 11 *Joint distributions (11 points; 3 points EC)*

The joint PMF of two random variables X and Y is shown in the table below.

$p_{X,Y}(x,y)$		y = **0**	**1**	**2**
	0	0.10	0.06	0.02
x	**1**	0.12	0.04	0.01
	2	0.20	0.30	0.15

[2] **(i)** Find $P(X = 0 \cap Y = 1)$.

$$P(X = 0 \cap Y = 1) = \boxed{0.06}$$

[3] **(ii)** Compute $P(X < 1)$.

$$\begin{aligned} P(X < 1) &= P(X = 0) \\ &= 0.1 + 0.06 + 0.02 = \boxed{0.18} \end{aligned}$$

[3] **(iii)** Compute $P(Y \geq 1)$.

$$\begin{aligned} P(Y \geq 1) &= 1 - P(Y < 1) = 1 - P(Y = 0) \\ &= 1 - (0.1 + 0.12 + 0.2) \\ &= 1 - 0.42 = \boxed{0.58} \end{aligned}$$

Note that you could also add up the six probabilities corresponding to $Y \geq 1$ to find the same answer, but this is more tedious.

(iv) Compute $P(Y = 0|X = 0)$. *[3]*

$$\begin{aligned} P(Y = 0|X = 0) &= \frac{0.1}{0.1 + 0.06 + 0.02} \\ &= \frac{0.1}{0.18} \\ &\approx \boxed{0.556} \end{aligned}$$

(v) (Extra Credit) Find the marginal distribution $p_X(x)$. *[3]*

$$\begin{aligned} &P(X = 0) = 0.18 \\ &P(X = 1) = 0.12 + 0.04 + 0.01 = 0.17 \\ &P(X = 2) = 0.2 + 0.3 + 0.15 = 0.65 \end{aligned}$$

Thus,

$$p_X(x) = \begin{cases} 0.18, & x = 0 \\ 0.17, & x = 1 \\ 0.65, & x = 2 \\ 0, & \text{otherwise} \end{cases}$$

CEE 260/MIE 273 | Jimi Oke | Fall 2021

Department of Civil and Environmental Engineering

University of Massachusetts Amherst