CS203 (2023) – First assignment

Total marks: 40

- **Note.** Answers without clear and concise explanations will not be taken into account. Use of immoral means will get severe punishment.

Name:____________________________

Roll No:____________________________

Questions

1. **(10 marks)** Let $\{A_i\}_{i=1}^n$ be a family of sets indexed from 1 to n. Let $I \subseteq [n]$ be a subset of index set. Let B be the event when only and all A_i's from I have happened, i.e.,

$$B = (\cap_{i \in I} A_i) \cap (\cap_{i \notin I} A_i^c) .$$

Notice that B is subset of $\cap_{i \in I} A_i$, but need not be equal to it. Show that,

$$P(B) = \sum_{J \supseteq I} (-1)^{|J|-|I|} P(\cap_{i \in J} A_i).$$

Hint: can you define some new sets in terms of A_i, such that, this problem looks like inclusion-exclusion?

Solution:

$B = (\bigcap_{i \in I} A_i) \cap (\bigcap_{i \notin I} A_i^c) = (\bigcap_{i \in I} A_i) \cap (\bigcup_{i \notin I} A_i)^c$(De Morgans law)
Since, $(\bigcup_{i \notin I} A_i)^c)$ and $(\bigcup_{i \notin I} A_i))$ are disjoint and mutually exhaustive, by partion theorem we can write,
$P(\bigcap_{i \in I} A_i) = P((\bigcap_{i \in I} A_i) \cap (\bigcup_{i \notin I} A_i)) + P((\bigcap_{i \in I} A_i) \cap (\bigcup_{i \notin I} A_i)^c)$
$P(\bigcap_{i \in I} A_i) = P(\bigcup_{i \notin I}((\bigcap_{j \in I} A_j) \cap A_i)) + P((\bigcap_{i \in I} A_i) \cap (\bigcup_{i \notin I} A_i)^c)$

$P(\bigcap_{i\in I} A_i) = P(\bigcup_{i\notin I}(\bigcap_{j\in I} A_j \cap A_i)) + P(B)$

$P(B) = P(\bigcap_{i\in I} A_i) - P(\bigcup_{i\notin I}(\bigcap_{j\in I} A_j \cap A_i))$

For each $i \notin I$, define $B_i = \bigcap_{j\in I} A_j \cap A_i$. From the inclusion-exclusion principle, we have:

$P(\bigcup_{i\notin I} B_i) = \sum_{K\subseteq I^c, K\neq\varnothing}(-1)^{|K|+1}P(\bigcap_{i\in K} B_i)$

$= \sum_{K\subseteq I^c, K\neq\varnothing}(-1)^{|K|+1}P(\bigcap_{i\in K} A_i \cap \bigcap_{j\in I} A_j)$

Taking $J = K \cup I$ and using the fact that $K \cap I = \varnothing$, we get:

$P(\bigcup_{i\notin I}(\bigcap_{j\in I} A_j \cap A_i)) = P(\bigcup_{i\notin I} B_i) = \sum_{J\supset I}(-1)^{|J|-|I|+1}P(\bigcap_{i\in J} A_i) -> [1]$

Since $\bigcup_{i\notin I}(\bigcap_{j\in I} A_j \cap A_i) \subseteq \bigcap_{i\in I} A_i$ and $B \cap (\bigcup_{i\notin I}(\bigcap_{j\in I} A_j \cap A_i)) = \varnothing$, we have:

$P(B) = P(\bigcap_{i\in I} A_i) - P(\bigcup_{i\notin I}(\bigcap_{j\in I} A_j \cap A_i))$

$= P(\bigcap_{i\in I} A_i) - \sum_{J\supset I}(-1)^{|J|-|I|+1}P(\bigcap_{i\in J} A_i)$ (From [1])

$= P(\bigcap_{i\in I} A_i) + \sum_{J\supset I}(-1)^{|J|-|I|}P(\bigcap_{i\in J} A_i)$

$= \sum_{J\supseteq I}(-1)^{|J|-|I|}P(\bigcap_{i\in J} A_i)$

$\square$

2. **(15 marks)** To study the efficacy of two tests for a set of three diseases, ICMR conducted trials with 10000 patients having one of these diseases. We restrict our universe to set of people having one of these diseases. The number of people who had these diseases, and the outcome of the result of these two tests on those patients is given in the table below.

Disease	Numbers having this disease	Result ++	Result +-	Result -+	Result −
d_1	3215	2110	301	704	100
d_2	2125	396	132	1187	410
d_3	4660	510	3568	73	509

Assume that these number accurately reflect the probability of people having a certain disease and efficacy of the tests. For a new patient (assuming he has one of the diseases), given test results, we want to estimate the probability of having d_1,d_2 or d_3. In other words fill out the following table.

Outcome	d_1	d_2	d_3
+ +			
+ -			
- +			
- -			

Solution: We need to find P[d$_1$ | ++], $P[d_2 \mid ++], P[d_3 \mid ++]$

$P[d_i \mid ++] := \frac{P(d_i \cap ++)}{P(++)}, \forall i \in \{1,2,3\}$

$P(++) = \frac{2110+396+510}{10000} = \frac{3016}{10000}$

$P(d_1 \cap ++) = \frac{2110}{10000}$

$P(d_2 \cap ++) = \frac{396}{10000}$

$P(d_3 \cap ++) = \frac{510}{10000}$

$P[d_1 \mid ++] = \frac{\frac{2110}{10000}}{\frac{3016}{10000}} = \frac{2110}{3016}$

$P[d_2 \mid ++] = \frac{\frac{396}{10000}}{\frac{3016}{10000}} = \frac{396}{3016}$

$P[d_3 \mid ++] = \frac{\frac{510}{10000}}{\frac{3016}{10000}} = \frac{510}{3016}$

$P(+-) = \frac{301+132+3568}{10000} = \frac{4001}{10000}$

$P(d_1 \cap +-) = \frac{301}{10000}$

$P(d_2 \cap +-) = \frac{132}{10000}$

$P(d_3 \cap +-) = \frac{3568}{10000}$

$P[d_1 \mid +-] = \frac{\frac{301}{10000}}{\frac{4001}{10000}} = \frac{301}{4001}$

$P[d_2 \mid +-] = \frac{\frac{132}{10000}}{\frac{4001}{10000}} = \frac{132}{4001}$

$P[d_3 \mid +-] = \frac{\frac{3568}{10000}}{\frac{4001}{10000}} = \frac{3568}{4001}$

$P(-+) = \frac{704+1187+73}{10000} = \frac{1964}{10000}$

$P(d_1 \cap -+) = \frac{704}{10000}$

$P(d_2 \cap -+) = \frac{1187}{10000}$

$P(d_3 \cap -+) = \frac{73}{10000}$

$P[d_1 \mid -+] = \frac{\frac{704}{10000}}{\frac{1964}{10000}} = \frac{704}{1964}$

$P[d_2 \mid -+] = \frac{\frac{1187}{10000}}{\frac{1964}{10000}} = \frac{1187}{1964}$

$P[d_3 \mid -+] = \frac{\frac{73}{10000}}{\frac{1964}{10000}} = \frac{73}{1964}$

$P(--) = \frac{100+410+509}{10000} = \frac{1019}{10000}$

$P(d_1 \cap --) = \frac{100}{10000}$

$P(d_2 \cap --) = \frac{410}{10000}$

$P(d_3 \cap --) = \frac{509}{10000}$

$P[d_1 \mid --] = \frac{\frac{100}{10000}}{\frac{1019}{10000}} = \frac{100}{1019}$

$P[d_2 \mid --] = \frac{\frac{410}{10000}}{\frac{1019}{10000}} = \frac{410}{1019}$

$P[d_3 \mid --] = \frac{\frac{509}{10000}}{\frac{1019}{10000}} = \frac{509}{1019}$

Outcome	d_1	d_2	d_3
+ +	$\frac{2110}{3016}$	$\frac{396}{3016}$	$\frac{510}{3016}$
+ -	$\frac{301}{4001}$	$\frac{132}{4001}$	$\frac{3568}{4001}$
- +	$\frac{704}{1964}$	$\frac{1187}{1964}$	$\frac{73}{1964}$
- -	$\frac{100}{1019}$	$\frac{410}{1019}$	$\frac{509}{1019}$

Outcome	d_1	d_2	d_3
+ +	0.699	0.131	0.169
+ -	0.075	0.033	0.892
- +	0.358	0.604	0.037
- -	0.098	0.402	0.499

$\square$

(5 + 10 marks) Your friend is trying to invest in NFTs. In particular, he is investing in a scheme where you get a random NFT from a set of n NFTs by giving Rs 50. The probability of getting a particular NFT is $\frac{1}{n}$. If you get a new NFT that wasn't in your collection already, then it is added to your collection, otherwise that buy is a waste. The company has an offer where they give Rs. 100 for each NFT if you have the complete set, and hence "double" the money. Let X be the random variable that counts the number of times your friend has to request NFT's so that he has a complete set. You have decided to show your friend that this is a losing deal and save him from the scam. Calculate the expectation and variance of X.

Solution:

Let A_1 be the random variable that we get the first distinct NFT (which is clearly $A_1 = 1$).
Let A_i be a random variable that counts the number of NFTs to be requested to have i different NFTs after getting i-1 different NFTs,(it has no dependence on A$_j where j < I) \forall i \in \{2, 3, ...n\}$
So, by definition of X, $X = A_1 + A_2 + ... + A_n = \sum_{i=1}^{n} A_i$
By linearity of expectation $E[X] = E[\sum_{i=1}^{n} A_i] = \sum_{i=1}^{n} E[A_i]$
$E[A_i] = \sum_{r=1}^{\infty} r(\frac{i-1}{n})^{r-1}(\frac{n-i+1}{n}) -> [1]$
$E[A_i] = \sum_{r=0}^{\infty}(r+1)(\frac{i-1}{n})^{r}(\frac{n-i+1}{n})$
$\implies$ E[A$_i] = \sum_{r=1}^{\infty}(r)(\frac{i-1}{n})^{r}(\frac{n-i+1}{n}) + \sum_{r=0}^{\infty}(\frac{i-1}{n})^{r}(\frac{n-i+1}{n}) -> [2]$

Multiply eq [1] by $\frac{i-1}{n}$ and subtract it from [2]

$[2] - \frac{i-1}{n}[1] = (\frac{n-i+1}{n})\sum_{r=0}^{\infty}(\frac{i-1}{n})^{r}$

$\implies (E[A_i])(\frac{n-i+1}{n}) = (\frac{n-i+1}{n})(\frac{1}{1-(\frac{i-1}{n})})$

$\implies E[A_i] = \frac{n}{n-i+1}$
$\implies E[X] = \sum_{i=1}^{n} \frac{n}{n-i+1}$
$\implies E[X] = n\sum_{i=1}^{n} \frac{1}{i}$
Expected gain of the person $100n - 50E[X] = 50n(2 - \sum_{i=1}^{n} \frac{1}{i}) = 50n(1 - \sum_{i=2}^{n} \frac{1}{i})$
Cleary the gain would be negative for n>3, hence it's a losing deal.
For $n \to \infty$, E[X]=n(2-ln(n))

Since, we know that $A_i's$ are independent, We can write the $var[X] = \sum_{i=1}^{n} var[A_i]$
Proof:
Var $\left[\sum_{i=1}^{n} A_i\right] = \mathbb{E}\left[\left(\sum_{i=1}^{n} A_i\right)^2\right] - \mathbb{E}\left[\sum_{i=1}^{n} A_i\right]^2$
$= \sum_{i=1}^{n} \mathbb{E}[A_i^2] + \sum_{i\neq j} \mathbb{E}[A_iA_j] - \left(\sum_{i=1}^{n} \mathbb{E}[A_i]\right)^2$
$\mathbb{E}[A_iA_j] = \mathbb{E}[A_i]\mathbb{E}[A_j]$, since $A_i's$ are independent
$= \sum_{i=1}^{n} \mathbb{E}[A_i^2] + \sum_{i\neq j} \mathbb{E}[A_i]\mathbb{E}[A_j] - \sum_{i=1}^{n} \mathbb{E}[A_i]^2 - \sum_{i\neq j} \mathbb{E}[A_i]\mathbb{E}[A_j]$
$= \sum_{i=1}^{n} \mathrm{Var}[A_i]$.
Now, $var[A_i] = E[A_i^2] - E[A_i]^2$
$E[A_i^2] = \sum_{r=1}^{\infty} r^2 (\frac{i-1}{n})^{r-1} (\frac{n-i+1}{n})$
Let $b_i = \frac{i-1}{n}, \forall i \in \{1, 2, 3, ...n\}$
$E[A_i^2] = (1 - b_i)\sum_{r=1}^{\infty} r^2 (b_i)^{r-1} -> [3]$
$E[A_i^2] = (1 - b_i)\sum_{r=0}^{\infty} (r + 1)^2 (b_i)^{r}$
$E[A_i^2] = (1 - b_i)\sum_{r=1}^{\infty} (r)^2 (b_i)^{r} + 2 * (1 - b_i)\sum_{r=1}^{\infty} r(b_i)^r + (1 - b_i)\sum_{r=0}^{\infty} (b_i)^r -> [4]$
Multiply eq [3] by b_i and subtract it from [4]
$E[A_i^2](1 - b_i) = +2 * (1 - b_i)\sum_{r=1}^{\infty} r(b_i)^r + (1 - b_i)\sum_{r=0}^{\infty} (b_i)^r$
Dividing by $1 - b_i$ throughout
$E[A_i^2] = 2 * \sum_{r=1}^{\infty} r(b_i)^r + \sum_{r=0}^{\infty} (b_i)^r -> [5]$
$E[A_i^2] = 2 * \sum_{r=0}^{\infty} (r + 1)(b_i)^{r+1} + \sum_{r=0}^{\infty} (b_i)^r$
$E[A_i^2] = 2 * \sum_{r=1}^{\infty} (r)(b_i)^{r+1} + 2 * \sum_{r=0}^{\infty} (b_i)^{r+1} + \sum_{r=0}^{\infty} (b_i)^r -> [6]$
Multiply eq [4] by b_i and subtract it from [5]
$E[A_i^2](1 - b_i) = \sum_{r=0}^{\infty} (b_i)^{r+1} + \sum_{r=0}^{\infty} (b_i)^r$
$E[A_i^2](1 - b_i) = \frac{b_i}{1-b_i} + \frac{1}{1-b_i}$
$E[A_i^2] = \frac{b_i+1}{(1-b_i)^2}$
As we know $var[A_i] = E[A_i^2] - E[A_i]^2$
We can write,$E[A_i] = \frac{1}{1-b_i}$
$var[A_i] = \frac{b_i+1}{(1-b_i)^2} - (\frac{1}{1-b_i})^2$
$var[A_i] = \frac{b_i}{(1-b_i)^2} = \frac{1}{(1-b_i)^2} - \frac{1}{(1-b_i)} = (\frac{n}{n-i+1})^2 - (\frac{n}{n-i+1})$
$var[X] = \sum_{i=1}^{n} var[A_i] = \sum_{i=1}^{n} (\frac{n}{n-i+1})^2 - (\frac{n}{n-i+1})$

$var[X] = (n)^2(\frac{1}{1^2} + \frac{1}{2^2} ..... \frac{1}{n^2}) - n(\frac{1}{1} + \frac{1}{2} ..... \frac{1}{n})$
var[X]=(n)$^2 \sum_{i=1}^{n}(\frac{1}{i})^2 - n \sum_{i=1}^{n}(\frac{1}{i})$ $\square$