DDA2001: Assignment 2

1. The assignment is due at Sunday **11:59 pm, March 06, 2022**.
2. Please submit your solution in **PDF** form. **Any other forms of solution will not be accepted and will be graded as 0.** Please leave enough time to make sure you have uploaded your solution as requirement before due.
3. If you submit the assignment late, you will get 0 for this assignment. **No excuses will be accepted for any late submission.**
4. Please make sure that your file could be downloaded successfully from BB after uploading your solution file.

---

1. **(4×2 points)** Suppose that a stock price is uniformly distributed on the interval $[10, 20]$. Suppose that you hold one unit of such a stock. Let X be the stock price.

   (a) Show the value of $E[X] - P(X \leq 13)$.

   (b) Show the value of $Var(X)$.

   Solution: Denote X as the price for one unit of stock.
   (a) Since $E(X) = \frac{10+20}{2} = 15$, and $P[X < 13] = \frac{13-10}{20-10} = \frac{3}{10}$,
   so we can get objective is $15 - \frac{3}{10} = 14\frac{7}{10}$
   (b) Since

   $$E(X^2) = \int_{10}^{20} \frac{1}{20-10} x^2 dx = \frac{1}{3} \frac{20-10}{20^3 - 10^3} = \frac{700}{3}.$$

   So $Var(X) = E(X^2) - E(X)^2 = \frac{700}{3} - 225 = \frac{25}{3}$. □

2. **(5×2 points)** Suppose that the amount of time one spends in a bank is *exponentially* distributed with mean equals to ten minutes.

   (a) What is the probability that a customer will spend more than fifteen minutes in the bank?

   (b*) What is the probability that a customer will spend more than fifteen minutes in the bank given that she is still in the bank after ten minutes?

   Solution: (a). If X represents the amount of time that the customer spends in the bank, given that $X \sim Expo(\lambda), \mathbb{E}(X) = \frac{1}{\lambda} = 10$, then the probability that a customer will spend more than fifteen minutes in the bank is:

   $$Pr(X > 15) = e^{-15\lambda} = e^{-1.5} \approx 0.220.$$

(b). Given that a customer is in the bank after ten minutes, the probability that she will spend more than fifteen minutes in the bank is:

$$\begin{aligned}
Pr(X > 5 + 10 | X > 10) &= \frac{Pr(X > 5 + 10, X > 10)}{Pr(X > 10)} \\
&= \frac{e^{-15\lambda}}{e^{-10\lambda}} \\
&= e^{-5\lambda} Pr(X > 5) \\
&= e^{-0.5} \\
&\approx 0.604.
\end{aligned}$$

□

3. **(6×3 points)** Imagine that a graduating SDS student is looking for a job. His/her likelihood of receiving an off-campus interview invitation for each job application is independent and identically distributed. Suppose that this likelihood only depends on how well he/she did in DDA2001. Specifically, with a grade A in DDA2001, the probability of obtaining an invitation for each application is $p_A = 0.95$, whereas with a grade C in DDA2001, the probability of obtaining an invitation for each application is $p_C = 0.15$.

   (a) Give the probability mass function (pmf) of the number of applications that a student has to submit such that he/she can get the first invitation. (Assume this student has a probability p of obtaining an invitation for each application. Express your answer in terms of p.)

   (b) On average, how many applications should an A student submit such that he can get an off-campus interview invitation? How about a C student?

   (c) Assuming that there are only N companies available for job application. Find the probability that an A student will not get any off-campus interview invitation from these N companies. Find the probability that a C student in DDA2001 will get exactly one invitation from these N companies.

Solution:

(a). $Pr(Y = k) = p(1-p)^{k-1}$.

(b). Y is a geometric random variable.

$$\begin{aligned}
E[Y] &= \sum_{n=1}^{\infty} np(1-p)^{n-1} \\
&= p \sum_{n=1}^{\infty} n(1-p)^{n-1} \\
&= \frac{p}{(1-(1-p))^2} \\
&= \frac{1}{p}.
\end{aligned}$$

For an A student, $E[Y] = 1.0526$, for a C student, $E[Y] = 6.6667$.

(c). $P(\text{an A student does not get an invitation in } N \text{ trials}) = \sum_{k=N+1}^{\infty} p(1-p)^{k-1} = (1-p)^N \sum_{k'=0}^{\infty} p(1-p)^{k'} = (1-p)^N = 0.05^N$
P(C gets an invitation in N trials) $= 0.15N \cdot 0.85^{N-1}$.

$\square$

4. **(5×2 points)** Suppose that the number of typographical errors on any single page of a book has a Poisson distribution with parameter $\lambda = 1$.

   (a) Calculate the probability that there is at least two errors on one specific page.

   (b) Calculate the expectation of the number of errors on 3 specific pages.

   Solution:

   (a). The probability that there is at least one error on this page is:

$$Pr(X \geq 1) = 1 - Pr(X = 0) - Pr(X = 1) = 1 - e^{-1} - e^{-1} \approx 0.264.$$

   (b). The expectation of the number of errors 3 specific pages is:

$$\begin{aligned}
3\mathbb{E}[X] &= 3\sum_{i=0}^{\infty} \frac{ie^{-\lambda}\lambda^i}{i!} \\
&= 3\sum_{i=1}^{\infty} \frac{e^{-\lambda}\lambda^i}{(i-1)!} \\
&= 3\lambda e^{-\lambda}\sum_{i=1}^{\infty} \frac{\lambda^{i-1}}{(i-1)!} \\
&= 3\lambda e^{-\lambda}\sum_{k=0}^{\infty} \frac{\lambda^k}{k!} \\
&= 3\lambda e^{-\lambda}e^{\lambda} \\
&= 3\lambda = 3.
\end{aligned}$$

$\square$

5. **(6×2 points)** Each of the members of a 7-judge panel independently makes a correct decision with probability 0.7. Suppose the panel's final decision is made by majority rule. Let X_i be the Bernoulli random variables for each member's judgement, where $X_i = 1$ indicates a correct decision and $X_i = 0$ indicates an incorrect decision.

   (a) What is the probability that the panel makes the correct decision?

   (b*) Given that $X_1 = X_2 = X_3 = X_4 \in \{0, 1\}$, show the probability that the panel made the correct decision.

   Solution: (a). Let C be the event that the jury makes the correct decision, then

$$\mathrm{P}(C) = \sum_{i=4}^{7} \binom{7}{i} (0.7)^i (0.3)^{7-i}$$

   (b). Let $F = [X_1 = X_2 = X_3 = X_4]$ be the event that four of the judges X_1, X_2, X_3 and X_4 agreed, then

$$\begin{aligned}
\mathrm{P}(C|F) =& \frac{\mathrm{P}(CF)}{\mathrm{P}(F)} \\
=& \frac{(0.7)^4}{(0.7)^4 + (0.3)^4} \approx 0.967.
\end{aligned}$$

6. **(4 points)** Entry to a certain University is determined by a national test. The scores on this test are normally distributed with a mean of 500 and a standard deviation of 100. Tom wants to be admitted to this university and he knows that he must score better than at least 70% of the students who took the test. Tom takes the test and scores 585. Will he be admitted to this university? (You can check the CDF of normal distribution from online sources to answer this question.)

   Solution: Let x be the random variable that represents the scores. x is normally ditsributed with a mean of 500 and a standard deviation of 100.

   The total area under the normal curve represents the total number of students who took the test.

   If we multiply the values of the areas under the curve by 100, we obtain percentages. For x = 585,

   $$z = (585 - 500)/100 = 0.85. \tag{1}$$

   The proportion P of students who scored below 585 is given by

   $$P(\text{area to the left of } z = 0.85) = 0.8023 = 80.23\% \tag{2}$$

   Tom scored better than 80.23% of the students who took the test and he will be admitted to this University.

   □

7. **(4 points)** Liam collected data on the sales of ice cream cones and air conditioners in his hometown. He found that when ice cream sales were low, air conditioner sales tended to be low and that when ice cream sales were high, air conditioner sales tended to be high. According to the discovery, Liam has the following conclusion:

   A. Sales of ice cream cones and air conditioner are positively correlated.

   B. Selling more ice cream cones causes more air conditioners to be sold.

   Show whether each conclusion is true?

   Solution: A □

8. **(5×3 points)** Justify the following questions, and if yes, prove it; if no, give one counterexample.

   (a) Does independence imply the uncorrelatedness?

   (b) Does uncorrelatedness imply independence?

   (c) Does correlation imply causality?

   Solution:

a) Yes. For convenience, we assume two discrete random variables X and Y and in fact the discussion about continuous random variables is similar. Then we have

$$\begin{aligned}\mathbb{E}(XY) &= \sum_{i,j} X_i Y_j \mathrm{P}(X_i Y_j) \\ &= \sum_{i,j} X_i Y_j \mathrm{P}(X_i)\mathrm{P}(Y_j) \\ &= \sum_{i} X_i \mathrm{P}(X_i) \sum_{j} Y_j \mathrm{P}(Y_j) \\ &= \mathbb{E}(X)\mathbb{E}(Y).\end{aligned}$$

b) No. If $X \sim \mathcal{N}(0,1)$ and $Y = X^2$, then $\mathrm{Cov}(X,Y) = \mathbb{E}(X^3) = 0$ but X and Y are dependent.

c) No. Sales of scarfs and ice cream are usually negatively correlated but there is no causality.

9. **(4 points)** A new credit card has been issued to 2000 customers. Of these customers, 1500 hold a Visa, 500 hold an AA card, and 40 hold a Visa and AA card. Find the probability that a random chosen customer holds an AA, given they hold a Visa.

Solution: The probability that a random chosen customer holds an AA, given they hold a Visa is:

$$\frac{40}{1500} = 0.0267.$$

□

10. **(4×3 points)** Company A is offering 40 software products and 30 hardware products. Company B is offering $x \in \{0, 1, \cdots, 50\}$ software products and $50-x$ hardware products.

If a product is selected at random, what is the probability that

(a) this product is a hardware product given that it is from company B? (in terms of x)

(b) this product is from company A given that it is a software product?

(c) For what values of x, will the probability in part (a) be greater than the probability in part (b)?

Solution: Let us arrange all the given information on a table as follows.

Table 1: Production information

	Software	Hardware	Total
Company A	40	30	70
Company B	x	50 - x	50
Total	40 + x	80 - x	120

(a) The total number of products is: $70 + 50 = 120$.

- Let event S: product selected is a software product.
- Let event H: product selected is a hardware product.

- Let event A: product selected is from company A.
- Let event B: product selected is from company B.

We are asked to find the conditional probability $P(H|B)$.

$$P(H|B) = \frac{P(H \cap B)}{P(B)} = \frac{\frac{50-x}{120}}{\frac{x+(50-x)}{120}} = \frac{50-x}{50}. \tag{3}$$

(b) We are asked to find the conditional probability $P(A|S)$.

$$P(A|S) = \frac{P(A \cap S)}{P(S)} = \frac{\frac{40}{120}}{\frac{40+x}{120}} = \frac{40}{40+x}. \tag{4}$$

(c) We need to solve the inequality as follows.

$$\frac{50-x}{50} > \frac{40}{40+x} \tag{5}$$

So, we can have

$$0 < x < 10. \tag{6}$$

x is a positive integer, hence

$$x = \{1, 2, 3, 4, 5, 6, 7, 8, 9\}. \tag{7}$$

□

11. **(3 points)** Linearity of expectation is the property that the expected value of the sum of random variables is equal to the sum of their individual expected values, regardless of whether they are independent. Why should you know this property? Show how useful it is by examples.

Solution: The hat-check problem can be easily solved with the linearity of expectation. (See example in lecture 4) □