Probability & Statistics for EECS: Midterm Exam

Due on Nov 21, 2023 at 23:59

Name:
Student ID:

Oct 22, 2023

Problem 1

(10 points) How many number of distinct positive integer-valued vectors $(x_1, x_2, \ldots, x_5)$ satisfying the equation and inequalities

$$x_1 + x_2 + \cdots + x_5 = 48$$
$$x_1 > 3, x_2 \geq 7, x_3 > 4, x_4 \geq 6, x_5 > 4.$$

Denote $y_1, y_2, \ldots, y_5$ be positive random variables that

$$y_1 = x_1 - 4, y_2 = x_2 - 7, y_3 = x_3 - 5, y_4 = x_4 - 6, y_5 = x_5 - 5.$$

Thus the original problem equals the number of non-negative integer-vectors $(y_1, y_2, y_3, y_4, y_5)$ satisfying

$$y_1 + y_2 + \cdots + y_5 = 48 - 4 - 7 - 5 - 6 - 5 = 21$$
$$y_i \geq 0, \quad \text{for any } i.$$

According to Bose-Einstein counting, the number of $(x_1, x_2, \cdots, x_5)$ is

$$\binom{n+k-1}{n-1} = \binom{5+21-1}{5-1} = \binom{25}{4}.$$

Problem 2

(10 points) A system composed of 5 homogeneous devices is shown in the following figure. It is said to be functional when there exists at least one end-to-end path that devices on such path are all functional. For such a system, if each device, which is independent of all other devices, functions with probability p, then what is the probability that the system functions? Such a probability is also called the system reliability.

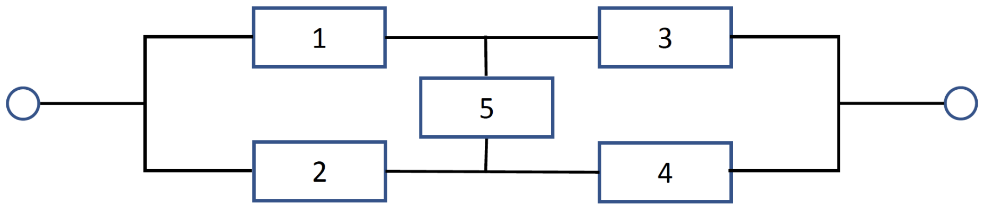

Figure 1: An illustration of the system composed of 5 homogeneous devices.

Method 1: In order to calculate the probability of system activation, we can classify according to the number of activated devices:

- If two devices are activated (1/3, 2/4): $P_2 = 2 \times p^2(1-p)^3$
- If three devices are activated (all cases except 1/2/5, 3/4/5): $P_3 = (C_5^3 - 2) \times p^3(1-p)^2 = 8p^3(1-p)^2$
- If four devices are activated (Four of any five devices are activated): $P_4 = 5 \times p^4(1-p)$
- If five devices are activated: $P_5 = p^5$

So the probability that the system is activated is:

$$\begin{aligned} P &= P_2 + P_3 + P_4 + P_5 \\ &= 2p^5 - 5p^4 + 2p^3 + 2p^2 \end{aligned}$$

Method 2: There are two situations whether the device 5 is functional.

1. ***case 1:*** device 5 is functional, the probability is p, the system is functional if and only if device 1 or device 2 is functional and device 3 or device 4 is functional, the probability is $(p+p-p^2)*(p+p-p^2)$. Therefore, the probability that the system functions is $p*(p+p-p^2)*(p+p-p^2)$.
2. ***case 2:*** device 5 is not functional, the probability is $1-p$, the system is functional if and only if both device 1 and device 3 are functional or both device 2 and device 4 are functional, the probability is $p^2+p^2-p^4$. Therefore, the probability that the system functions is $(1-p)*(p^2+p^2-p^4)$.

In conclusion, taking both case 1 and case 2 into consideration, the probability that the system functions is $p*(p+p-p^2)*(p+p-p^2)+(1-p)*(p^2+p^2-p^4) = 2p^5 - 5p^4 + 2p^3 + 2p^2$.
Let $p = 1/2$, we have the final answer as $\frac{1}{2}$.

Problem 3

(10 points) A frog starts at 0 on the number line and makes a sequence of jumps to the right. In any one jump, independent of previous jumps, this frog leaps a positive integer distance m with probability $\frac{1}{2^m}$. Find the probability that this frog will eventually land at 1997 on the number line.

Intuition. At any point of the jumping sequence, the probabilities of landing *at* 1997 (or whatever other positive numbers) and landing *past* 1997 are exactly the same (verified by simple calculation). Therefore, the result must be $\frac{1}{2}$.

Assume a frog starts from the origin and reaches the point 1997 on the number line after $n \geq 1$ jumps, with each jump consisting of a positive integer number of steps $x_1, \ldots, x_n$. Then we have

$$x_1 + \ldots + x_n = 1997, x_i \geq 1, i \in \{1, \ldots, n\}.$$

According to the problem, the probability of the frog jumping a distance of x_i is $\frac{1}{2^{x_i}}$. Therefore, the probability of the event "the frog starts from the origin and after $n \geq 1$ jumps of fixed distances $x_1, \ldots, x_n$, reaches the point 2024 on the number line" is

$$\frac{1}{2^{x_1}} \cdot \frac{1}{2^{x_2}} \cdots \frac{1}{2^{x_n}} = \frac{1}{2^{x_1+\ldots+x_n}} = \frac{1}{2^{1997}}.$$

Next, we will examine how many positive integer solutions satisfy the indeterminate equation $x_1 + \ldots + x_n = 2024$ under a given n. As known from the method of partitions, there are

$$\binom{1997-1}{n-1} = \binom{1996}{n-1}$$

such positive integer solutions. Therefore, under a given n, the probability of the event "the frog starts from the origin and after $n \geq 1$ jumps, reaches the point 2024 on the number line" is

$$\frac{1}{2^{1997}}\binom{1996}{n-1}.$$

Since the frog jumps at least one step each time, it jumps at least once and at most 1997 times to reach the point 1997 on the number line, thus $1 \leq n \leq 1997$. Therefore, the probability of the event "the frog jumps to the point 1997 on the number line" is

$$\sum_{n=1}^{1997} \frac{\binom{1996}{n-1}}{2^{1997}} = \frac{1}{2^{1997}} \cdot \sum_{n=1}^{1997} \binom{1996}{n-1} = \frac{1}{2^{1997}} \cdot 2^{1996} = \frac{1}{2}.$$

Similarly, we can calculate the probability of the frog jumping to the point $k \geq 1$ on the number line as

$$\sum_{n=1}^{k} \frac{\binom{k-1}{n-1}}{2^k} = \frac{1}{2^k} \cdot \left(\sum_{n=1}^{k} \binom{k-1}{n-1}\right) = \frac{1}{2^k} \cdot 2^{k-1} = \frac{1}{2}.$$

In the penultimate step, we used the combinatorial identity

$$\binom{k-1}{0} + \binom{k-1}{1} + \ldots + \binom{k-1}{k-1} = 2^{k-1}.$$

Problem 3 continued on next page. . .

Alternative. We first define $f(k)$ as the probability that this frog will eventually land at k on the number line. Via first step analysis, we have

$$f(k) = \frac{1}{2}f(k-1) + \frac{1}{2^2}f(k-2) + \ldots + \frac{1}{2^{k-1}}f(1).$$

Note that we have

$$\begin{aligned}
f(k) &= \frac{1}{2}f(k-1) + \frac{1}{2^2}f(k-2) + \ldots + \frac{1}{2^{k-1}}f(1) \\
&= \frac{1}{2}f(k-1) + \frac{1}{2}\left(\frac{1}{2^1}f(k-2) + \ldots + \frac{1}{2^{k-2}}f(1)\right) \\
&= \frac{1}{2}f(k-1) + \frac{1}{2}f(k-1) \\
&= f(k-1).
\end{aligned}$$

Therefore,

$$f(k) = f(k-1) = \ldots = f(1) = \frac{1}{2}.$$

Problem 4

(10 points) A six-sided fair dice is rolled three times independently. What is more likely: a sum of 15 or a sum of 16? (You need to compute the corresponding probability for each case).

Let X_i be the number of ith roll, and $X = X_1 + X_2 + X_3 + X_4$ be total number. The PGF of X_1 is $\mathrm{E}\left(t^{X_1}\right) = \frac{1}{6}\sum_{i=1}^{6} t^i$. Then we have

$$\mathrm{E}\left(t^X\right) = \mathrm{E}\left(t^{X_1}t^{X_2}t^{X_3}t^{X_4}\right) = \frac{1}{6^4}\left(\sum_{i=1}^{6} t^i\right)^4.$$

Since $\mathrm{E}\left(t^X\right) = \sum_{k=4}^{24} P\left(X = k\right) t^k$, $P(X = k)$ is equal to the coefficient of t^k in $\frac{1}{6^4}\left(\sum_{i=1}^{6} t^i\right)^4$, or simpler, in $\left(\sum_{i=1}^{6} t^i\right)^4$. Therefore, we obtain

$$P(X = 15) = \frac{140}{1296} > P(X = 16) = \frac{125}{1296}.$$

Problem 5

(10 points) Consider a coin that comes up heads with probability p and tails with probability $1-p$. Let q_n be the probability that after n independent tosses, there have been an odd number of heads. Find q_n.
If after n independent tosses, there have been an *odd number of heads*, then there can be two cases.

1. After $n-1$ independent tosses, there have been an odd number of heads, and the nth toss gets a tail.
2. After $n-1$ independent tosses, there have been an even number of heads, and the nth toss gets a head.

Denote O_n as the event that there have been an odd number of heads after n independent tosses, and T_n as the event that the nth toss gets a tail, then from the analysis above, we have

$$\begin{aligned} q_n = P(O_n) &= P(O_n|O_{n-1})P(O_{n-1}) + P(O_n|O_{n-1}^c)P(O_{n-1}^c) \\ &= P(T_n)P(O_{n-1}) + P(T_n^c)P(O_{n-1}^c) \\ &= (1-p)\cdot q_{n-1} + p\cdot(1-q_{n-1}) \end{aligned}$$

from which we get

$$q_n = (1-2p)q_{n-1} + p.$$

Reformulate the equation we get

$$q_n - \frac{1}{2} = (1-2p)\cdot\left(q_{n-1} - \frac{1}{2}\right),$$

and we can see that $\{q_n - \frac{1}{2}, n \geq 0\}$ is a geometric progression. Since $q_0 = 0$, for any $n \in \{0, 1, \dots\}$, we have

$$\begin{aligned} q_n - \frac{1}{2} &= (1-2p)^n\left(q_0 - \frac{1}{2}\right) = -\frac{1}{2}\times(1-2p)^n \\ \Rightarrow q_n &= \frac{1}{2} - \frac{1}{2}(1-2p)^n. \end{aligned}$$

Another method:
From $q_n = (1-2p)\,q_{n-1} + p$, another solution is to use characteristic function

$$\begin{aligned} x^2 &= (1-2p)x + p \\ \Rightarrow x_1 &= \frac{(1-2p)+\sqrt{4p^2+1}}{2}, x_2 = \frac{(1-2p)-\sqrt{4p^2+1}}{2} \end{aligned}$$

and thus we have

$$q_n = a\left(\frac{(1-2p)+\sqrt{4p^2+1}}{2}\right)^n + b\left(\frac{(1-2p)-\sqrt{4p^2+1}}{2}\right)^n.$$

Since $q_0 = 0$ and $q_1 = p$, we have

$$\begin{aligned} q_0 &= 0 = a + b \\ q_1 &= p = a\left(\frac{(1-2p)+\sqrt{4p^2+1}}{2}\right) + b\left(\frac{(1-2p)-\sqrt{4p^2+1}}{2}\right) \\ \Rightarrow a &= \frac{p}{\sqrt{4p^2+1}}, b = -\frac{p}{\sqrt{4p^2+1}}. \end{aligned}$$

Therefore, we have

$$q_n = \frac{p}{\sqrt{4p^2+1}}\left(\frac{(1-2p)+\sqrt{4p^2+1}}{2}\right)^n - \frac{p}{\sqrt{4p^2+1}}\left(\frac{(1-2p)-\sqrt{4p^2+1}}{2}\right)^n$$

Problem 6

(10 points) An urn contains w white balls and b black balls, which are randomly drawn one by one without replacement. Let X denote the number of black balls drawn before drawing r $(1 \le r \le w)$ white balls.

(a) (5 points) Find the PMF of X.

(b) (5 points) Find $E(X)$.

(a) To calculate the PMF firstly we select a white ball as the r-th white ball, which has w possibilities. Then select k black balls and $r-1$ white balls from the left balls, which have $\binom{w-1}{r-1}\binom{b}{k}$ choices. If we randomly select the first ball and the left balls, we will have $(w+b)\binom{w+b-1}{k+r-1}$ choices. Therefore, the probability of drawing k black balls is:

$$P(X=k) = \frac{w}{w+b}\frac{\binom{w-1}{r-1}\binom{b}{k}}{\binom{w+b-1}{k+r-1}} \tag{1}$$

Other solution: the event can be describe as the following two events happens concurrently: A: obtain k black balls and $r-1$ white balls in the first $k+r-1$ trail, event B get white ball in the $k+r$-th trail. Therefore

$$P(A,B) = P(A)P(B|A) = \frac{w-(r-1)}{w+b-(k+r-1)}\frac{\binom{w}{r-1}\binom{b}{k}}{\binom{w+b}{k+r-1}}.$$

(b) Define the indicator I_i: the i-th black ball is selected before the r-th white ball. Since the probability that the i-th black ball is selected between the gap of drawing two white balls are all the same (including the two end). Therefore we can know the expectation of I_i is $\frac{r}{w+1}$ since there are $w+1$ gaps and the black ball should be drawn from the first r gaps. Further we can obtain the expectation of X as follows:

$$E(X) = E\left(\sum_{i=1}^{b} I_i\right) = \sum_{i=1}^{b} E(I_i) = \frac{br}{w+1} \tag{2}$$

Problem 7

(20 points) Bob's database of friends contains n entries, but due to a software bug, the addresses correspond to the names in a totally random fashion. Bob writes a holiday card to each of his friends and sends it to the (software-corrupted) address. Let X denote the number of friends of him who will get the correct card.

(a) (5 points) Find $E(X)$.

(b) (5 points) Find $\mathrm{Var}(X)$.

(c) (5 points) Find the PMF of X.

(d) (5 points) When $n \to \infty$, show that the distribution of X converges to a Poisson distribution.

(a) Denote event A_i = "the ith friend of Bob gets the correct card" and let $I_i = I(A_i)$ be the indicator of event A_i. For any $i \in \{1, \cdots, n\}$, we have

$$\mathrm{E}(I_i) = P(A_i) = 1/n,$$

then it follows that

$$\mathrm{E}(X) = \mathrm{E}\left[\sum_{i=1}^{n} I_i\right] = \sum_{i=1}^{n} \mathrm{E}(I_i) = n \cdot \frac{1}{n} = 1.$$

(b) To obtain the variance of X, we use the equation $\mathrm{Var}(X) = \mathrm{E}\left(X^2\right) - \mathrm{E}^2(X)$. First, we calculate the second moment $\mathrm{E}(X^2)$ as follows:

$$\mathrm{E}\left(X^2\right) = \mathrm{E}\left[\left(\sum_{i=1}^{n} I_i\right)^2\right] = \mathrm{E}\left[\sum_{i=1}^{n} I_i^2 + 2\sum_{1\le i<j\le n} I_i I_j\right] = \sum_{i=1}^{n} \mathrm{E}\left[I_i\right] + 2\sum_{1\le i<j\le n} \mathrm{E}\left[I_i I_j\right].$$

Since $\mathrm{E}(I_i) = 1/n$ (by (a)) and

$$\mathrm{E}\left[I_i I_j\right] = P\left(A_i \cap A_j\right) = P\left(A_i\right) P\left(A_j | A_i\right) = \frac{1}{n} \cdot \frac{1}{n-1} = \frac{1}{n(n-1)},$$

we have

$$\mathrm{E}\left(X^2\right) = \sum_{i=1}^{n} \frac{1}{n} + 2\sum_{1\le i<j\le n} \frac{1}{n(n-1)} = n \cdot \frac{1}{n} + 2 \cdot \binom{n}{2} \cdot \frac{1}{n(n-1)} = 1 + 1 = 2.$$

Therefore, we obtain the variance of X:

$$\mathrm{Var}(X) = \mathrm{E}\left(X^2\right) - \mathrm{E}^2(X) = 2 - 1 = 1.$$

(c) First, we pick k ($k \in \{0, \ldots, n\}$) friends who will get the correct card from the n friends, there are $\binom{n}{k}$ different choices. Other $n-k$ friends get wrong cards.

We make some denotations first:

C = "the k chosen friends all get the correct card from the n cards"

W = "the other $n-k$ unchosen friends all get the wrong card from the rest $n-k$ cards". Then

$$P(X = k) = \binom{n}{k} P(C, W) = \binom{n}{k} P(W|C)\, P(C).$$

First, we compute $P(W|C)$:

$P(W|C) = 1 - P(W^c|C)$
$= 1 - P($"at least one of the $n-k$ unchosen friends get the correct card from the rest $n-k$ cards"$|C)$.

Denote $A_i =$ "the ith friend of the $n-k$ friends get the correct card", then using the Inclusion-Exclusion Formula, we have

$$\begin{aligned}
P\left(W^c \middle| C\right) =& P\left(\cup_{i=1}^{n-k} A_i|C\right) \\
=& \sum_{1\le i\le n-k} P\left(A_i|C\right) - \sum_{1\le i<j\le n-k} P\left(A_i\cap A_j|C\right) + \sum_{1\le i<j<p\le n-k} P\left(A_i\cap A_j\cap A_p|C\right) - \cdots \\
&+ (-1)^{n-k+1} P\left(A_1\cap\cdots\cap A_{n-k}|C\right) \\
=& (n-k)\, P\left(A_1|C\right) - \binom{n-k}{2} P\left(A_1\cap A_2|C\right) + \binom{n-k}{3} P\left(A_1\cap A_2\cap A_3|C\right) - \cdots \\
&+ (-1)^{n-k+1} P\left(\bigcap_{i=1}^{n-k} A_i|C\right).
\end{aligned}$$

Since

$$\begin{aligned}
P\left(A_1|C\right) &= \frac{1}{n-k} \\
P\left(A_1\cap A_2|C\right) &= \frac{(n-k-2)!}{(n-k)!} = \frac{1}{(n-k)(n-k-1)} \\
P\left(\bigcap_{i=1}^{n-k} A_i|C\right) &= \frac{1}{(n-k)!}
\end{aligned}$$

we get

$$P\left(W^c|C\right) = 1 - \frac{1}{2!} + \frac{1}{3!} - \cdots + (-1)^{n-k+1}\frac{1}{(n-k)!}$$

$$\Rightarrow P\left(W|C\right) = \frac{1}{2!} - \frac{1}{3!} + \cdots + (-1)^{n-k}\frac{1}{(n-k)!} = \sum_{i=0}^{n-k}\frac{(-1)^i}{i!}.$$

Then, we compute $P(C)$:

Number the k chosen friends as 1, 2, ..., k. Denote $C_i = \{$"the ith friend of the k chosen friends get the correct card from the n cards"$\}$, then

$$\begin{aligned}
P\left(C\right) &= P\left(C_1, C_2, \cdots, C_k\right) \\
&= P\left(C_1\right) P\left(C_2|C_1\right) P\left(C_3|C_2, C_1\right)\cdots P\left(C_k|C_{k-1},\cdots,C_1\right) \\
&= \frac{1}{n}\cdot\frac{1}{n-1}\cdot\frac{1}{n-2}\cdot\cdots\cdot\frac{1}{n(n-1)\cdots(n-k+1)} \\
&= \frac{(n-k)!}{n!}
\end{aligned}$$

Finally, we can get that for any $k\in\{0,\ldots,n\}$,

$$P(X=n) = \binom{n}{k} P\left(W|C\right) P\left(C\right) = \frac{1}{k!}\sum_{i=0}^{n-k}\frac{(-1)^i}{i!}.$$

(d) When $n\to\infty$, according to the above result, we have

$$\lim_{n\to\infty} P(X=n) = \lim_{n\to\infty}\frac{1}{k!}\sum_{i=0}^{n-k}\frac{(-1)^i}{i!} = \frac{e^{-1}}{k!} = \frac{(1)^k e^{-1}}{k!} \sim \text{Pois}(1).$$

Problem 8

(20 points) Consider the following 7-door version of the Monty Hall problem. There are 7 doors, behind one of which there is a car (which you want), and behind the rest of which there are goats (which you don't want). Initially, all possibilities are equally likely for where the car is. You choose a door. Monty Hall then opens 3 goat doors, and offers you the option of switching to any of the remaining 3 doors.

(a) (10 points) Assume that Monty Hall knows which door has the car, will always open 3 goat doors and offer the option of switching, and that Monty chooses with equal probabilities from all his choices of which goat doors to open. Should you switch? What is your probability of success if you switch to one of the remaining 3 doors?

(b) Generalize the above to a Monty Hall problem where there are $n \geq 3$ doors, of which Monty opens m goat doors, with $1 \leq m \leq n-2$.

(a) Assume the doors are labeled such that you choose Door 1 (to simplify notation), and suppose first that you follow the "stick to your original choice" strategy. Let S be the event of success in getting the car, and let C_j be the event that the car is behind Door j. Conditioning on which door has the car, we have

$$P(S) = \frac{1}{7}.$$

Let M_{ijk} be the event that Monty opens Doors i, j, k. Then

$$P(S) = \sum_{i,j,k} P(S|M_{i,j,k}), 2 \leq i < j < k \leq 7$$

By symmetry, this gives $P(S|M_{i,j,k}) = P(S) = \frac{1}{7}$. Thus the conditional probability that the car is behind 1 of the remaining 3 doors is 6/7, which gives 2/7 for each. So you should switch, thus making your probability of success 2/7 rather than 1/7

(b) The problem becomes the following: Consider the following n-door version of the Monty Hall problem. There are n doors, behind one of which there is a car (which you want), and behind the rest of which there are goats (which you don't want). Initially, all possibilities are equally likely for where the car is. You choose a door. Monty Hall then opens m goat doors, and offers you the option of switching to any of the remaining $n-m-1$ doors.

Without loss of generality, we can assume the contestant picked door 1 (if she didn't pick door 1, we could simply relabel the doors, or rewrite this solution with the door numbers permuted). Let S be the event of success in getting the car, and let C_j be the event that the car is behind door j. Conditioning on which door has the car, by LOTP we have

$$P(S) = P(S|C_1) \cdot \frac{1}{n} + \cdots + P(S|C_n) \cdot \frac{1}{n}.$$

Suppose you employs the non-switching strategy. The only possibility of success is that the car is indeed behind door 1, which implies

$$P_{\text{non-switching}}(S) = 1 \cdot \frac{1}{n} + 0 \cdot \frac{1}{n} + \cdots + 0 \cdot \frac{1}{n} = \frac{1}{n}.$$

Suppose you employs the switching strategy. If the car is behind door 1, then switching will fail, so $P(\text{get car}|C_1) = 0$. Otherwise, since Monty always reveals m goat, the probability of getting a car by switching to a remaining unopened door is $\frac{1}{n-m-1}$. Thus,

$$P_{\text{switching}}(S) = 0 \cdot \frac{1}{n} + \frac{1}{n-m-1} \cdot \frac{1}{n} + \cdots + \frac{1}{n-m-1} \cdot \frac{1}{n} = \frac{n-1}{(n-m-1)n}.$$

This value is greater than $\frac{1}{n}$, so you should switch.