IE 529 - Homework 1.1 Solutions: Prob and Stats Review

I. Warm-up:
Consider three events, A, B, and C satisfying $A \cup B \cup C = \mathcal{S}$ (where $\mathcal{S}$ represents the entire sample space), $P(A) = .3$, $P(B) = .3$ and $P(C) = .5$. Assume events B and C are mutually exclusive.

True or False:

a. B and C are independent events.

False.

b. $P(B \cap C) = 0$.

True.

c. $P(B \cup C) = .8$.

True.

d. $P((A \cup B) \cap (A \cup C)) = 0$.

False, $P((A \cup B) \cap (A \cup C)) = P(A) \neq 0$.

e. Events A, B and C are mutually exclusive.

False.

f. $P(A \cup B \cup C) = 1$.

True, $P(A \cup B \cup C) = P(\mathcal{S}) = 1$.

g. Given $P(A \cup B) = .5$, then $P(A \cap B) = .06$.

False.

h. Given $P(A \cap B) = .06$ holds, then A and B are independent events.

False.

i. Given $P(A \cap B) = .06$ holds, then A^C and B are independent events.

False.

j. Given $P(A \cap B) = .06$ holds, then $P(A|B) = .2$.

True.

II. Basic calculations:
Given the following probability density function for a continuous random variable, y,

$$f(y) = (1/3)e^{\frac{-y}{3}}, \text{ for } y \geq 0.$$

a. Calculate the **mean** of the distribution, $E(y)$.

Note, this is an exponential random variable $Exp(\lambda)$, with $\lambda = 1/3$. Therefore $E(y) = \frac{1}{\lambda} = 3$.

Alternatively,

$$E(y) = \int_0^\infty y((1/3)e^{\frac{-y}{3}})dy = -e^{\frac{-y}{3}}(y+3)\Big|_0^\infty = 3.$$

b. Calculate the **probability** that $y > 3$.

$$P(y > 3) = \int_3^\infty (1/3)e^{\frac{-y}{3}}dy = -e^{\frac{-y}{3}}\Big|_3^\infty = \frac{1}{e} \approx 0.368.$$

c. Find the value m such that $P(y \leq m) = 0.5$. (Note: m is the distribution median).

$$P(y \leq m) = 1 - \left(-e^{\frac{-y}{3}}\Big|_m^\infty\right) = .5 \quad \Rightarrow \quad m = -3ln(.5) = 3ln(2) \approx 2.0794$$

d. Compare and comment on your answers to parts (a), (b) and (c) above. (Hint: a sketch of $f(y)$ versus y might be helpful).

The distribution is non-symmetric, skewed to the left. Thus the median is lower than the mean.

III. Conditional probabilities:
An analytical method to detect three different types of pollutants (organic contaminants, volatile solvents, and chlorinated compounds) in water was tested on a large batch of water samples. From the tests it was found that 60% of the samples were polluted with organic contaminants (OC), 27% were polluted with volatile solvents (VS), and 13% were polluted with chlorinated compounds (C).

So $P(OC) = 0.6$, $P(VS) = 0.27$, and $P(C) = 0.13$.

The test is known to have high accuracy and will provide a *positive* signal (P) with probability .997 if there are high levels of organic contaminants, with probability .995 if there are high levels of volatile solvents, and with probability .897 if there are high levels of chlorinated compounds. If no pollutant is present, the test will not signal.

So $P(P|OC) = 0.997$, $P(P|VS) = 0.995$, and $P(P|C) = 0.897$.

a. If a test sample is selected randomly, what is the probability that it will give a positive signal?

$$P(P) = P(P|OC)P(OC) + P(P|VS)P(VS) + P(P|C)P(C) \approx 0.984$$

b. If a test signals positively, what is the probability that chlorinated compounds are present at high levels?

$$P(C|P) = \frac{P(C)P(P|C)}{P(P)} \approx 0.1185.$$

c. If a test signals positively, what is the probability that volatile solvents are present at high levels?

$$P(VS|P) = \frac{P(VS)P(P|VS)}{P(P)} \approx 0.273.$$

d. Suppose a consumer is concerned about a combination of volatile solvents and organic contaminants in their water, and has been told the actual probability of both being present is 0.162. Does this imply the presence of these pollutants is in fact independent? (Justify)

$$P(VS \cap OC) = 0.162 = P(VS)P(OC) \quad \Rightarrow \quad \mathit{Independence}.$$

IV. Bivariate distributions:
The elapsed times taken by each of two construction workers to complete his/her assigned jobs on a site are denoted by the continuous random variables, x and y, respectively. A joint probability density function, $f(x, y)$, is given to model these times, which is defined by

$$f(x,y) = \begin{cases} c(x+y), & 1 < x < 4 \text{ and } 1 < y < x \\ 0 & \text{elsewhere,} \end{cases}$$

where c is a constant.

a. (4pts) Calculate the value of c that makes $f(x, y)$ a proper bivariate density function.

$$1 = c\int_1^4\int_1^x (x+y)dydx = c\int_1^4 (xy+\frac{1}{2}y^2)\Big|_1^x dx = c\int_1^4 \frac{3}{2}x^2-x-\frac{1}{2}dx = \frac{c}{2}(x^3-x^2-x)\Big|_1^4 = c\frac{45}{2}.$$

Therefore $c = \frac{2}{45}$.

b. (8pts) Find the **marginal** probability distributions $f_X(x)$ and $f_Y(y)$.

$$f_X(x) = \int_1^x \frac{2}{45}(x+y)dy = \frac{1}{45}(3x^2-2x-1),$$

$$f_Y(y) = \int_y^4 \frac{2}{45}(x+y)dx = \frac{1}{45}(x^2+2xy)\Big|_y^4 = \frac{1}{45}(16+8y-3y^2).$$

c. (4pts) Are x and y independent? Why or why not?

$$f_{XY}(x) \neq f_X(x)f_Y(y) \quad \Rightarrow \quad \textit{Not Independent.}$$

d. (10pts) Suppose the proportion of "dead time" (i.e., the time when no assigned jobs are begin performed by either worker) is given by the new random variable Z, defined by

$$z = 1 - \frac{c(x+y)}{4},$$

where c is the constant determined in part **a.** Find the expected value of Z, i.e., $E(z)$.

$$Z = 1 - \frac{(X+Y)}{90},$$

$$\Rightarrow E(Z) = E(1) - \frac{1}{90}E(X) - \frac{1}{90}E(Y).$$

$$E(X) = \frac{1}{45}\int_1^4 (3x^2-2x-1) = \frac{63}{20},$$

$$E(Y) = \frac{1}{45}\int_1^4 (16+8y-3y^2) = \frac{43}{20}.$$

Therefore $E(Z) = 1 - \frac{63}{180} - \frac{43}{180} \approx 0.9411$.

V. More fundamental calculations:
Suppose the **times** it takes randomly selected students to solve one particular problem may be viewed as independent, identically distributed, continuous random variables. Specifically, suppose the time it takes each student is denoted by t_k for $k = 1, 2, \ldots$ and is described by the exponential distribution

$$f(t_k) = \begin{cases} \lambda e^{-\lambda t_k} & t_k \geq 0 \\ 0 & t_k < 0 \end{cases}$$

Compute the probability that *student 1* takes twice as long, or longer, to solve the problem as *student 2*.

$$\begin{aligned}
P(2t_2 \leq t_1) &= \int_0^\infty P(2t_2 \leq t_1) f_{t_1}(t_1) dt_1 \\
&= \int_0^\infty P(t_2 \leq \frac{1}{2} t_1) f_{t_1}(t_1) dt_1 \\
&= \int_0^\infty (1 - e^{-\lambda(\frac{1}{2}t_1)}) \lambda e^{-\lambda t_1} dt_1 \\
&= -e^{-\lambda t_1} + \frac{2}{3} e^{-\frac{3}{2}\lambda t_1} \Big|_0^\infty \\
&= 1 - \frac{2}{3} \\
&= \frac{1}{3}.
\end{aligned}$$

Or alternatively:

$$\begin{aligned}
P(t_1 \geq 2t_2) &= \int_0^\infty \int_{2t_2}^\infty f(t_1) f(t_2) dt_1 dt_2 \\
&= \int_0^\infty \int_{2t_2}^\infty \lambda^2 e^{-\lambda t_1} e^{-\lambda t_2} dt_1 dt_2 \\
&= \int_0^\infty \lambda e^{-\lambda t_2} (-e^{-\lambda t_1}) \Big|_{2t_2}^\infty dt_2 \\
&= \int_0^\infty \lambda e^{-\lambda t_2} (e^{-2\lambda t_2}) dt_2 \\
&= \int_0^\infty \lambda e^{-3\lambda t_2} dt_2 \\
&= -\frac{1}{3} e^{-3\lambda t_2} \Big|_0^\infty \\
&= \frac{1}{3}.
\end{aligned}$$