<u>**ST3009 - Statistical Methods for Computer Science**</u>
<u>**Week 5 Questions**</u>

Brandon Dooley - #16327446

***Question 1 –** A box contains 5 red and 5 blue marbles. Two marbles are withdrawn randomly. If they are the same color, then you win \$1.10; if they are different colors, then you lose \$1.00. Calculate:*

(a) Let the event Y be the event that the marbles are the same colour. The probability of Y occurring is the probability of pulling one colour, and then the same colour again on the next selection which is as follows:

$$P(Y) = \frac{\binom{2}{1} * \binom{5}{2}}{\binom{10}{2}} = \frac{4}{9}$$

Using this we can then calculate the expected value of winnings W:

$$E[W] = 1.10 * \frac{4}{9} + \left(-1.00 * \left(1 - \frac{4}{9}\right)\right) = -\frac{1}{15} = -0.0667$$

(b) For a discrete random variable an expression for variance is as follows:

$$Var(X) = E[X^2] - (E[X])^2$$

Let X be our winnings i.e X = W. Therefore using what we calculated as the expected value in part (a) the variance on the amount you win can be calculated as follows:

$$Var(W) = E[W^2] - (E[W])^2$$

Hence, we can calculate:

$$E[W^2] = (1.10)^2 * \frac{4}{9} + \left((-1.00)^2 * \left(1 - \frac{4}{9}\right)\right) = \frac{82}{75}$$

And using this value we can then calculate Var(W):

$$Var(W) = \frac{82}{75} - \left(-\frac{1}{15}\right)^2 = \frac{49}{45} = 1.089$$

***Question 2 –** Suppose you carry out a poll following an election. You do this by selecting n people uniformly at random and asking whether they voted or not, letting Xi = 1 if person i voted and Xi = 0 otherwise. Suppose the probability that a person voted is 0.6.*

(a) $E[X_i]$ is the expected value of asking person I whether they voted or not. This can be calculated as follows. X_i is an indicator variable for event E (person has voted) i.e $X_i = 1$ if event E occurs and $X_i = 0$ otherwise. Therefore we can calculate the expected value as:

$$E[X_i] = 1 * P(E) + 0 * \left(1 - P(E)\right) = P(E) = 0.6$$

$$E[X_i] = (1 * 0.6) + 0 * (1 - 0.6) = P(E) = 0.6$$

Similarly to question 1 (b) we can calculate the variance of X_i using the equation:

$$Var(X_i) = E\left[{X_i}^2\right] - (E[X_i])^2$$

First we must calculate $E\left[{X_i}^2\right]$:

$$E\left[{X_i}^2\right] = 1^2 * 0.6 + 0^2 * (1 - 0.6) = 0.6$$

Using this we can then calculate $Var(X_i)$:

$$Var(X_i) = E\left[{X_i}^2\right] - (E[X_i])^2$$

$$Var(X_i) = 0.6 - (0.6)^2 = 0.24$$

(b) $E[Y]$ is the expected value of the number of people that did vote when n people are uniformly selected at random. This can also be written in terms of X_i as follows (using the linearity of the expected value to move the E[] inside the sum).

$$E[Y] = E\left[\sum_{i=1}^{n} X_i\right] = \sum_{i=1}^{n} E[X_i] = n * E[X_i] = \ n * 0.6$$

Using the fact that all people are sampled independently we can derive for all i:

$$E[X_i] = E[X] = 0.6$$

Thus $E[Y]$ is clearly not the same as $E[X]$ as shown and also because as you were to survey more than one person the expected value would change.

(c) $E\left[\frac{1}{n}Y\right]$ is the expected value for one n^th of the survey sample. Using the equation derived in part (a) and the linearity of the expectation this can be written as:

$$E[Y] = n * E[X_i]$$

$$E\left[\frac{1}{n}Y\right] = \frac{1}{n}E[Y] = \frac{1}{n} * n * 0.6 = 0.6$$

From this we can infer that $E\left[\frac{1}{n}Y\right] = E[X]$ as it is the same as selecting one person out of a sample of n.

(d) We can calculate the variance of $\frac{1}{n}Y$ using the equation:

$$Var(X_i) = E\left[{X_i}^2\right] - (E[X_i])^2$$

$$Var\left(\frac{1}{n}Y\right) = \frac{1}{n^2}Var(Y) = \frac{1}{n^2}Var(\sum_{i=1}^{n} X_i)$$

Since people are sampled independently we can infer:

$$Var\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} Var(X_i) = n * Var(X_i) = Var(X)$$

Hence:

$$Var\left(\frac{1}{n}Y\right) = \frac{1}{n^2}Var(X)$$

***Question 3 –** Suppose that 2 balls are chosen without replacement from an urn consisting of 5 white and 8 red balls. Let Xi equal 1 if the i'th ball selected is white, and let it equal 0 otherwise.*

(a) Since X_i can only take the values of 0 and one the joint probability mass function of X_1 and X_2 would look like the following:

	$X_1 = 0$	$X_1 = 1$	P(Y=y)
$X_2 = 0$	0.3589	0.2564	0.6153
$X_2 = 1$	0.2564	0.1282	0.3846
P(X=x)	0.6153	0.3846	1

$$P(X_1 = 0 \cap X_2 = 0) = \frac{8}{13} * \frac{7}{12} = 0.3589$$
$$P(X_1 = 0 \cap X_2 = 1) = \frac{8}{13} * \frac{5}{12} = 0.2564$$
$$P(X_1 = 1 \cap X_2 = 0) = \frac{5}{13} * \frac{8}{12} = 0.2564$$
$$P(X_1 = 1 \cap X_2 = 1) = \frac{5}{13} * \frac{4}{12} = 0.1282$$

(b) For the events to be independent they must satisfy the following formula:

$$P(X_1 \cap X_2) = P(X_1) * P(X_2)$$

$$0.1282 = 0.3846 * 0.3846$$

$$0.1282 \neq 0.1479$$

Therefore we can determine that the events X_1 and X_2 are not independent as they do not satisfy the formal definition of independence.

(c) The expected value of X_2 can be calculated as follows:

$$E[X_2] = 1 * P(X_2) + 0 * \left(1 - P(X_2)\right)$$

$$E[X_2] = 1 * 0.3846 + 0 * (1 - 0.3846)$$

$$E[X_2] = 0.3846$$

(d) For two discrete random variables X and Y the conditional expectation of X given. Y = y is as follows:

$$E[X \mid Y = y] = \sum_{x} xP(X = x|Y = y)$$

Therefore, given that $P(X_2 = 1) = 0.3846$:

$$E[X_1 \mid X_2 = 1] = 0 * \left(\frac{0.2564}{0.3846}\right) + 1 * \left(\frac{0.1282}{0.3846}\right)$$