Practice Exam: Probability Part

Master Statistics and Data Science

Instructor: Roula Tsonaka

1. The random variables X and Y take on the values 1, 2, and 3, as indicated in the following table:

	X = 1	X = 2	X = 3
Y = 1	2/36	2/36	3/36
Y = 2	1/36	10/36	3/36
Y = 3	4/36	5/36	6/36

(*a*) Compute the marginal frequency function or probability mass function for X.

(*b*) Give the conditional frequency function for X given that $Y = 2$.

(*c*) Calculate $E(X \mid Y = 2)$ and $E[E(X \mid Y)]$.

(*d*) Find the probability $P(Y > 1 \mid X \leq 1)$.

(*e*) Compute $\text{Var}(2 \cdot X + 5)$.

(*d*) Compute $\text{Cov}(X, Y)$.

Solution:

(*a*) The marginal frequency function is:

$$P(X = 1) = \frac{2}{36} + \frac{1}{36} + \frac{4}{36} = 7/36$$
$$P(X = 2) = \frac{2}{36} + \frac{10}{36} + \frac{5}{36} = 17/36$$
$$P(X = 3) = \frac{3}{36} + \frac{3}{36} + \frac{6}{36} = 12/36.$$

(b) The conditional frequency function for X given that $Y = 2$ is:

$$P(X = 1 \mid Y = 2) = \frac{P(X = 1, Y = 2)}{P(Y = 2)} = \frac{1/36}{14/36} = 1/14$$
$$P(X = 2 \mid Y = 2) = \frac{P(X = 2, Y = 2)}{P(Y = 2)} = \frac{10/36}{14/36} = \frac{10}{14}$$
$$P(X = 3 \mid Y = 2) = \frac{P(X = 2, Y = 2)}{P(Y = 2)} = \frac{3/36}{14/36} = 3/14.$$

(c) We can compute $E(X \mid Y = 2)$ as follows:

$$\begin{aligned} E(X \mid Y = 2) &= \sum_{x=1}^{3} x \times p(x \mid Y = 2) \\ &= 1 \times 1/14 + 2 \times 10/14 + 3 \times 3/14 = 30/14. \end{aligned}$$

We know from the law of total expectation that: $E[E(X \mid Y)] = E(X)$. Thus, $E(X) = \sum_1^3 x \times p_x(x) = 1 \times 7/36 + 2 \times 17/36 + 3 \times 12/36 = 77/36$.

(d) Using the definition of conditional probability:

$$\begin{aligned} P(Y > 1 \mid X \leq 1) &= \frac{P(Y = 2, X = 1) + P(Y = 3, X = 1)}{P(X = 1)} \\ &= \frac{1/36 + 4/36}{7/36} = 5/7. \end{aligned}$$

(e) Using the properties of the variance:

$$\text{Var}(2 \cdot X + 5) = 4 \times \text{Var}(X).$$

We have:

$$\begin{aligned} \text{Var}(X) &= E(X^2) - [E(X)]^2 \\ &= \sum_{x=1}^{3} x^2 \times p(x) - (77/36)^2 \\ &= 1^2 \times 7/36 + 2^2 \times 17/36 + 3^2 \times 12/36 - (77/36)^2 \\ &= 7/36 + 68/36 + 108/36 - (7/36)^2 = 183/36 - (77/36)^2 = 0.5084. \end{aligned}$$

Thus $\text{Var}(2 \cdot X + 5) = 4 \times \text{Var}(X) = = 2.033951$.

(*d*) Using the properties of the variance:

$$\text{Cov}(X,Y) = E(X \times Y) - E(X) \times E(Y)$$

We have $E(X) = 77/36$. Then:

$$E(Y) = \sum_{y=1}^{3} 1 \times 7/36 + 2 \times 14/36 + 3 \times 15/36 = 80/36.$$

Finally,

$$\begin{aligned} E(X \times Y) &= \sum_{x=1}^{3}\sum_{y=1}^{3} x \times y \times p(x,y) \\ &= 1 \times 1 \times 2/36 + 1 \times 2 \times 1/36 + 1 \times 3 \times 4/36 \\ &+ 2 \times 1 \times 2/36 + 2 \times 2 \times 10/36 + 2 \times 3 \times 5/36 \\ &+ 3 \times 1 \times 3/36 + 3 \times 2 \times 3/36 + 3 \times 3 \times 6/36 \\ &= 2/36 + 2/36 + 12/36 + 4/36 + 40/36 + 30/36 + 9/36 + 18/36 + 54/36 \\ &= 171/36 = 4.75. \end{aligned}$$

So we can compute the covariance as:

$$\begin{aligned} \text{Cov}(X,Y) &= E(X \times Y) - E(X) \times E(Y) \\ &= 171/36 - 77/36 \times 80/36 = -0.00308642. \end{aligned}$$

2. A new study is being designed to develop a human infection model that could be used for testing vaccines for SARS-CoV-2. Volunteers will be inoculated with a challenge virus and the chance to be infected with the virus is 70%. As the exact number of volunteers demonstrating mild clinical infection with SARS-CoV-2 is unknown, we will plan three different groups of volunteers to be inoculateds.

   I. three volunteers will be enrolled and inoculated with the virus.
   II. seven volunteers will be enrolled and inoculated with the virus.
   III. ten volunteers will be enrolled and inoculated with the virus.

   Answer to the following questions, assuming that the volunteers are independent:

   (*a*) What is the probability that none of the three volunteers in Group I will get infected?

(b) If after the inoculation of 10 volunteers (Group I and Group II), five volunteers get infected, what is the chance that 9 out of the 10 volunteers in Group III will get infected?

Solution:

(a) Let X the rv that counts the number infected. $X \sim \text{Binomial}(n, p = 0.7)$. The probability that none of the three volunteers in Group 1 will get infected is: `dbinom(0, size = 3, prob =0.7) = 0.027.`

(b) Let X_1 be the rv that counts the number infected in Group 1 and Group 2 and X_2 the rv that counts the number infected in Group 3. We have: $X_1 \sim \text{Binomial}(n = 10, p = 0.7)$ and $X_2 \sim \text{Binomial}(n = 10, p = 0.7)$. We need the conditional probability: $P(X_2 = 9 \mid X_1 = 5)$. Given that all volunteers are independent: $P(X_2 = 9 \mid X_1 = 5) = P(X_2 = 9) =$ `dbinom(9, size = 10, prob =0.7) = 0.1210608.`

3. Three employers at the Ministry of Education have the duty to archive the national examination papers written by the students in a cabinet. The three employers have different work contracts for the same job: Employer A has a 60% appointment, employer B has a 30% appointment and employer C has a 10% appointment. This means that employer A works 60% of the time on this task, employer B works 30% of the time on this task, etc. We assume that all three work with the same speed. One day the inspector picks a file at random and discovers that it has been misplaced. Given that employer A works most of the time on this task, he is blamed to be responsible for the mistake. In his defense his boss, states that employer A is the most careful of all and that last year he misplaced only 0.3% of the examination papers he was supposed to file. The corresponding percentages for employer B and employer C are: 0.7% and 1%, respectively. Given this new information, how likely is it that employer A has misplaced the file?

Solution:
Let A, B and C the events that employers A, B and C, respectively have archived the file and M the event that the file has been misplaced. Based on the description, we have $P(A) = 0.6$, $P(B) = 0.3$ and $P(C) = 0.1$. In addition, $P(M \mid A) = 0.003$, $P(M \mid B) = 0.007$ and $P(M \mid C) = 0.01$.

We want to compute the probability: $P(A \mid M)$. According to the Bayes' theorem we have:

$$\begin{aligned} P(A \mid M) &= \frac{P(A \cap M)}{P(M)} \\ &= \frac{P(M \mid A)P(A)}{P(M \mid A) \cdot P(A) + P(M \mid B) \cdot P(B) + P(M \mid C) \cdot P(C)} \\ &= \frac{0.003 \cdot 0.6}{0.003 \cdot 0.6 + 0.007 \cdot 0.4 + 0.01 \cdot 0.1} = 0.3673469. \end{aligned}$$

4. In the Van Gogh Museum in Amsterdam, an art exhibition will take place in March-May 2021, where the top three winners of the Fall 2020 competition will display their paintings. There are 11 artists who have each submitted a portfolio containing 7 paintings for the competition. Each of the top three winners will display in a row, four of their paintings (of their own choice) on one of the three walls that the museum manager will devote for this exhibition. For each wall, there are 31 different lighting options the artists can choose from to show off their work. How many different displays are possible?

   Solution:
   First we need to specify the different triplets of winners. Out of the 11 candidates there will be three winners and thus there are $11 \times 10 \times 9 = 990$ possible triplets. This is an example of sampling without replacement where order matters. Then each of the top three winners can select 4 out of his paintings to put in a row on one of the walls of the museum. Each artist can propose one of the: $7 \times 6 \times 5 \times 4 = 840$ possible arrangements. This is again an example of sampling without replacement where order matters. Finally, each of the three artists will select any of the 31 lighting options and thus there are $31 \times 31 \times 31 = 29791$ possible lighting arrangements. This is an example of sampling with replacement where order matters. In total using the multiplication principle, there are $(11 \times 10 \times 9) \times (7 \times 6 \times 5 \times 4)^3 \times 31^3$ possible displays.

5. One thousand cards are drawn (with replacement) from a standard deck of 52 playing cards. Let X be the random variable denoting the total number of aces drawn. Note that there are 4 aces in a standard deck.

   Answer the following questions:

(a) If you pick one card how likely is it that it is an ace?

(b) Give the frequency or probability mass function of the random variable X.

(c) We are interested in the probability that you get between 65 and 90 aces when you draw 1000 cards. Show the exact way to compute this probability. You do not need to give the end result.

(d) Compute approximately using the Central Limit Theorem the probability that you get between 65 and 90 aces when you draw 1000 cards.

(d) If the probability of selecting an ace when you draw one card is 10%, find c such that $P(X \leq c) \approx 0.1$?

<u>*Solution:*</u>

(a) Let Y the random variable that denotes if the drawn card is an ace. Then $P(Y) = 4/52 = 1/13$.

(b) The random variable X follows the Binomial distribution with $n = 1000$ and $p = 1/13$. The frequency function is:
$p_X(x) = P(X = x) = \left(\begin{array}{c} 1000 \\ x \end{array} \right) p^x(1-p)^{1000-x}$.

(c) We want to compute $P(65 \leq X \leq 90) = P(X = 65) + P(X = 66) + \ldots P(X = 90)$, where each term is computed using the frequency function of the Binomial distribution under (b).

(d) For the random variable X we have that $E(X) = 1000 \times 1/13. = 76.92308$ and $Var(X) = 1000 \times 1/13 \times 12/13 = 71.00592$.
According to CLT X is approximately normal with mean 76.92308 and variance 71.0059.

$$
\begin{aligned}
P(65 \leq X \leq 90) &= P(X \leq 90) - P(X \leq 65) \\
&= P(\frac{X - 76.92308}{\sqrt{71.0059}} \leq \frac{90 - 76.92308}{\sqrt{71.0059}}) - P(\frac{X - 76.92308}{\sqrt{71.0059}} \leq \frac{65 - 76.92308}{\sqrt{71.0059}}) \\
&= \Phi(1.55188) - \Phi(-1.41495) \\
&= 0.9394 - (1 - 0.9207)) = 0.8601.
\end{aligned}
$$

(d) We need to find c such that $P(X \leq c) = 0.1$. We have $E(X) = 1000 \times 0.1 = 100$. and $Var(X) = 1000 \times 0.1 \times 0.9. = 90$. Thus $\Phi(\frac{c-100}{\sqrt{90}}) = 0.10$ and $\frac{c-100}{\sqrt{90}} = \Phi^{-1}(0.10)$ where $\Phi^{-1}(0.10) = -1.29$ and $c = 87.76199$.

Motivate your answers!

Good luck!