STA371G Homework Assignment 1

(40 Points. Due in class on Wednesday, 02/04/2015. Group homework.)
Please write down the name and EID of each group member. Each group consists of up to three members.

Problem 1 (10 points)

A construction company has to complete a project no later than three months from now or there will be significant cost overruns. The manager of the construction company believes that there are four possible values for the random variable X, the number of months from now it will take to complete the project: 2, 2.5, 3 and 3.5. The manager currently thinks that the probabilities for these four possibilities are in the ratio of 1 to 3 to 4 to 2. That is to say, $X = 2.5$ is three times more likely than $X = 2$.

(a) Find the probability distribution of X.

The random variable X has four possible random outcomes: 2, 2.5, 3 and 3.5. Since the summation of the probabilities of all possible outcomes must be equal to one, we have

$$P(X = 2) + P(X = 2.5) + P(X = 3) + P(X = 3.5) = 1.$$

Since $P(X = 2.5) = 3P(X = 2)$, $P(X = 3) = 4P(X = 2)$ and $P(X = 3.5) = 2P(X = 2)$, we have

$$P(X = 2) + 3P(X = 2) + 4P(X = 2) + 2P(X = 2) = 10P(X = 2) = 1.$$

Therefore

$$P(X = 2) = 0.1$$
$$P(X = 2.5) = 0.3$$
$$P(X = 3) = 0.4$$
$$P(X = 3.5) = 0.2$$

(b) What is the probability that this project will be completed in no more than three months from now?

$$P(X \leq 3) = P(X = 2) + P(X = 2.5) + P(X = 3) = 0.8$$

(c) What is the expected completion time of this project from now?

The expected completion time of the project is also the mean of the random variable, which can be calculated as

$$
\begin{aligned}
E(X) &= 2P(X=2) + 2.5P(X=2.5) + 3P(X=3) + 3.5P(X=3.5) \\
&= 2 \times 0.1 + 2.5 \times 0.3 + 3 \times 0.4 + 3.5 \times 0.2 \\
&= 2.85
\end{aligned}
$$

(d) How much variability exists around the expected completion time? (Hint: calculate the variance/standard deviation)

To measure the variability around $E(X)$, we use the variance $Var(X)$ or standard deviation $sd(X)$. The variance is

$$
\begin{aligned}
Var(X) =& (2 - E(X))^2 P(X=2) + (2.5 - E(X))^2 P(X=2.5) \\
&+ (3 - E(X))^2 P(X=3) + (3.5 - E(X))^2 P(X=3.5) \\
=& (2 - 2.85)^2 \times 0.1 + (2.5 - 2.85)^2 \times 0.3 \\
&+ (3 - 2.85)^2 \times 0.4 + (3.5 - 2.85)^2 \times 0.2 \\
=& 0.2025
\end{aligned}
$$

The standard deviation is

$$sd(X) = \sqrt{Var(X)} = \sqrt{0.2025} = 0.45$$

Note that we can also calculate the variance as

$$
\begin{aligned}
Var(X) =& E(X^2) - (E(X))^2 \\
=& (2)^2 P(X=2) + (2.5)^2 P(X=2.5) \\
&+ (3)^2 P(X=3) + (3.5)^2 P(X=3.5) - 2.85^2 \\
=& 2^2 \times 0.1 + 2.5^2 \times 0.3 \\
&+ 3^2 \times 0.4 + 3.5^2 \times 0.2 - 2.85^2 \\
=& 0.2025
\end{aligned}
$$

(e) The project requires a special equipment that the company has already paid 3550 dollars to rent. To continue renting this equipment until the end of the project, for each additional half month, the company will have to pay a rental fee of 500 dollars. What is the expected total rental cost for this equipment? How much variability exists around the expected total rental cost?

Denote Y as the total rental cost. We have

$$Y = 3550 + 500 \times 2 \times X = 3550 + 1000X.$$

The expected total rental cost is

$$E(Y) = 3550 + 1000 \times E(X) = 3550 + 1000 \times 2.85 = 6400 \text{ dollars}.$$

The variance of Y is

$$Var(Y) = 1000^2 \times Var(X) = 1000^2 \times 0.2025 = 2.025 \times 10^5.$$

The standard deviation of Y is

$$sd(Y) = \sqrt{Var(Y)} = 450.$$

Problem 2 (10 points)

Suppose that Jennifer has a 90% probability to eat breakfast and David has a 30% probability to eat breakfast. Moreover, assume whether Jennifer eats breakfast is not related to whether David eats breakfast.

(a) What's the probability that both Jennifer and David eat breakfast tomorrow morning?

$$90\% \times 30\% = 27\%$$

(b) What's the probability that Jennifer eats breakfast or David eats breakfast tomorrow morning?

$$90\% + 30\% - 27\% = 93\%$$

(c) What's the probability that either Jennifer or David (not both of them) eats breakfast tomorrow morning?

$$90\% \times (1 - 30\%) + (1 - 90\%) \times 30\% = 66\%$$

Problem 3 (10 points)

Suppose that 60% of the students at UT are football fans, 10% of the students at UT are from McCombs Business School, and 80% of the McCombs students are football fans.

(a) (3 points) If we choose a student at random from UT, what is the probability that this student is a football fan from McCombs Business School?

Let $F = 1$ if the student is a football fan and $F = 0$ otherwise, and $M = 1$ if the student is from McCombs and $M = 0$ otherwise. Since $P(F = 1) = 0.60$, $P(M = 1) = 0.10$ and $P(F = 1|M = 1) = 0.80$, we have

$$P(F = 1, M = 1) = P(M = 1)P(F = 1|M = 1) = 0.10 \times 0.80 = 8\%$$

(b) (3 points) If we choose a student at random from UT, what is the probability that this student is from McCombs Business School but *not* a football fan?

$$P(F = 0, M = 1) = P(M = 1) - P(F = 1, M = 1) = 0.10 - 0.08 = 2\%$$

(c) (3 points) If we choose a student at random from UT, what is the probability that this student is neither a football fan nor from McCombs Business School?

$$P(F = 0, M = 0) = P(F = 0) - P(F = 0, M = 1) = (1 - 0.6) - 0.02 = 38\%$$

(d) (3 points) If we choose a football fan student at random from UT, what is the probability that this football fan student is from McCombs Business School?

$$P(M = 1|F = 1) = \frac{P(F = 1, M = 1)}{P(F = 1)} = \frac{0.08}{0.60} = 13.3\%$$

(e) (3 points) If we choose a student who is *not* from McCombs Business School at random, what is the probability for this non-McCombs student to be a football fan?

$$P(F = 1|M = 0) = \frac{P(F = 1, M = 0)}{P(M = 0)} = \frac{P(F = 1) - P(F = 1, M = 1)}{1 - P(M = 1)} = \frac{0.6 - 0.08}{1 - 0.1} = 57.8\%$$

Comment: one easy way to do this kind of problem is using a table shown below.

	McCombs (M=1)	Not McCombs (M=0)	
Football Fan (F=1)	0.08	0.52	0.60
Not Football Fan (F=0)	0.02	0.38	0.40
	0.10	0.90	1

Problem 4

A company ships its products from its warehouse in Newark, NJ to Austin, TX by air. If the shipment arrives within 24 hours, it is considered on time. Let Y be a random variable that $Y = 1$ if the shipment arrives within 24 hours and $Y = 0$ if the shipment gets delayed.

Based on the company's historical data, if the weather conditions in both Newark and Austin are normal, then the shipment arrives on time with a 95% probability; if either Newark or Austin has adverse weather conditions, then the shipment arrives on time with a 40% probability; and if both Newark and Austin have adverse weather conditions, then the shipment gets delayed with a 95% probability.

Let X be a random variable that $X = 0$ if neither Newark nor Austin has adverse weather conditions, $X = 1$ if either Newark or Austin has adverse weather conditions, and $X = 2$ if both Newark and Austin have adverse weather conditions.

(a) If the company ships its product on a day when both Newark and Austin have adverse weather conditions, then what is the probability of $Y = 1$, i.e., the shipment arrives within 24 hours?

Since $P(Y = 0|X = 2) = 95\%$, the probability of $Y = 1$ conditioning on $X = 2$ (both Newark and Austin have adverse weather conditions) is

$$P(Y = 1|X = 2) = 1 - P(Y = 0|X = 2) = 1 - 0.95 = 0.05$$

(b) Suppose it is forecasted that Newark and Austin have adverse weather conditions with the probabilities of 60% and 10%, respectively, on February 4. We further assume that the weather conditions of these two cities are independent given the forecast. According to the forecast, what is $P(X=2)$, i.e., the probability that both Newark and Austin have adverse weather conditions on February 4? How about $P(X=0)$, i.e., the probability that neither Newark nor Austin has adverse weather conditions on February 4? How about $P(X=1)$? (Hint: using the joint probability multiplication rule for independent random variables.)

Since the whether conditions of Newark and Austin are assumed to be independent given the forecast, using the multiplication rule for the joint probability of independent random variables, we have

$$P(X=2) = 0.6 \times 0.1 = 0.06$$
$$P(X=0) = (1-0.6) \times (1-0.1) = 0.36$$
$$P(X=1) = 0.6 \times (1-0.1) + (1-0.6) \times 0.1 = 0.58$$

We can also calculate $P(X=1)$ with

$$P(X=1) = 1 - P(X=2) - P(X=0) = 0.58$$

(c) Based on the information provided in **(b)**, if the company ships its products on February 4, what is $P(Y=1)$, i.e., the probability that the shipment arrives on time? (Hint: first find out all $P(Y=1|X=x)$ and all $P(X=x)$, then ....)

Using the relationship between the marginal and joint probabilities, we have

$$P(Y=1) = P(Y=1, X=0) + P(Y=1, X=1) + P(Y=1, X=2)$$

Using the relationship between the joint and conditional probabilities, we further have

$$\begin{aligned}
P(Y=1) =& P(Y=1, X=0) + P(Y=1, X=1) + P(Y=1, X=2) \\
=& P(Y=1|X=0)P(X=0) \\
&+ P(Y=1|X=1)P(X=1) \\
&+ P(Y=1|X=2)P(X=2) \\
=& 0.95 \times 0.36 + 0.4 \times 0.58 + 0.05 \times 0.06 \\
=& 0.577
\end{aligned}$$

Therefore, based on both the company's historical data and the weather forecast, we predict that there is a 57.7% chance that the shipment sent out on February 4 arrives on time.

Comment: one easy way to do this kind of problem is using a table shown below.

	Delayed (Y=0)	On time (Y=1)	
Neither bad (X=0)		0.95*0.36	0.36
Either bad (X=1)		0.40*0.58	0.58
Both bad (X=2)	0.95*0.06		0.06
			1