Homework Assignment 1

(Due in class on Thursday, 01/23/2014)

STA 371G, Statistics and Modeling, Spring 2014

Problem 1

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

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

(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

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

Problem 2

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 January 28, 2014. 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 January 28, 2014? How about $P(X = 0)$, i.e., the probability that neither Newark nor Austin has adverse weather conditions on January 28, 2014? 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 January 28, 2014, 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 January 28, 2014 arrives on time.