SI 140A-02 Probability & Statistics for EECS, Fall 2024
Homework 3

Name: **Wenye Xiong**
Student ID: 2023533141

Due on Oct. 22, 2024, 11:59 UTC+8

Read all the instructions below carefully before you start working on the assignment, and before you make a submission.

- You are required to write down all the major steps towards making your conclusions; otherwise you may obtain limited points of the problem.
- Write your homework in English; otherwise you will get no points of this homework.
- Any form of plagiarism will lead to 0 point of this homework.

Problem 1

A fair die is rolled repeatedly, and a running total is kept (which is, at each time, the total of all the rolls up until that time). Let p_n be the probability that the running total is ever exactly n (assume the die will always be rolled enough times so that the running total will eventually exceed n, but it may or may not ever equal n).
(a) Write down a recursive equation for p_n (relating p_n to earlier terms p_k in a simple way). Your equation should be true for all positive integers n, so give a definition of p_0 and p_k for $k < 0$ so that the recursive equation is true for small values of n.
(b) Find p_7.
(c) Give an intuitive explanation for the fact that $p_n \to 1/3.5 = 2/7$ as $n \to \infty$.

Solution

(a):
Let p_n be the probability that the running total is ever exactly n. And we can see that to reach exactly n from a certain number, k for example, we are actually looking for the probability that the running total is ever exactly $n - k$ at some point. So we have the recursive equation:

$$p_n = \frac{1}{6} \sum_{k=1}^{6} p_{n-k}$$

where we define $p_0 = 1$ since the running total is always 0 at the beginning, and $p_k = 0$ for $k < 0$ since the running total can never be negative.
(b):
Using the recursive equation, we can calculate p_7 as follows:

$$p_1 = \frac{1}{6} \times 1 = \frac{1}{6},\ p_2 = \frac{1}{6} \times (1 + \frac{1}{6}),\ p_3 = \frac{1}{6} \times (1 + \frac{1}{6})^2$$
$$p_4 = \frac{1}{6} \times (1 + \frac{1}{6})^3,\ p_5 = \frac{1}{6} \times (1 + \frac{1}{6})^4,\ p_6 = \frac{1}{6} \times (1 + \frac{1}{6})^5$$

So we have $p_7 = \frac{1}{6} \times ((1 + \frac{1}{6})^6 - 1) \approx 0.254$.
(c):
Consider the average number added to the total each time we roll the die. The average number is $(1 + 2 + 3 + 4 + 5 + 6)/6 = 3.5$. So every two times we roll the die, we are expected to cross a range of 7, and have two lands on these 7 numbers with same probability, which is 2/7. So the probability that the running total is ever exactly n will approach 2/7 as n goes to infinity.

Problem 2

A message is sent over a noisy channel. The message is a sequence $x_1, x_2, \dots, x_n$ of n bits ($x_i \in \{0, 1\}$). Since the channel is noisy, there is a chance that any bit might be corrupted, resulting in an error (a 0 becomes a 1 or vice versa). Assume that the error events are independent. Let p be the probability that an individual bit has an error ($0 < p < 1/2$). Let $y_1, y_2, \dots, y_n$ be the received message (so $y_i = x_i$ if there is no error in that bit, but $y_i = 1 - x_i$ if there is an error there).
To help detect errors, the nth bit is reserved for a parity check: x_n is defined to be 0 if $x_1 + x_2 + \cdots + x_{n-1}$ is even, and 1 if $x_1 + x_2 + \cdots + x_{n-1}$ is odd. When the message is received, the recipient checks whether y_n has the same parity as $y_1 + y_2 + \cdots + y_{n-1}$. If the parity is wrong, the recipient knows that at least one error occurred; otherwise, the recipient assumes that there were no errors.

(a) For $n = 5, p = 0.1$, what is the probability that the received message has errors which go undetected?

(b) For general n and p, write down an expression (as a sum) for the probability that the received message has errors which go undetected.

(c) Give a simplified expression, not involving a sum of a large number of terms, for the probability that the received message has errors which go undetected.

Solution

(a):
Notice that in every situation, $\sum_{i=1}^{n} x_n$ is an even number. So if the received message has even number of errors, the parity check will pass, and the errors will remain undetected. Otherwise, the errors will be detected. So the probability that the received message has errors which go undetected is the probability that the number of errors is even.
The number of errors is a binomial distribution $Bin(n, p)$ with $n = 5$ and $p = 0.1$. So the probability that the received message has errors which go undetected is:

$$P(\text{undetected}) = \binom{5}{2} p^2(1-p)^3 + \binom{5}{4} p^4(1-p) \approx 0.073$$

(b):
The probability that the received message has errors which go undetected is the probability that the number of errors is even. And the number of errors is a binomial distribution $Bin(n, p)$ with n bits and p probability of error. So the probability that the received message has errors which go undetected is:

$$P(\text{undetected}) = \sum_{k=1}^{\frac{n}{2}} \binom{n}{2k} p^{2k}(1-p)^{n-2k}$$

(c):
Consider two terms:

$$E = \sum_{k=0}^{\frac{n}{2}} \binom{n}{2k} p^{2k}(1-p)^{n-2k}$$
$$O = \sum_{k=0}^{\frac{n-1}{2}} \binom{n}{2k+1} p^{2k+1}(1-p)^{n-2k-1}$$

Easy to find that $E + O = \sum_{k=0}^{n} \binom{n}{k} p^k(1-p)^{n-k} = 1$. And $E - O = \sum_{k=0}^{n}(-1)^k \binom{n}{k} p^k(1-p)^{n-k} = (1-2p)^n$. So $E = \frac{1+(1-2p)^n}{2}$, and the probability that the received message has errors which go undetected is $E - (1-p)^n = \frac{1+(1-2p)^n}{2} - (1-p)^n$.
Therefore, the probability that the received message has errors which go undetected is $\frac{1+(1-2p)^n}{2} - (1-p)^n$.

Problem 3

In Monty Hall problem, now suppose the car is not placed randomly with equal probability behind the three doors. Instead, the car is behind door one with probability p_1, behind door two with probability p_2, and behind door three with probability p_3. Here $p_1 + p_2 + p_3 = 1$ and $p_1 \geq p_2 \geq p_3 > 0$. You are to choose one of the three doors, after which Monty will open a door he knows to conceal a goat. Monty always chooses randomly with equal probability among his options in those cases where your initial choice is correct. What strategy should you follow?

Solution

Your strategy should alter depending on the relationship between the probabilities of having car behind the door you choose and the door left unopened.
At the beginning, we choose door one. This is because the probability of the car being behind door one is p_1, which is the largest among the three probabilities. We then assume that Monty opens door two. Let C_i be the event that the car is behind door i, G_i be the event that the Goat is behind door i, and O_i be the event that Monty opens door i. By Bayes' theorem, we have:

$$P(C_1|O_2, G_2) = P(C_1|O_2) = \frac{P(O_2|C_1)P(C_1)}{P(O_2)} = \frac{\frac{1}{2}p_1}{\frac{1}{2}p_1 + p_3} = \frac{p_1}{p_1 + 2p_3}$$

We can also get that:

$$P(C_3|O_2, G_2) = P(C_3|O_2) = \frac{2p_3}{p_1 + 2p_3}$$

So your strategy should be to switch to door three if $p_3 > \frac{p_1}{2}$, and stick to door one otherwise.
Similarly, we discuss the case where Monty opens door three. We can get that:

$$P(C_1|O_3, G_3) = P(C_1|O_3) = \frac{p_1}{p_1 + 2p_2}$$

$$P(C_2|O_3, G_3) = P(C_2|O_3) = \frac{2p_2}{p_1 + 2p_2}$$

So your strategy should be to switch to door two if $p_2 > \frac{p_1}{2}$, and stick to door one otherwise.
We can eventually get the probability of $P(C_i|M_j) = \frac{p_i}{p_i + 2p_j}$, when you choose the door i initially. The strategy is to switch to the door k if $p_k > \frac{p_i}{2}$, and stick to the door i otherwise.
And we can write the probability of success (that is you can try to switch or to stay) as $S_{i,k} = \frac{1}{2}|\frac{p_i - 2p_k}{p_i + 2p_k}| + \frac{1}{2}$.
Finally, we can get the total probability of success when choose the door i initially.

$$\begin{aligned} P(S_i) &= \frac{1}{2}((\frac{1}{2}p_i + p_k)|\frac{p_i - 2p_k}{p_i + 2p_k}| + (\frac{1}{2}p_i + p_j)|\frac{p_i - 2p_j}{p_i + 2p_j}|) + \frac{1}{2} \\ &= \frac{1}{4}(|p_i - 2p_k| + |p_i - 2p_j)| + \frac{1}{2} \end{aligned}$$

We can easily see that when $i = 3$, $P(S_i)$ has the largest value because p_3 is the smallest among the three probabilities. So you should choose door three initially.
Also, because p_3 is the smallest, so $\frac{p_3}{2} < p_1$ and $\frac{p_3}{2} < p_2$, so you should always switch to the other door. And the probability of success is $P(S_3) = p_1 + p_2$.
In conclusion, **you should choose door three initially, and always switch to the other door**. The probability of success is $p_1 + p_2$.

Problem 4

(a) 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.
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 $1 \leq m \leq n-2$.

Solution

(a):
We first talk about the probability of success if you stick to the door you initially choose. Let S be the event that we successfully get the car. We have a prior probability $P(S) = \frac{1}{7}$. And we update it after Monty opens 3 goat doors, let's say they are door i,j,k with $2 \leq i \leq j \leq k \leq 7$. Define O_{ijk} be the event that Monty opens door i,j,k. We have:

$$P(S) = \sum_{i,j,k} P(S|O_{ijk})P(O_{ijk})$$

By symmetry, we have $P(S|O_{ijk}) = \frac{1}{7}$ for all $2 \leq i \leq j \leq k \leq 7$
Then we talk about the probability of success if you switch to one of the remaining 3 doors. We now know that the probability of the car being behind the door you initially choose is 1/7, therefore the probability of the car being behind one of the remaining 3 doors is 6/7. Because of the symmetry, the probability of the car being behind each of the remaining 3 doors is 2/7. So the probability of success if you switch to one of the remaining 3 doors is 2/7.
Therefore, you should switch to one of the remaining 3 doors, and the probability of success if you switch to one of the remaining 3 doors is 2/7.
(b):
This part is nearly the same as what we did in part (a). We first talk about the probability of success if you stick to the door you initially choose. Let S be the event that we successfully get the car. We have a prior probability $P(S) = \frac{1}{n}$. And we update it after Monty opens m goat doors, let's say they are door $i_1, i_2, \cdots, i_m$ with $1 \leq i_1 < i_2 < \cdots < i_m \leq n$. Define $O_{i_1 i_2 \cdots i_m}$ be the event that Monty opens door $i_1, i_2, \cdots, i_m$. We have:

$$P(S) = \sum_{i_1, i_2, \cdots, i_m} P(S|O_{i_1 i_2 \cdots i_m})P(O_{i_1 i_2 \cdots i_m})$$

By symmetry, we have $P(S|O_{i_1 i_2 \cdots i_m}) = \frac{1}{n}$ for all $1 \leq i_1 < i_2 < \cdots < i_m \leq n$
Then we talk about the probability of success if you switch to one of the remaining $n-m$ doors. We now know that the probability of the car being behind the door you initially choose is $1/n$, therefore the probability of the car being behind one of the remaining $n-m$ doors is $(n-1)/n$. Because of the symmetry, the probability of the car being behind each of the remaining $n-m$ doors is $(n-1)/n(n-m)$. So the probability of success if you switch to one of the remaining $n-m$ doors is $(n-1)/n(n-m)$.
Compare $\frac{1}{n}$ with $\frac{n-1}{n(n-m)}$, we have $\frac{1}{n} \leq \frac{n-1}{n(n-m)}$ when $m \geq 1$, so you should switch to one of the remaining $n-m$ doors, and the probability of success if you switch to one of the remaining $n-m$ doors is $(n-1)/n(n-m)$.

Problem 5

A/B testing is a form of randomized experiment that is used by many companies to learn about how customers will react to different treatments. For example, a company may want to see how users will respond to a new feature on their website (compared with how users respond to the current version of the website) or compare two different advertisements.

As the name suggests, two different treatments, Treatment A and Treatment B, are being studied. Users arrive one by one, and upon arrival are randomly assigned to one of the two treatments. The trial for each user is classified as "success" (e.g., the user made a purchase) or "failure". The probability that the nth user receives Treatment A is allowed to depend on the outcomes for the previous users. This set-up is known as a *two-armed bandit*.

Many algorithms for how to randomize the treatment assignments have been studied. Here is an especially simple (but fickle) algorithm, called a "stay-with-a-winner" procedure:

(i) Randomly assign the first user to Treatment A or Treatment B, with equal probabilities.

(ii) If the trial for the nth user is a success, stay with the same treatment for the $(n+1)$ st user; otherwise, switch to the other treatment for the $(n+1)$st user.

Let a be the probability of success for Treatment A, and b be the probability of success for Treatment B. Assume that $a \neq b$, but that a and b are unknown (which is why the test is needed). Let p_n be the probability of success on the nth trial and a_n be the probability that Treatment A is assigned on the nth trial (using the above algorithm).

(a) Show that

$$p_n = (a-b)a_n + b, a_{n+1} = (a+b-1)a_n + 1 - b$$

(b) Use the results from (a) to show that p_{n+1} satisfies the following recursive equation:

$$p_{n+1} = (a+b-1)p_n + a + b - 2ab$$

(c) Use the result from (b) to find the long-run probability of success for this algorithm, $\lim_{n\to\infty} p_n$, assuming that this limit exists.

Solution

(a):

Because we only have two treatments, we have $b_n = 1 - a_n$. And we have

$$p_n = a_n a + b_n b = a_n a + (1-a_n)b = (a-b)a_n + b.$$

We can also write a recursive equation for a_n because the probability of taking Treatment A on the $n+1$st trial depends on the outcome of the nth trial. Define A_n be the event that Treatment A is assigned on the nth trial, and B_n be the event that Treatment B is assigned on the nth trial. We have:

$$a_{n+1} = P(success|A_n)P(A_n) + P(failure|B_n)P(B_n) = a \cdot a_n + (1-b) \cdot (1-a_n) = (a+b-1)a_n + 1 - b$$

(b):

Combining the two equations we got in (a), we have:

$$\begin{aligned} p_{n+1} &= (a-b)a_{n+1} + b = (a-b)((a+b-1)a_n + 1 - b) + b = (a-b)((a+b-1)\tfrac{p_n - b}{a-b} + 1 - b) + b \\ &= (a+b-1)(p_n - b) + (a-b)(1-b) + b = (a+b-1)p_n + a + b - 2ab \end{aligned}$$

(c):

Let $p = \lim_{n\to\infty} p_n$, then we have:

$$\begin{aligned} p &= (a+b-1)p + a + b - 2ab \\ p &= \tfrac{a+b-2ab}{2-a-b} \end{aligned}$$

Problem 6

(**Optional Challenging Problem I**) By LOTP for problems with recursive structure, we generate many different equations.

(a) Explain the principle behind the method of characteristic equation.

(b) Solve the following difference equation:

$$p \cdot f_{i+1} - f_i + q \cdot f_{i-1} = -1, 1 \leq i \leq N-1$$

where $0 < p < 1, q = 1-p, N$ is a constant, $f_0 = 0, f_N = 0$.

(c) Solve the following difference equation:

$$f_{i+1} = b \cdot f_i + a \cdot f_{i-1} + h, i \geq 1$$

where h is a constant.

(d) Solve the following difference equation:

$$f_{i+1} = b \cdot f_i + a \cdot f_{i-1} + g(i), i \geq 1$$

where $g(i)$ is a function of i.

Problem 7

Optional Challenging Problem II

(a) An event E_{n+1} is mutually independent of the set of events $E_1, \ldots, E_n$ if for any subset $I \subseteq [1, n]$

$$P\left(E_{n+1} \mid \bigcap_{j \in I} E_j\right) = P\left(E_{n+1}\right)$$

(b) A dependence graph for the set of events $E_1, \ldots, E_n$ is a graph $G = (V, E)$ such that $V = \{1, \ldots, n\}$, and for $i = 1, \ldots, n$, event E_i is mutually independent of the events $\{E_j \mid (i,j) \notin E\}$.

(c) Assume there exist real numbers $x_1, \ldots, x_n \in [0, 1]$ such that, for any $i (1 \leq i \leq n)$,

$$P\left(E_i\right) \leq x_i \prod_{j:(i,j) \in E} \left(1 - x_j\right)$$

Then show the following inequality hold:

$$P\left(\bigcap_{i=1}^{n} E_i^c\right) \geq \prod_{i=1}^{n} \left(1 - x_i\right)$$

(d) Find the possible applications of the above inequality in the field of EECS.