Midterm Examination Solution

Name: ______________________

Student ID: ______________________

1. (18 marks) A standard 52-card deck comprises of 13 ranks in each of the four suits. Here ace is treated as rank 1. Alice randomly draws four cards from the deck at once without replacement. She is interested in the probability that the numbers on the four cards form a **consecutive** and **increasing** sequence (e.g. $(1,\ 2,\ 3,\ 4)$ and $(9,\ 10,\ J,\ Q)$ are what we want, but $(2,\ 3,\ 5,\ 6)$ or $(J,\ Q,\ K,\ 1)$ are not).

   (a) (6 marks) Define the sample space and calculate its size.

   (b) (6 marks) Define the event and calculate its size.

   (c) (6 marks) What is the probability that the numbers of four randomly drawn cards form a consecutive and increasing sequence?

   Solution:

   (a) The sample space is the combination of four cards from the 52 cards. The sample space consists of all $\binom{52}{4}$ outcomes.

   (b) The event is that numbers on four cards forms a consecutive sequence. Let E denote the event. We can write E as a disjoint union of $E_1, E_2, \ldots, E_{10}$ where E_c consists of those outcomes in which four cards form a consecutive sequence starting from rank c. Then E_c consists of $4 \times 4 \times 4 \times 4$ outcomes. The size of E is $\sum_{c=1}^{10} 4^4 = 2560$.

   (c) The probability that numbers on four cards form a consecutive sequence is that

   $$\mathrm{P}(E) = 2560/\binom{52}{4} \approx 0.0095$$

   .

2. (6 marks) It is estimated that 50% of emails are spam emails. Some software has been applied to filter these spam emails before they reach your inbox. A certain brand of software claims that it can detect 99% of spam emails, and the probability for a false positive (a non-spam email detected as spam) is 5%. Now if an email is detected as spam, then what is the probability that it is in fact a non-spam email?

Solution. Define events A = event that an email is detected as spam, B = event that an email is spam, B^c = event that an email is not spam.

We know $P(B) = P(B^c) = 0.5$, $P(A|B) = 0.99$, $P(A|B^c) = 0.05$.

Hence by the Bayes' formula we have

$$P(B^c|A) = \frac{P(A|B^c)P(B^c)}{P(A|B)P(B) + P(A|B^c)P(B^c)} = \frac{0.05 \times 0.5}{0.05 \times 0.5 + 0.99 \times 0.5} = \frac{5}{104}$$

3. (6 marks) For a family of three, the probability of contracting a certain infectious disease has the following law:

   P(The child has this disease)=0.6,
   P(The mother has this disease | The child has this disease)=0.5,
   P(The father has this disease | The child and mother both have this disease)=0.4.

   Find the probability that the mother and child both have this disease but the father does not.

   Solution. Define events, A = event that the child has this disease, B = event that the mother has this disease, C = event that the father has this disease.

   We know $P(A) = 0.6, P(B|A) = 0.5, P(C|BA) = 0.4$. Hence we have

$$\begin{aligned} P(ABC^c) &= P(C^c BA) = P(C^c \mid BA)P(BA) \\ &= P(C^c \mid BA)P(B \mid A)P(A) \\ &= (1 - P(C \mid BA))P(B \mid A)P(A) \\ &= 0.6 \times 0.5 \times 0.6 = 0.18 \end{aligned}$$

4. (10 marks) If A and B are any events in the sample space S. Prove that

$$P(A \cup B) = P(A) + P(B) - P(A \cap B).$$

   Solution. Since $A \cap \overline{B}$, $A \cap B$ and $\overline{A} \cap B$ are mutually exclusive events, thus

$$\begin{aligned} P(A \cup B) &= P\big[(A \cap B^c) \cup (A \cap B) \cup (A^c \cap B)\big] \\ &= P(A \cap B^c) + P(A \cap B) + P(A^c \cap B) \\ &= \big[P(A \cap B^c) + P(A \cap B)\big] + \big[P(A^c \cap B) + P(A \cap B)\big] - P(A \cap B) \\ &= P\big[(A \cap B^c) \cup (A \cap B)\big] + P\big[(A^c \cap B) \cup (A \cap B)\big] - P(A \cap B) \\ &= P(A) + P(B) - P(A \cap B) \end{aligned}$$

5. (10 marks) In cyber warfare, malware is introduced to a target system via an intermediate host. The target system is the computer system of the target company. The intermediate host is a home PC, smart phone or MP3 player of an employee of the target company. The intermediate host can be infected from a number of sources; via Bluetooth (S_1), via the Internet (S_2) or from a Memory card (S_3).

   The frequencies of infection by the different sources are shown in Table 1:

Source of infection	Probability of exposure
Bluetooth S_1	$P(S_1) = 0.4$
Internet S_2	$P(S_2) = 0.4$
Memory card S_3	$P(S_3) = 0.2$

Table 1:

Source	Intermediate host		
	Home PC H_1	Smart phone H_2	MP3 player H_3
Bluetooth S_1	$P(H_1\|S_1) = 0.1$	$P(H_2\|S_1) = 0.2$	$P(H_3\|S_1) = 0.7$
Internet S_2	$P(H_1\|S_2) = 0.3$	$P(H_2\|S_2) = 0.6$	$P(H_3\|S_2) = 0.1$
Memory card S_3	$P(H_1\|S_3) = 0.2$	$P(H_2\|S_3) = 0.5$	$P(H_3\|S_3) = 0.3$

Table 2:

The probabilities of infection given exposure to a source are in Table 2, where H_i denotes an infected host:

Lastly, the probability a target is infected from a specific host is $P(T|H_i)$, $i = 1, 2, 3$, where T denotes an infected target and these probabilities are given by

$$P(T|H_1) = 0.4 \qquad P(T|H_2) = 0.2 \qquad P(T|H_3) = 0.8.$$

Calculate the probability that the target is infected.

Solution. Let $P(T)$ be the probability that the target is infected. By the total probability formula, we obtain

$$P(T) = P(H_1)P(T|H_1) + P(H_2)P(T|H_2) + P(H_3)P(T|H_3).$$

Similarly,

$$P(H_i) = P(S_1)P(H_i|S_1) + P(S_2)P(H_i|S_2) + P(S_3)P(H_i|S_3), \quad i = 1, 2, 3.$$

Then,

$$P(H_1) = 0.4 \times 0.1 + 0.4 \times 0.3 + 0.2 \times 0.2 = 0.20,$$
$$P(H_2) = 0.4 \times 0.2 + 0.4 \times 0.6 + 0.2 \times 0.5 = 0.42,$$

and

$$P(H_3) = 0.4 \times 0.7 + 0.4 \times 0.1 + 0.2 \times 0.3 = 0.38.$$

Thus,

$$P(T) = 0.2 \times 0.4 + 0.42 \times 0.2 + 0.38 \times 0.8 = 0.468$$

6. (14 marks) Let N be the number of distinct values observed when a 6-sided die is rolled 4 times. For example, if the outcome is 2114 then the observed values are $\{1, 2, 4\}$ and $N = 3$.

   (a) (8 marks) What is $\mathrm{E}[N]$?

(b) (6 marks) What is the expected value of N conditioned on all the observed values are odd?

Solution:

(a) We can write $N = N_1+N_2+N_3+N_4+N_5+N_6$ where N_i takes value 1 if a 1 was observed from the i-th roll and 0 if not. Then $\mathrm{E}[N_1] = \mathrm{P}[N_1 = 1]$ is the probability that a 1 was observed. This is one minus the probability that a 1 was not observed, which is $(5/6)^4$ by independence. Thus, $\mathrm{E}[N_1] = 1-(5/6)^4$. Clearly, $\mathrm{E}[N_1] = \cdots = \mathrm{E}[N_6]$, then $\mathrm{E}[N] = 6(1-(5/6)^4) \approx 3.107$.

(b) Let A be the event of "all the observed values are odd". We can write $\mathrm{E}[N \mid A] = \mathrm{E}[N_1 \mid A] + \mathrm{E}[N_2 \mid A] + \mathrm{E}[N_3 \mid A] + \mathrm{E}[N_4 \mid A] + \mathrm{E}[N_5 \mid A] + \mathrm{E}[N_6 \mid A]$. $\mathrm{E}[N_2 \mid A] = \mathrm{E}[N_4 \mid A] = \mathrm{E}[N_6 \mid A] = \mathrm{E}[\emptyset] = 0$. By symmetry, $\mathrm{E}[N_1 \mid A] = \mathrm{E}[N_3 \mid A] = \mathrm{E}[N_5 \mid A] = 1-(2/3)^4$, then $\mathrm{E}[N \mid A] = 3(1-(2/3)^4) \approx 2.407$.

7. (20 marks) Suppose there are two fair 4-sided dice with the numbers $\{1, 2, 3, 4\}$. Roll the two dice at the same time. Let X denote the **sum** of the numbers on the two dice, and Y denote the **absolute value of the difference** between the numbers of the two dice.

(a) (8 marks) Calculate the joint PMF $P_{XY}(x, y)$ and the marginal PMFs $P_X(x)$ and $P_Y(y)$.

(b) (7 marks) Calculate the expectation and variance of $X + 2Y$.

(c) (5 marks) Are X and Y independent? Justify your answer.

Solution:

(a) There are 16 conditions for rolling two four-sided fair dice. By analyzing their sums and absolute value of differences, we have: the value of X can be $2, 3, 4, 5, 6, 7, 8$; the value of Y can be $0, 1, 2, 3$. So the joint PMF of X and Y should be:

$y \backslash x$	2	3	4	5	6	7	8
0	1/16	0	1/16	0	1/16	0	1/16
1	0	2/16	0	2/16	0	2/16	0
2	0	0	2/16	0	2/16	0	0
3	0	0	0	2/16	0	0	0

By summing each column, we have $P_X(x)$:

x	2	3	4	5	6	7	8
$P(x)$	1/16	2/16	3/16	4/16	3/16	2/16	1/16

By summing each row, we have $P_Y(y)$:

y	0	1	2	3
$P(y)$	4/16	6/16	4/16	2/16

(b) The value of $X + 2Y$ can be $2, 4, 5, 6, 7, 8, 9, 10, 11$, and its PMF is:

$x + 2y$	2	4	5	6	7	8	9	10	11
$P(x + 2y)$	1/16	1/16	2/16	1/16	2/16	3/16	2/16	2/16	2/16

So, $E[X + 2Y] = 2 \times \frac{1}{16} + 4 \times \frac{1}{16} + \cdots + 11 \times \frac{2}{16} = 7.5$.

$Var[X + 2Y] = (7.5-2)^2 \times \frac{1}{16} + (7.5-4)^2 \times \frac{1}{16} + \cdots + (7.5-11)^2 \times \frac{2}{16} = 6.25$.

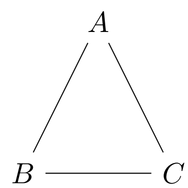
(c) X and Y are not independent. We can give many counterexamples.

For example $P(x=2, y=0) = \frac{1}{16}$, but $P(x=2) = \frac{1}{16}$ and $P(y=0) = \frac{4}{16}$.

So $P(x=2, y=0) \neq P(x=2) \times P(y=0)$. X and Y are not independent.

8. (16 marks) This question concerns faults and failures.

(a) (6 marks) Computers A, B and C are linked through three cables as shown in the figure below. Each cable fails with probability 10% independently of the others. Let E_{xy} be the event that "there is at least a working connection between computers x and y.". E_{xy}^c is the complement of E_{xy}, which is the event that "there is no working connections between computers x and y.".



What is the probability of E_{BC}?

(b) (10 marks) Suppose for a specific link, it fails at an average rate of two times per year. Bob observed at least one failure within the last year. Given this information, what is the expected number of failures on this link within the last year?

Solution:

(a) Let F_{xy} be the event "Cable xy fails". The event E_{BC} does not happen when cable BC fails and at least one of the cables AB, AC also fails, thus:

$$\mathrm{P}(E_{BC}) = 1 - \mathrm{P}(F_{BC})(1 - \mathrm{P}(F_{BA}^c)\,\mathrm{P}(F_{AC}^c)) = 1 - 0.1 \cdot (1 - 0.9^2) = 0.981$$

(b) If we model the number of failures on this link in last year as a Poisson(2) random variable N, we are looking for the expectation of N given $N > 0$. By the total expectation formula

$$\mathrm{E}[N] = \mathrm{E}[N|N>0]\,\mathrm{P}(N>0) + \mathrm{E}[N|N=0]\,\mathrm{P}(N=0).$$

We know that $\mathrm{E}[N] = 2$, $\mathrm{E}[N|N=0] = 0$, and $\mathrm{P}(N=0) = e^{-2} \cdot 2^0/0! = e^{-2}$. By the axioms $\mathrm{P}(N>0) = 1 - \mathrm{P}(N=0) = 1 - e^{-2}$, so

$$\mathrm{E}[N|N>0] = \frac{\mathrm{E}[N]}{\mathrm{P}(N>0)} = \frac{2}{1-e^{-2}} \approx 2.313.$$