2022Fall Probability & Mathematical Statistics	2022/09/25

Homework 3

Professor: Ziyu Shao | *Due: 2022/10/09 10:59pm*

1. Fred decides to take a series of n tests, to diagnose whether he has a certain disease (any individual test is not perfectly reliable, so he hopes to reduce his uncertainty by taking multiple tests). Let D be the event that he has the disease, $p = P(D)$ be the prior probability that he has the disease, and $q = 1 - p$. Let T_j be the event that he tests positive on the jth test.

   (a) Assume for this part that the test results are conditionally independent given Fred's disease status. Let $a = P(T_j \mid D)$ and $b = P(T_j \mid D^c)$, where a and b don't depend on the jth test. Find the posterior probability that Fred has the disease, given that he tests positive on all n of the n tests.

   (b) Suppose that Fred tests positive on all n tests. However, some people have a certain gene that makes them always test positive. Let G be the event that Fred has the gene. Assume that $P(G) = 1/2$ and that D and G are independent. If Fred does not have the gene, then the test results are conditionally independent given his disease status. Let $a_0 = P(T_j \mid D, G^c)$ and $b_0 = P(T_j \mid D^c, G^c)$, where a_0 and b_0 don't depend on j. Find the posterior probability that Fred has the disease, given that he tests positive on all n of the tests.

2. 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.

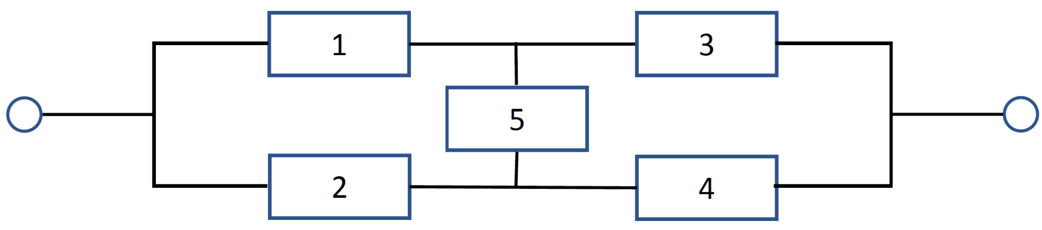

Figure 1: An illustration of the system composed of 5 homogeneous devices.

3. 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 $n \to \infty$.

4. A sequence of $n \geq 1$ independent trials is performed, where each trial ends in "success" or "failure" (but not both). Let p_i be the probability of success in the i^{th} trial, $q_i = 1 - p_i$, and $b_i = q_i - 1/2$, for $i = 1, 2, \ldots, n$. Let A_n be the event that the number of successful trials is even.

(a) Show that for $n = 2$, $P(A_2) = 1/2 + 2b_1b_2$.

(b) Show by induction that

$$P(A_n) = 1/2 + 2^{n-1}b_1b_2 \ldots b_n$$

(This result is very useful in cryptography. Also, note that it implies that if n coins are flipped, then the probability of an even number of Heads is 1/2 if and only if at least one of the coins is fair.) *Hint*: Group some trials into a super-trial.

(c) Check directly that the result of (b) is true in the following simple cases: $p_i = 1/2$ for some i; $p_i = 0$ for all i; $p_i = 1$ for all i.

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.

6. (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 $1 \leq m \leq n-2$.

7. Consider the Monty Hall problem, except that Monty enjoys opening door 2 more than he enjoys opening door 3, and if he has a choice between opening these two doors, he opens door 2 with probability p, where $\frac{1}{2} \leq p \leq 1$.
To recap: there are three doors, behind one of which there is a car (which you want), and behind the other two 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, which for concreteness we assume is door 1. Monty Hall then opens a door to reveal a goat, and offers you the option of switching. Assume that Monty Hall knows which door has the car, will always open a goat door and offer the option of switching, and as above assume that if Monty Hall has a choice between opening door 2 and door 3, he chooses door 2 with probability p ($\frac{1}{2} \leq p \leq 1$).

(a) Find the unconditional probability that the strategy of always switching succeeds (unconditional in the sense that we do not condition on which of doors 2 or 3 Monty opens).

(b) Find the probability that the strategy of always switching succeeds, given that Monty opens door 2 (Assume we always choose door 1 first).

(c) Find the probability that the strategy of always switching succeeds, given that Monty opens door 3 (Assume we always choose door 1 first).

8. Let X be the number of purchases that Fred will make on the online site for a certain company (in some specified time period). Suppose that the PMF of X is $P(X = k) = e^{-\lambda}\lambda^k/k!$ for $k = 0, 1, 2, \cdots$.

   (a) Find $P(X \geq 1)$ and $P(X \geq 2)$ without summing infinite series.

   (b) Suppose that the company only knows about people who have made at least one purchase on their site (a user sets up an account to make a purchase, but someone who has never made a purchase there doesn't appear in the customer database). If the company computes the number of purchases for everyone in their database, then these data are draws from the conditional distribution of the number of purchases, given that at least one purchase is made. Find the conditional PMF of X given $X \geq 1$. (This conditional distribution is called a truncated Poisson distribution.)

9. A message is sent over a noisy channel. The message is a sequence $x_1, x_2, \ldots, x_n$ 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 $y_1, y_2, \ldots, y_n$ 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.

10. For x and y binary digits (0 or 1), let $x \bigoplus y$ be 0 if $x = y$ and 1 if $x \neq y$ (this operation is called exclusive or (often abbreviated to XOR), or addition mod 2).

    (a) Let $X \sim \text{Bern}(p)$ and $Y \sim \text{Bern}(1/2)$, independently. What is the distribution of $X \bigoplus Y$?

(b) With notation as in sub-problem (a), is $X \bigoplus Y$ independent of X? Is $X \bigoplus Y$ independent of Y? Be sure to consider both the case $p = 1/2$ and the case $p \neq 1/2$.

(c) Let $X_1, \ldots, X_n$ be i.i.d. (*i.e.*, independent and identically distributed) Bern(1/2) R.V.s. For each nonempty subset J of $\{1, 2, \ldots, n\}$, let

$$Y_J = \bigoplus_{j \in J} X_j.$$

Show that $Y_J \sim \text{Bern}(1/2)$ and that these $2^n - 1$ R.V.s are pairwise independent, but not independent.