MAE108: Probability and Statistical Methods for Mechanical and Environmental Engineering
Midterm 2

To receive full credit you have to provide a detailed explanation of your answers.

1. (**20 points**) Like so many care-free San Diegans, your friend decided to get married at an outdoor ceremony by the beach. Unfortunately, the weather forecast predicts rain for the day of the wedding. *Use the following information to determine the probability that it will actually rain on the wedding day.* In recent years, San Diego had 42 rainy days each year. When it actually rains, the weather forecast correctly predicts rain 90% of the time. When it doesn't rain, it incorrectly predicts rain 10% of the time.

   Hint: Assume that all days are equally likely to have rain, i.e., ignore the seasonality of weather patterns.

Solution:

The sample space Ω consists of two events, on the wedding day it either rains (event A) or it does not rain (event B). Since these two events are mutually exclusive and collectively exhaustive, the form a partition of the space Ω. A third event (F) occurs when the weather forecast predicts rain.

The probability of having rain on the wedding day, $\mathbb{P}(A)$, is determined as

$$\mathbb{P}(A) = \frac{42 \text{ days}}{365 \text{ days}} \approx 0.115. \qquad \textbf{(5 pts)}$$

The probability of not having rain on the wedding day, $\mathbb{P}(B)$, is

$$\mathbb{P}(B) = 1 - \mathbb{P}(A) = 1 - 0.115 = 0.885.$$

The statement "When it actually rains, the weather forecast correctly predicts rain 90% of the time" translates into the conditional probability

$$\mathbb{P}(F|A) = 0.9.$$

The statement "When it doesn't rain, the forecast incorrectly predicts rain 10% of the time" translates into the conditional probability

$$\mathbb{P}(F|B) = 0.1.$$

The problem asks us to compute the probability $\mathbb{P}(A|F)$. This is done by using Bayes' theorem,

$$\mathbb{P}(A|F) = \frac{\mathbb{P}(F|A)\mathbb{P}(A)}{\mathbb{P}(F|A)\mathbb{P}(A) + \mathbb{P}(F|B)\mathbb{P}(B)} = \frac{0.9 \cdot 0.115}{0.9 \cdot 0.115 + 0.1 \cdot 0.885} \approx 0.54. \qquad \textbf{(15 pts)}$$

2. (**40 points**) In a certain coastal region, the average time between severe earthquakes is 80 years. Geoscientists decided to model the time between severe earthquakes as a lognormal random variable T whose coefficient of variation is 0.4.

   (i) Write down the probability density function for T, and determine the parameters in this distribution. **(10 pts)**

(ii) Determine the probability that a severe earthquake will occur within 20 years from the previous one. **(10 pts)**

(iii) Suppose the last severe earthquake in the region took place 100 years ago. What is the probability that a severe earthquake will occur over the next year? **(20 pts)**

Solution:

(i) Since T has a lognormal distribution, its PDF has the form

$$f_T(t) = \frac{1}{\sqrt{2\pi\sigma_X^2}} \frac{1}{t} \exp\left[-\frac{(\ln t - \mu_X)^2}{2\sigma_X^2}\right],$$

where the (unknown) parameters representing the mean (μ_X) and variance (σ_X^2) of the underlying Gaussian random variable $X = \ln T$. These parameters have to be determined from the known mean ($\mu_T = 80$) and coefficient of variation (c.o.v. = 0.4) of the random variable T. To begin with, we note that c.o.v. = σ_T/μ_T. Hence $\sigma_T = 0.4 \cdot 80 = 32$. We showed in class that

$$\sigma_X^2 = \ln\left[1 + \frac{\sigma_T^2}{\mu_T^2}\right] \approx 0.148 \qquad \text{and} \qquad \mu_X = \ln \mu_T - \frac{\sigma_X^2}{2} \approx 4.31.$$

(ii) The probability that a severe earthquake will occur within 20 years from the previous one is (see the class notes and Table A1 in the textbook)

$$\mathbb{P}(T \leq 20) = \Phi\left(\frac{\ln 20 - \mu_X}{\sigma_X}\right) \approx \Phi\left(\frac{3.00 - 4.31}{0.385}\right) = 1 - \Phi(3.41) = 1 - 0.9997 = 0.0003.$$

(iii) "The last earthquake took place 100 years ago" is a *given* event ($T > 100$). We are asked to determine the probability of the earthquake occurring in the next year ($100 < T \leq 101$) conditioned on the event $T > 100$, i.e., $\mathbb{P}(100 < T \leq 101 | T > 100)$. Using the definition of conditional probability, we obtain

$$\mathbb{P}(100 < T \leq 101 | T > 100) = \frac{\mathbb{P}(100 < T \leq 101 \cap T > 100)}{\mathbb{P}(T > 100)}. \qquad \textbf{(5 pts)}$$

Since $(100 < T \leq 101) \cap (T > 100) = (100 < T \leq 101)$, this yields

$$\mathbb{P}(100 < T \leq 101 | T > 100) = \frac{\mathbb{P}(100 < T \leq 101)}{1 - \mathbb{P}(T \leq 100)}. \qquad \textbf{(5 pts)}$$

Finally, since T has the lognormal distribution with the parameters determined above, we obtain

$$\mathbb{P}(100 < T \leq 101 | T > 100) = \frac{\Phi\left(\frac{\ln 101 - \mu_X}{\sigma_X}\right) - \Phi\left(\frac{\ln 100 - \mu_X}{\sigma_X}\right)}{1 - \Phi\left(\frac{\ln 100 - \mu_X}{\sigma_X}\right)} \approx \frac{\Phi(0.798) - \Phi(0.772)}{1 - \Phi(0.772)}$$

$$\approx \frac{0.788 - 0.779}{1 - 0.779} \approx 0.04. \qquad \textbf{(10 pts)}$$

3. (**40 points**) The purchasing manager at a firm is tasked with buying equipment produced either by company A or by company B. Her decision is determined solely by the average daily cost of operating the equipment. *The manager has only the following information on which to base her decision.* For the equipment produced by company A, the number of repairs (X) during each day of operation is a random variable with mean $\mu_X = 0.96$. The daily cost of operating this equipment is $C_{\rm A} = 160+40X^2$. For the equipment produced by company B, the number of daily repairs (Y) is a random variable with mean $\mu_Y = 1.12$. The daily cost of operating this equipment is $C_{\rm B} = 128 + 40Y^2$.

   (i) Which probabilistic model describes random variables X and Y? Explain your choice. **(10 pts)**

   (ii) Which company's equipment has the lower average daily cost? **(30 pts)**

   Hint: Assume that the repairs take negligible time and each night the machine are cleaned so that they operate like new machine at the start of each day.

Solution:

(i) Since the random variables involved are discrete (the number of daily repairs), one has to chose from a set of discrete distributions. Since the only available information about these discrete random variables is their respective means, one has to choose *a Poisson distribution*,

$$p_X(x) = e^{-\mu_X} \frac{\mu_X^x}{x!} \qquad \text{and} \qquad p_Y(y) = e^{-\mu_Y} \frac{\mu_Y^y}{y!}. \qquad \textbf{(10 pts)}$$

(ii) We are asked to compute the averages of the daily costs $C_{\rm A}$ and $C_{\rm B}$. These are computed as follows

$$\mathbb{E}(C_{\rm A}) = 160 + 40\,\mathbb{E}(X^2) = 160 + 40\,(\sigma_X^2 + \mu_X^2). \qquad \textbf{(10 pts)}$$

We know that the variance of a Poisson distribution equals its mean, $\sigma_X^2 = \mu_X$. Hence

$$\mathbb{E}(C_{\rm A}) = 160 + 40\,(0.96 + 0.96^2) = 235.264. \qquad \textbf{(10 pts)}$$

Likewise,

$$\mathbb{E}(C_{\rm B}) = 128 + 40\,(\sigma_Y^2 + \mu_Y^2) = 222.976. \qquad \textbf{(10 pts)}$$

Thus, the equipment produced by company B has the lower average daily cost.