E1: Midterm Exam

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

October 16, 2024

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 **10 problems, 106 points** (with 5 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/Python to obtain a result. In the case of MATLAB/Python, 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 (15 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 mutually exclusive events is 1. Only if they are also collectively exhaustive.

(ii) **F** If two events A and B are collectively exhaustive, then $P(A) - P(B) = 1$.Only in the special case when $P(A)$ is certain and $P(B)$ is impossible.

(iii) **T** For a given online account, your password must have seven small or capital letters. Assuming you are the first user to create an account, the number of possibilities for your password is 52^7.

(iv) **F** The area under the curve of a PDF can be less than 1. It's always 1, otherwise it's not a PDF.

(v) **F** You begin your turn in a board game by rolling a six-sided die. If all outcomes are equally likely, then the probability of rolling a "2" is $\frac{2}{6}$. It is $\frac{1}{6}$.

(vi) **T** In a right-skewed distribution, the mean is not equal to the median. Mean lies to the right of median in right-skewed distribution.

(vii) **T** $P(A \cup B) = P(A) + P(B)$ for two events A and B that are mutually exclusive.

(viii) **T** If A and B are statistically independent events, then $P(AB)$ is always equal to $P(A)P(B)$.

(ix) **T** The complement of the union of collectively exhaustive events yields an empty set.

(x) **T** The maximum value of any cumulative distribution function is 1.

(xi) **F** Under certain conditions, a binomial distribution with parameters (n, p) can be approximated by a normal distribution with $\sigma = np(1-p)$. It is σ^2 rather that equals $np(1-p)$

(xii) **T** The probability of the number of tails that occur in 100 tosses of a coin (which can yield either a head or a tail) can be appropriately modeled by a binomial distribution. It's a Bernoulli sequence.

(xiii) **F** Given a normal distribution with parameters μ and σ, the median of the distribution is μ^2. The median is μ.

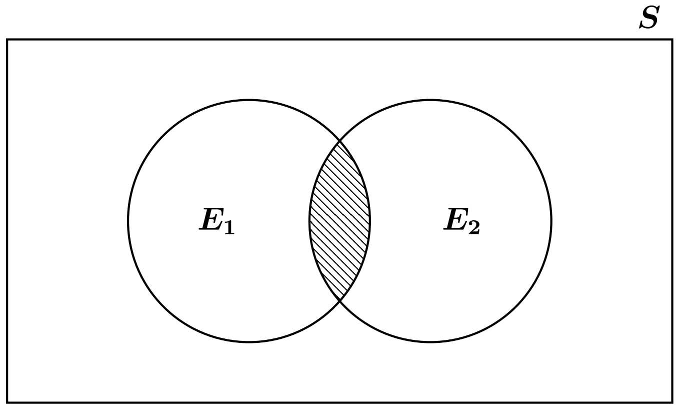
(xiv) **T** The standard normal variate Z has a variance of 1.

(xv) **T** The probability of the elapsed time between events that constitute a Poisson process can be modeled using an exponential distribution.

Problem 2 *Venn diagrams (11 points)*

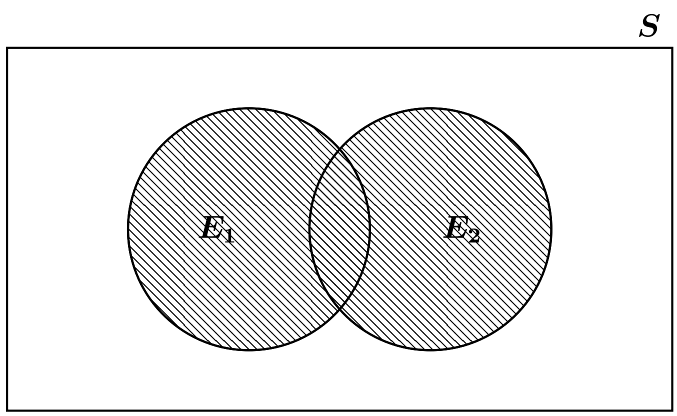
Write the combination of events (using set notation) depicted in each of the figures below.

(i) *[1]*



Answer: $\boldsymbol{E_1 \cap E_2}$

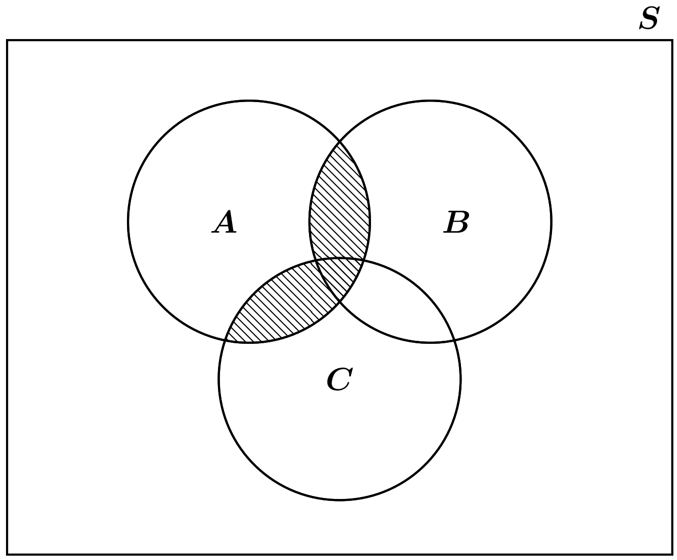
(ii) *[2]*



Answer: $\boldsymbol{E_1 \cup E_2}$

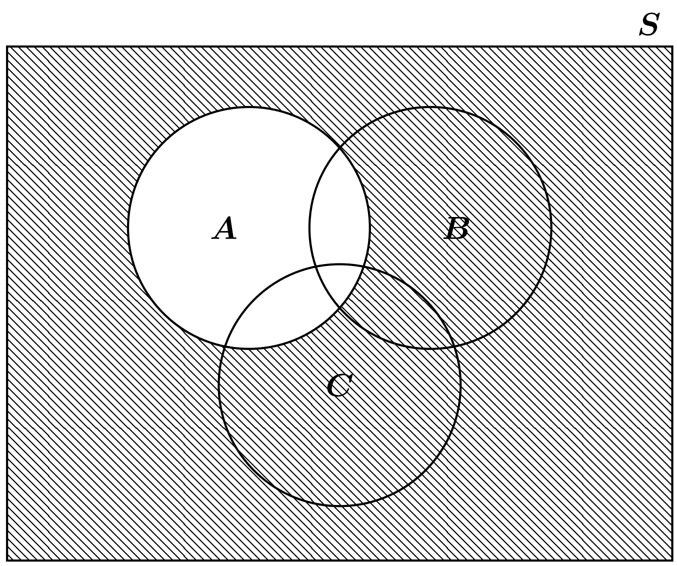
Award 1 pt if not completely correct, e.g. "$\boldsymbol{E}_1 + \boldsymbol{E}_2$" or "$\boldsymbol{E}_1 \cup \boldsymbol{E}_2 - \boldsymbol{E}_1\boldsymbol{E}_2$"

[2] **(iii)**



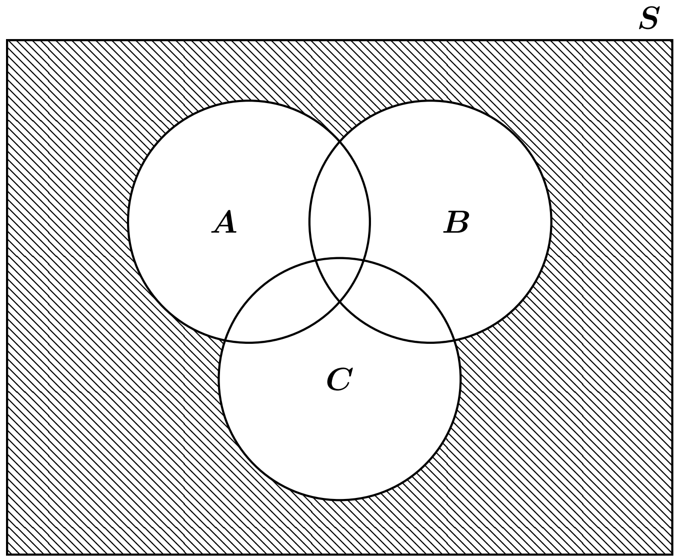
Answer: $\boldsymbol{AB} \cup \boldsymbol{AC}$ OR $\boldsymbol{A} \cap (\boldsymbol{B} \cup \boldsymbol{C})$

[2] **(iv)**



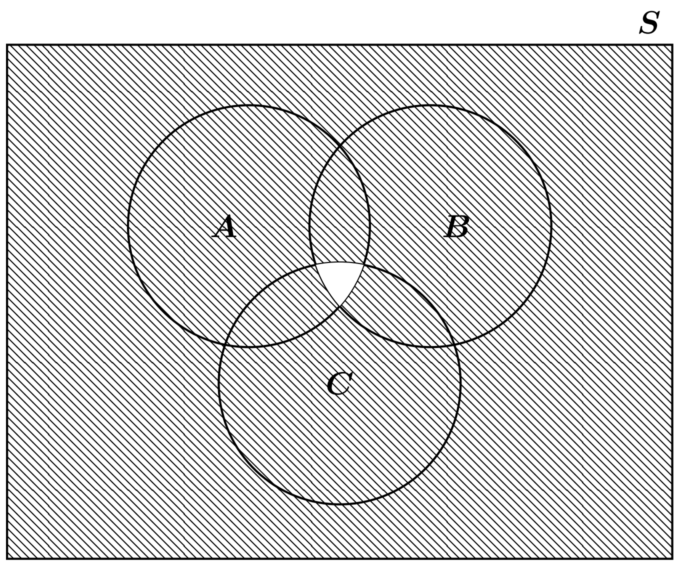
Answer: $\boldsymbol{A}^c$ OR $\overline{\boldsymbol{A}}$

(v) *[2]*



Answer: $\overline{\boldsymbol{A} \cup \boldsymbol{B} \cup \boldsymbol{C}}$

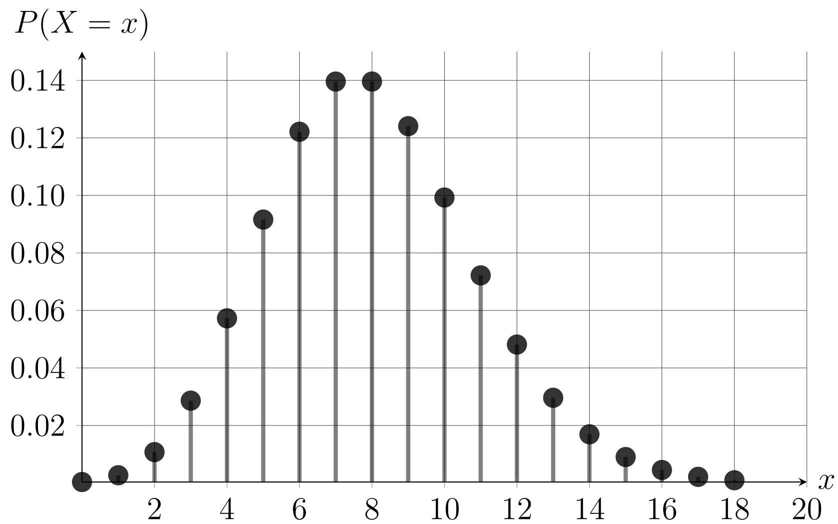
(vi) *[2]*



Answer: $\overline{ABC}$

Problem 3 *Short answer questions (22 points)*

(a) The PMF of a random variable X is given in the figure below. Use the figure to answer the following questions.



[1] **(i)** Estimate the probability $P(X = 4)$.

Answer: ~ 0.06 (Award point for any number between 0.05 and 0.06.)

[2] **(ii)** Estimate the probability $P(6 < X \leq 8)$.

Answer: $\sim 0.14 + 0.14 = 0.28$

[1] **(iii)** If the distribution shown has a single parameter λ, which of the following would be your best guess for λ? (Circle or underline the correct answer.)

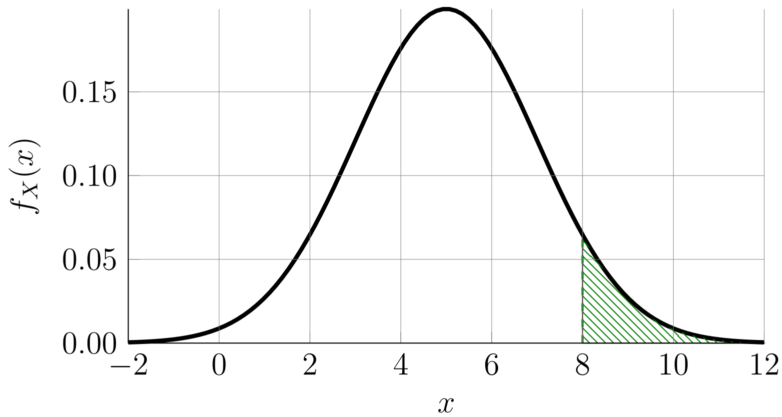
A. 2

B. 8

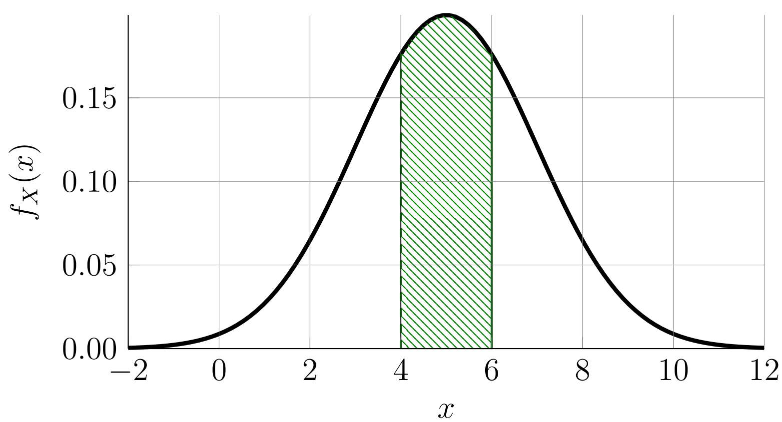
C. 10

D. 12

(b) Shade the area under the curve that represents the probability $P(X > 8)$. *[1]*



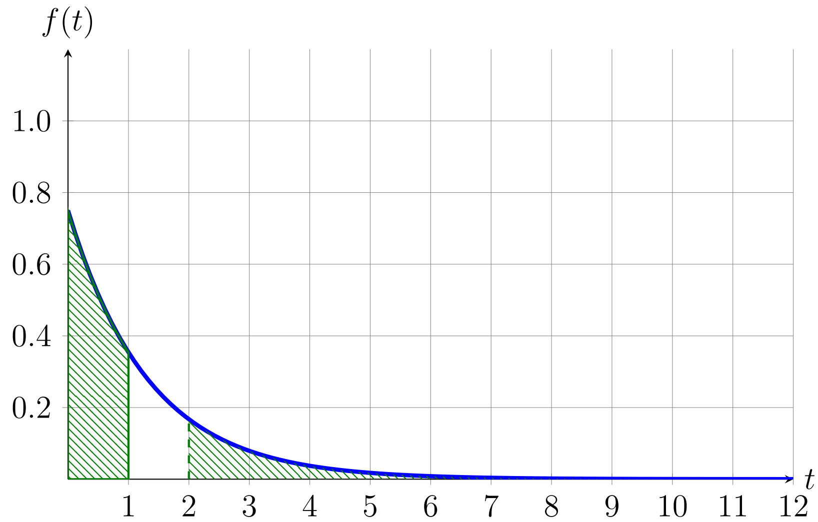
(c) Shade the area under the curve that represents the probability $P(4 \leq X < 6)$. *[1]*



(d) If the curve in part **(c)** is the PDF of a normal distribution, what is its mean value? *[1]*

Answer: 5

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



[1] **(f)** The figure shows the graph of the CDF of an exponential random variate. What is the 30th percentile of this distribution?

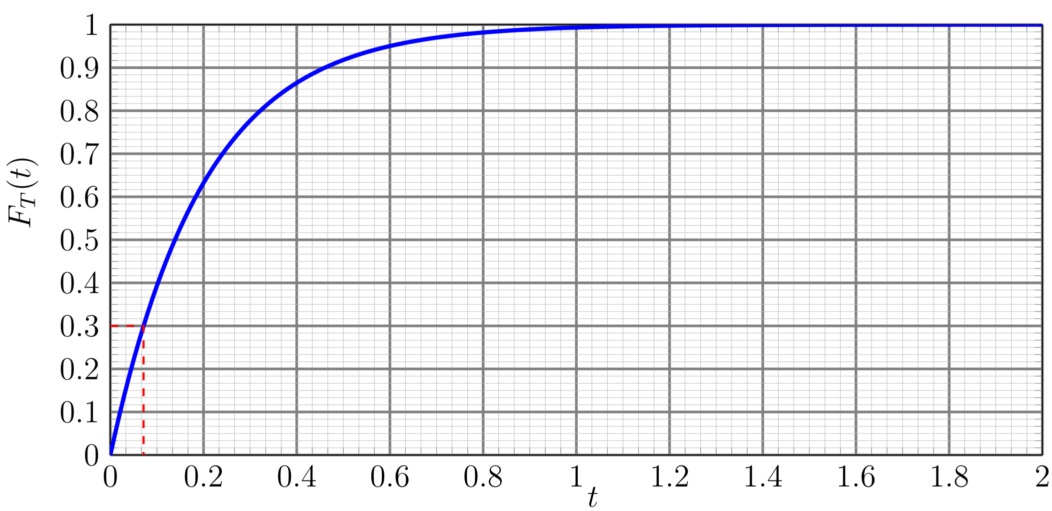
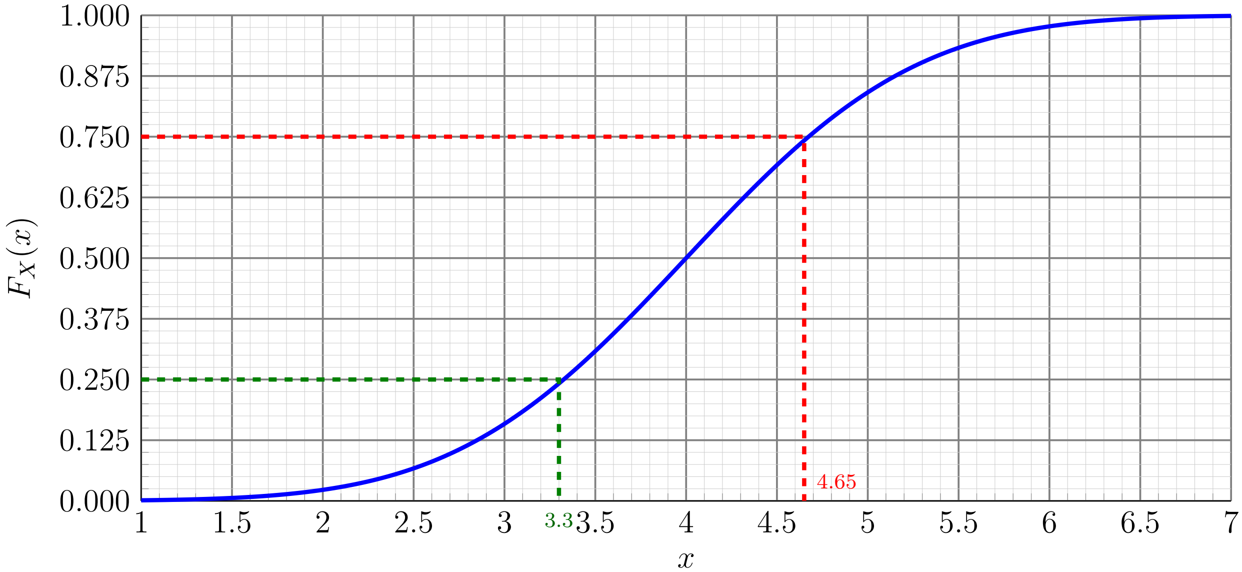


Figure 1: CDF of an exponential random variate

Answer: ~ 0.07

(g) Below is the CDF of a given normal distribution. Use the figure to answer the following 5 questions (i) – (v).



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

Answer: 4

(ii) What is the mode of this distribution? *[1]*

Answer: 4

(iii) Estimate the first quartile. *[1]*

Answer: 3.3

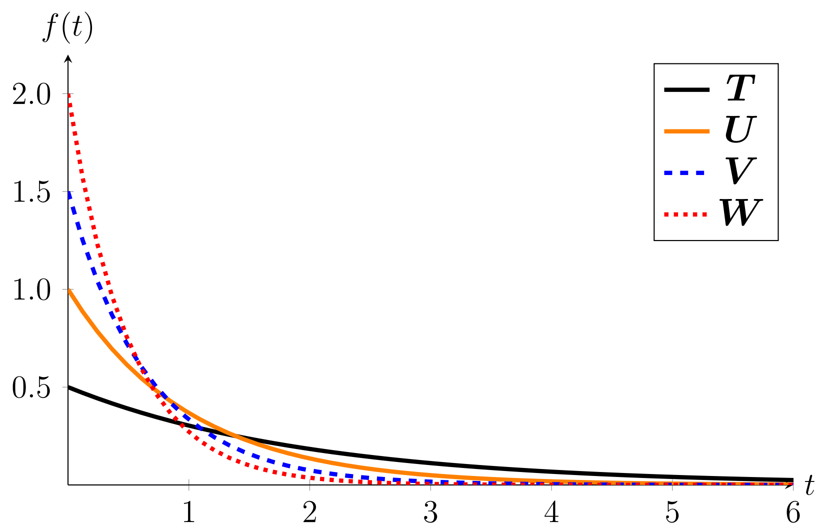
(iv) Estimate the third quartile. *[1]*

Answer: 4.65

(v) Compute the interquartile range (IQR). *[1]*

Answer: $IQR = Q_3 - Q_1 = 4.65 - 3.3 = 1.35$

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

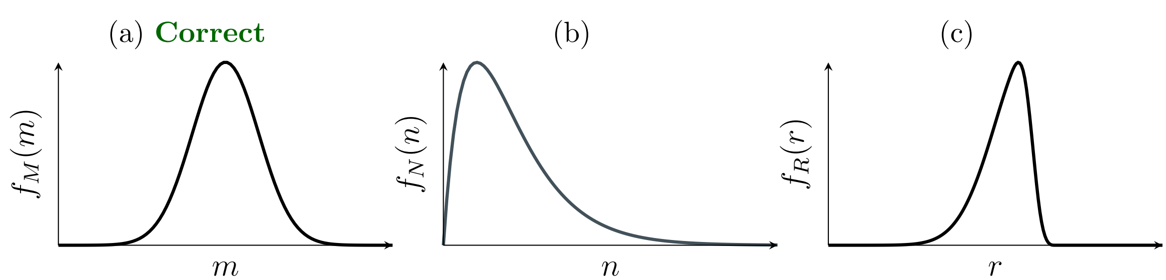


Answer: $\boldsymbol{T}$ (Reason: $\mu = 1/\lambda = 1/0.5 = 2$ hrs)

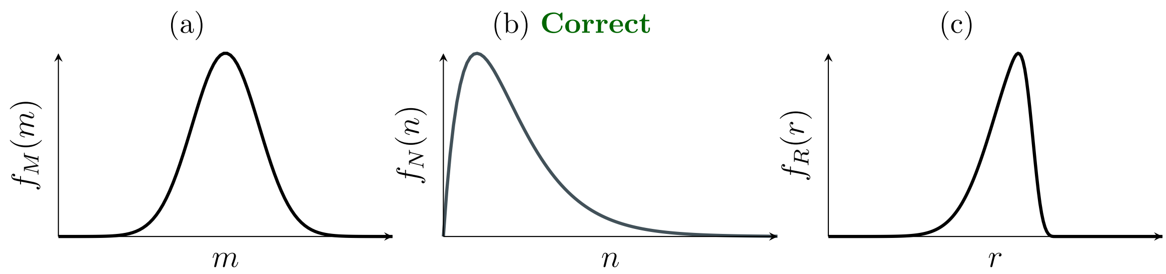
[1] **(i)** In the figure above, which random variable has standard deviation of 2 hours?

Answer: $\boldsymbol{T}$; for exponential distribution, mean = SD

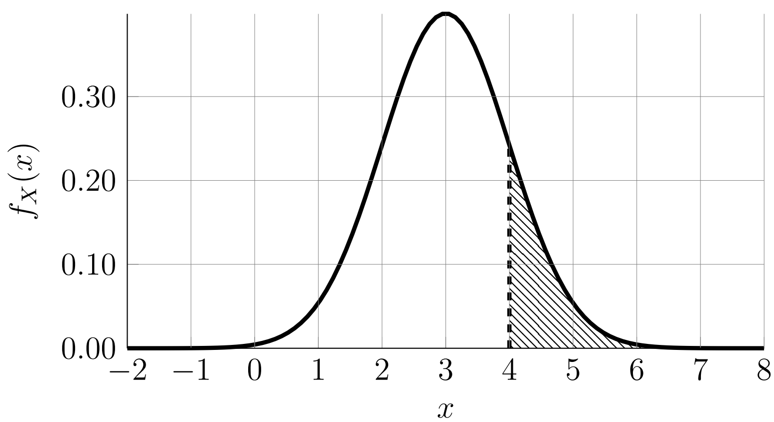
[1] **(j)** In which of the distributions does the mode appear to be equal to the mean? Circle (a), (b) or (c).



(k) Which of the distributions is right-skewed? Circle (a), (b) or (c). *[1]*



(l) The figure below shows the PDF of a normal distribution with a variance of 1. Compute the probability indicated by the shaded portion of the PDF below. *[3]*



$\mu = 3$ and $\sigma = 1$. Thus:

Answer:

$$
\begin{aligned}
P(X > 4) &= 1 - \Phi\left(\frac{4-3}{1}\right) \\
&= 1 - \Phi(1) \\
&\equiv \texttt{normcdf}(1, \text{`}\texttt{upper}\text{'}) \\
&= \boxed{.1587}
\end{aligned}
$$

Problem 4 *Counting (7 points)*

[3] **(a)** License plates in a certain state consist of 1 digit, followed by 3 letters and then 2 digits. How many different license plates can be manufactured?

Let n be the number of possible license plates. There are 10 digits available for the first character, and then 26 letters available for each of the next 3 characters and finally 10 digits available each for the last 2 characters. Thus:

$$\begin{aligned} n &= 10 \times 26 \times 26 \times 26 \times 10 \times 10 \\ &= 10^3 \times 26^3 \\ &= \boxed{17,576,000} \approx \boxed{1.76 \times 10^7} \end{aligned}$$

[2] **(b)** In how many ways can you arrange 6 items?

Let the number of distinct arrangements be n. Then:

$$n = 6! = \boxed{720}$$

[2] **(c)** How many distinct groups of 3 can you obtain from a larger group of 5 items?

Let the number of distinct groups be n. Then:

$$\begin{aligned} n &= \binom{5}{3} \\ &= \frac{5!}{3!(5-3)!} = \frac{5 \times 4 \times 3!}{3!2!} = \frac{20}{2} \\ &= \boxed{10} \end{aligned}$$

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

Given that $P(A) = 0.6$, $P(B) = 0.3$ and $P(C) = 0.1$ represent the production of machines in a factory. The conditional probabilities of damaged items are $P(D|A) = 0.02$, $P(D|B) = 0.03$ and $P(D|C) = 0.04$.

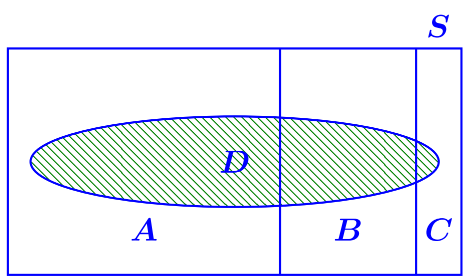
(a) Find the total probability $P(D)$. We apply the theorem of total probability. *[3]*

$$P(D) = P(D|A)P(A) + P(D|B)P(B) + P(D|C)P(C) \quad \textit{(1 pt for correct approach)}$$
$$= 0.02(0.6) + 0.03(0.3) + 0.04(0.1) \quad \textit{(1 pt for correct substitutions)}$$
$$= \boxed{0.025} \quad \textit{(1 pt for correct answer)}$$

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

$$\text{Bayes' Theorem: } P(B|D) = \frac{P(D|B)P(B)}{P(D)} = \frac{0.03(0.3)}{0.025} = \boxed{0.36}$$

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



- 1 pt if all 5 events are represented
- 1 pt if A, B and C are depicted as collectively exhaustive (span entire sample space)
- 1 pt is D spans the events A, B and C

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

80% of all vehicles inspected at a certain facility station pass the inspection. Assuming that successive vehicles pass or fail independently of one another, find the following:

[2] **(a)** The standard deviation of the binomial distribution governing the probability of the next 10 vehicles passing inspection.

$n = 10$ and $p = 0.8$. Thus,

$\sigma = \sqrt{np(1-p)} = \sqrt{10(0.8)(1-0.8)} = \sqrt{10(0.8)(0.2)} = \boxed{1.265}$

[3] **(b)** The probability that half of the next 10 vehicles inspected pass.

Here, we want to find:

$$\begin{aligned} P(X=5) &= \binom{10}{5}(0.8)^5(0.2)^5 \\ &\equiv \texttt{binopdf(5, 10, 0.8)} \\ &= \boxed{0.0264} \end{aligned}$$

[4] **(c)** The probability that at least 4 of the next 10 vehicles inspected fail.

Note that if at least 4 out of 10 fail, then at most, 6 pass. Thus, we want to find:

$$\begin{aligned} P(X \le 6) &= \sum_{k=0}^{6} \binom{10}{k}(0.8)^k(0.2)^{n-k} \\ &\equiv \texttt{binocdf(6, 10, 0.8)} \\ &= \boxed{.1209} \end{aligned}$$

(d) (Extra Credit) Use the normal distribution to estimate the probability that 95 of the next 100 vehicles inspected will pass inspection. (Show all your work to earn all the extra points.) *[5]*

When n is sufficiently large, the binomial distribution can be approximated by a normal distribution with $\mu = np$ and $\sigma = \sqrt{np(1-p)}$. Thus:

$$\mu = np = 100 * 0.8 = 80$$
$$\sigma^2 = np(1-p) = 100(.8)(.2) = 16$$
$$\therefore \sigma = \sqrt{16} = 4$$

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 95, i.e. $P(94.5 \leq X < 95.5)$:

$$P(94.5 \leq X < 95.5) = \texttt{normcdf}(95.5, 80, 4) - \texttt{normcdf}(94.5, 80, 4)$$
$$\approx \boxed{9.12 \times 10^{-5}}$$

This tells us that the chances of this particular event are very small. This makes sense, as the mean (expected) value of vehicles that should pass after 100 are inspected is 80.

An alternative way to interpret this question is to simply find $P(X > 95)$ and see what this gives:

$$P(X > 95) = \texttt{normcdf}(95, 80, 4, '\texttt{upper}')$$
$$\approx \boxed{8.84 \times 10^{-5}}$$

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)*

Hourly arrivals at a small coffee shop can be model using a Poisson distribution with a variance of 4.

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

Let X be the number of hourly arrivals. The mean of X is given by the rate parameter λ, which is the same as the variance in a Poisson distribution. Thus:

$$\mathbb{B}(X) = \lambda = 4 \text{ customers}$$

[2] **(b)** What is the probability that 5 customers will arrive in a given hour?

$$\begin{aligned} P(X = 5) &= \frac{\lambda^x}{x!} e^{-\lambda} \equiv \texttt{poisspdf}(5, 4) \\ &= \frac{4^5}{5!} e^{-4} \\ &= \boxed{0.1563} \end{aligned}$$

[4] **(c)** Find the probability that no more than 8 customers will arrive over a duration of 2 consecutive hours.

In a two-hour interval, $\lambda_2 = 2(4) = 8$. Thus,

$$\begin{aligned} P(X \leq 8) &= \sum_{k=0}^{8} \frac{\lambda^k}{k!} e^{-\lambda} \\ &\equiv \texttt{poisscdf}(8, 8) \\ &= \boxed{0.5925} \end{aligned}$$

Problem 8 *Normal distribution (7 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 50th percentile of daily high temperatures in June in LA? *[1]*

The 50th percentile is the median, which is equal to the mean in a normal distribution. Thus, it is 77°F.

(b) What is the probability that the high temperature on a random day in June in LA is lower than 60°F? *[3]*

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

$$\begin{aligned} P(X \leq 60) &\equiv \texttt{normcdf(60, 77, 5)} \\ &= \Phi\left(\frac{60-77}{5}\right) \\ &= \Phi(-3.4) \equiv \texttt{normcdf(-3.4)} \\ &= \boxed{3.37 \times 10^{-4}} \end{aligned}$$

(c) What is the probability that the high temperature on a random day in June in LA is between 60° and 80°F? *[3]*

$$\begin{aligned} P(X60 \leq X < 80) &= \Phi\left(\frac{80-77}{5}\right) - \Phi\left(\frac{60-77}{5}\right) \\ &\equiv \texttt{normcdf(80, 77, 5) - normcdf(60, 77, 5)} \\ &= \boxed{0.7254} \end{aligned}$$

The probability that the high on a random June day in LA is between 60 and 80 degrees F is about 73%.

Problem 9 *Lognormal distribution (10 points)*

The lifetime (in years) of a machine is lognormally distributed with $\mu = 2$ and $\sigma = 0.5$.

[2] (a) What is the median lifetime of the machine?

The median is given by $\exp(\mu) = \exp(2) = \boxed{7.389}$ years.

[2] (b) Find the mean lifetime.

Let X be the machine lifetime. Then the mean is given by:

$$\begin{aligned} \mathbb{E}(X) &= \exp\left(\mu + \frac{1}{2}\sigma^2\right) = \exp\left(2 + \frac{1}{2}\left(0.5^2\right)\right) \\ &\approx \boxed{8.373 \text{ years}} \end{aligned}$$

[3] (c) What is the probability that the lifetime is at least 5 years?

$$\begin{aligned} P(X \geq 5) &= 1 - P(X < 5) \\ &= 1 - \Phi\left(\frac{\ln 5 - 2}{0.5}\right) = 1 - \Phi(-.7811) \\ &\equiv \texttt{normcdf(-0.7811, 'upper')} \equiv \texttt{logncdf(5, 2, .5, 'upper')} \\ &= \boxed{0.7826} \end{aligned}$$

[3] (d) What is the probability that the lifetime is less than 2 years?

$$\begin{aligned} P(X \geq 2) &= \Phi\left(\frac{\ln 2 - 2}{0.5}\right) = \Phi(-2.6137) \\ &\equiv \texttt{normcdf(-2.6137)} \equiv \texttt{logncdf(2, 2, .5)} \\ &= \boxed{0.0045} \end{aligned}$$

Problem 10 *Exponential distribution (8 points)*

The delay time T of a bus is exponentially distributed with $\lambda = 1$ (mean rate of occurrence per hour).

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

The mean is given by

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

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

The variance is given by

$$\mathbb{V}(T) = \frac{1}{\lambda^2} = 1 \text{ hr}^2$$

(c) What is the probability that a train is delayed by no more than 15 minutes? *[3]*

15 minutes is 0.25 hr. Thus:

$$\begin{aligned} P(T \leq .25) &= 1 - e^{-\lambda(t)} = 1 - e^{-1(.25)} \\ &\equiv \texttt{expcdf}(.25, 1) = \boxed{0.2212} \end{aligned}$$

(d) Given that you have already waited for 30 minutes, what is the probability that a certain bus will be further delayed by more than 15 minutes? *[3]*

Use the memoryless property:

$$\begin{aligned} P(T > .5 + .25 | T > .5) &= \frac{P(T > .5 + .25) \cap P(T > .5)}{P(T > .5)} \\ &= \frac{P(T > .5 + .25)}{P(T > .5)} = \frac{P(T > .5) P(T > .25)}{P(T > .5)} = P(T > .25) \\ &= e^{-1(0.25)} \equiv \texttt{expcdf}(.25, 1, '\texttt{upper}') \approx \boxed{0.7788} \end{aligned}$$

CEE 260/MIE 273 | Jimi Oke | Fall 2024

Department of Civil and Environmental Engineering

University of Massachusetts Amherst