EECS 70 Discrete Mathematics and Probability Theory
Fall 2015 Jean Walrand, Satish Rao

HW 9

Due Wednesday Oct 28 at 10PM

1. Box of marbles

You are given two boxes: one of them containing 900 red marbles and 100 blue marbles, the other one contains 500 red marbles and 500 blue marbles.

1. If we pick one of the boxes randomly, and pick a marble what is the probability that it is blue?

   Solution: Let B be the event that the picked marble is blue, R be the event that it is red, A_1 be the event that the marble is picked from box 1, and A_2 be the event that the marble is picked from box 2. Therefore we want to calculate $Pr(B)$. By total probability

$$Pr(B) = Pr(B|A_1)P(A_1) + Pr(B|A_2)Pr(A_2) = 0.5 \times 0.1 + 0.5 \times 0.5 = 0.3$$

2. If we see that the marble is blue, what is the probability that it is chosen from box 1?

   Solution: In this part, we want to find $Pr(A_1|B)$. By Bayes' rule

$$Pr(A_1|B) = \frac{Pr(B|A_1)Pr(A_1)}{Pr(B|A_1)Pr(A_1) + Pr(B|A_2)Pr(A_2)} = \frac{0.1 \times 0.5}{0.5 \times 0.1 + 0.5 \times 0.5} = \frac{1}{6}$$

3. Suppose we pick one marble from box 1 and without looking at its color we put it aside. Then we pick another marble from box 1. What is the probability that the second marble is blue?

   Solution: Let B_1 be the event that first marble is blue, R_1 be the event that the first marble is red, and B_2 be the event that second marble is blue without looking at the color of first marble. We want to find $Pr(B_2)$. By total probability,

$$Pr(B_2) = Pr(B_2|B_1)Pr(B_1) + Pr(B_2|R_1)Pr(R_1) = \frac{99}{999} \times 0.1 + \frac{100}{999} \times 0.9 = 0.1$$

   More generally, one can see that the probability that the n-th marble picked from box 1 is blue with probability 0.1. This is clear by symmetry: all the permutations of the 1000 marbles have the same probability, so the probability that the n-th marble is blue is the same as the probability that the first marble is blue.

2. Erasures and probabilities

In this problem, we will consider the erasure correction scheme that you have seen before, but add a probability twist to it.

As usual, let's say Alice wants to send a message of length n to Bob, and she transmits $n+k$ packets instead. Bob can recover Alice's message if and only if the number of erased packets is at most k.

For this problem, you may assume that $n = 10$ and $k = 4$.

a) Let's say that each packet sent by Alice gets erased with a probability $p = 0.1$. What are the chances that, of the 14 packets that Alice transmits, all arrive intact? All but 1? All but 2?

Solution: Clearly, the probability that the first packet arrives intact is 0.9. Similarly, the probability that the second packet arrives intact is also 0.9. And so on all the way up to the 14^{th} packet. Thus, the probability that all 14 packets arrive intact is 0.9^{14}, or about a 22.88% chance.

Now let's try and find the chances that *all but one* packets arrive intact. That is, 13 packets arrive intact and 1 gets erased. Clearly, there are 14 possibilities for the packet that gets erased (it could be any of the 14 packets that Alice transmits). Each of these mutually disjoint cases has a probability $0.9^{13} \times 0.1$. Thus, the total probability that all but one packets arrive intact is $14 \times 0.9^{13} \times 0.1$, or about a 35.59% chance.

Now let's try and find the chances that *all but two* packets arrive intact. That is, 11 packets arrive intact and 2 get erased. Clearly, there are $\binom{14}{2}$ possibilities for the packets that get erased. Each of these mutually disjoint cases has a probability $0.9^{12} \times 0.1^2$. Thus, the total probability that all but two packets arrive intact is $\binom{14}{2} \times 0.9^{12} \times 0.1^2$, which represents about a 25.7% chance.

b) Under the assumptions of part (a) above, what are the chances that Bob will be able to successfully recover Alice's message?

Solution: Clearly, Bob can recover Alice's message as long as the number of erased packets is less than or equal to 4. Following the method outlined in the previous part, the odds of *exactly* k packets getting erased is $\binom{14}{k} \times 0.9^{(14-k)} \times 0.1^k$. Since our k can take any value such that $0 \leq k \leq 4$, the odds that Bob will be able to recover Alice's message are given by:

$$\Pr(\text{Bob recovers Alice's message}) = \sum_{k=0}^{4} \left[\binom{14}{k} \times 0.9^{(14-k)} \times 0.1^k \right],$$

which evaluates to a likelihood of about 99.077%.

c) **You only need to do either the original 2c or the alternate 2c!**

original 2c)

Now let's make our probabilistic model somewhat more complicated. Let's assume that erasures always occur in consecutive pairs. That is, you can't have a single erasure. Neither can you have a consecutive run of erasures of odd length. The rule is that each erasure is paired up with another. That is, whenever packet i is erased, either it is the first erasure in a pair (meaning that packet $i+1$ is also erased) or it is the second erasure in a pair (meaning that packet $i-1$ is also erased). The chances of a packet getting erased, assuming it's the first erasure in the pair, remain the same at 0.1. Now what are the chances that Bob will be able to recover Alice's message successfully?

Solution:

Due to the confusing phrasing of the question, we will be very lenient with grading this question

Original version: packet 14 cannot be dropped alone

Packet 14 cannot be erased on its own.

The probability model can be formulated as follows:

Let H(i) be the event that packet i is the first of a pair of erasures. Then

P(H(1)) = 0.1

P(H(2) | H(1)) = 0

P(H(2) | - H(1)) = 0.1

...

P(H(k) | H(k-1)) = 0
P(H(k) | - H(k-1)) = 0.1 for k = 2...13
P(H(14)) = 0 always

This means we always get an even number of erasures. We consider the following 3 cases:

Case 1: 0 erasures

In this case, no packets are erased, so the probability is 0.9^{13}. The number 13 comes from packets 1...13 all not being erased. The $14th$ packet cannot be dropped alone so we do not factor it into our probability.

Case 2: 2 erasures

In this case, we have 2 consecutive packets which are erased, and 12 which are not. Thus, for any specific outcome where packet 14 is not erased, e.g. "packets 3 and 4 erased, everything else not erased", we have probability of $0.9^{11} * 0.1$. However, if packet 14 is erased, i.e. packets 13 and 14 erased, then the probability is $0.9^{12} * 0.1$

Next, we need to sum over all outcomes with 2 erasures. The "first erasure" can happen at positions 1...13, so 13 choices in total. Thus, $P(2 \text{ erasures}) = 12 * 0.9^{11} * 0.1 + 1 * 0.9^{12} * 0.1$

Case 3: 4 erasures

In this case, we have 2 sets of 2 consecutive erasures. The probability of a specific outcome, assuming packet 14 is not erased, i.e. "packets 5,6, 8,9 erased, everything else not erased" is $0.9^9 * 0.1^2$. However, if packet 14 is erased, then the probability is $0.9^{10} * 0.1^2$

Next, we need to count the numbers for each outcome. To do this, we will represent each outcome as a tuple of 3 numbers (x,y,z). Where x is number of unerased packets before the first erased pair, y is number of unerased packets between the second and third pair, and z is number of unerased packets after the third pair. There is a bijection between the set $\{(x,y,z) : x \geq 0, y \geq 0, z \geq 0, x+y+z = 10\}$ and the set of all possible outcomes with 4 erasures. This means that we only need to count the number of tuples (x,y,z).

But notice this is just stars and bars, where we are dividing 10 objects among 3 buckets (x,y,z). Thus, there are $\binom{12}{2} = 66$ possible tuples.

We are not done yet! We still need to find out which of these have packet 14 erased. There are 11 such outcomes, corresponding to the 13,14 erased, and another erasure pair starting at any of 1...11. Thus, $P(4 \text{ erasures}) = 55 * 0.9^9 * 0.1^2 + 11 * 0.9^{10} * 0.1^2$

Thus, the total probability of being able to recover the message is the sum over 3 cases, which is $0.9^{13} + 12 * 0.9^{11} * 0.1 + 1 * 0.9^{12} * 0.1 + 55 * 0.9^9 * 0.1^2 + 11 * 0.9^{10} * 0.1^2 \approx 0.910$

alternate 2c)

Alice sends consecutive packets to Bob. With probability p, the channel erases the first and second packets. and the channel moves to the third packet. With probability $q = 1 - p$, the first packet is not erased and the channel moves to the second packet. The process repeats forever.

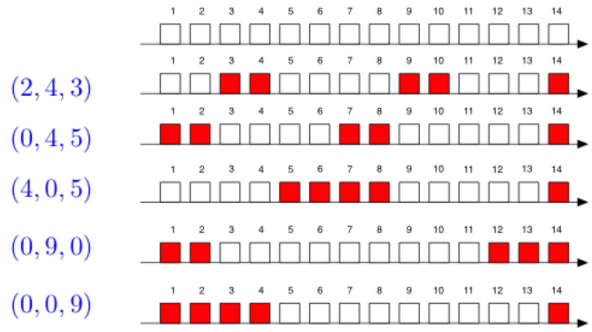
Thus, the channel erases packets in pairs. Let x,y,z be the gaps between erasures. Thus, if $(x,y,z) = (0,3,1)$, packets 1, 2 are erased (zero gap before the first erasure), then packets 6, 7 are erased (gap of three packets not erased after packet 2), then packets 9, 10 are erased, and so on.

What is the probability that more than four packets are erased among the first 14 packets?

Solution:

Alternative Version: packet 14 can be dropped alone

Solution 1 Under this interpretation, we need to consider 3 more cases. We also need to change our original cases.



Case 1: 0 erasures In this case, no packets are erased, so the probability is 0.9^{14}. The number 14 comes from packets 1...14 all not being erased. The $14th$ packet is considered this time because it can be erased.

Case 2: 2 erasures

Since packet 14 can be erased alone, we have probability $0.9^{12} * 0.1$. We have possible position for this pair of erasure, ranging from (1,2)... (13,14), so total probability is $0.9^{12} * 0.1 * 13$

Case 3: 4 erasures

Since packet 14 can be erased alone, we have probability $P(4 \text{ erasures}) = 0.9^{10} * 0.1^2 * 66$, where 0.1^2 is due to 2 pairs of erasures, and 0.9^{10} is due to 10 packets not beinge rased.

We need to count the numbers for each outcome. To do this, we will represent each outcome as a tuple of 3 numbers (x, y, z). Where x is number of unerased packets before the first erased pair, y is number of unerased packets between the second and third pair, and z is number of unerased packets after the third pair. There is a bijection between the set $\{(x,y,z) : x \geq 0, y \geq 0, z \geq 0, x+y+z = 10\}$ and the set of all possible outcomes with 4 erasures. This means that we only need to count the number of tuples (x, y, z).

But notice this is just stars and bars, where we are dividing 10 objects among 3 buckets (x,y,z). Thus, there are $\binom{12}{2} = 66$ possible tuples.

Case 4: 1 erasure

An odd-number of erasures happen if and only if packet 14 is dropped, so this case has probability $0.9^{13} * 0.1$

Case 5: 3 erasures

This amounts to packet 14 being the "head" of a pair of erasures and 1 addition pair being erased. Thus, for any specific outcome, e.g. "packets 3 and 4 erased, packet 14 erased, everything else not erased", we have probability of $0.9^{11} * 0.1^2$.

Next, we need to sum over all outcomes. The "first erasure" can happen at positions 1...12, so 12 choices in total. Thus, $P(2 \text{ erasures}) = 0.9^{11} * 0.1^2 * 12$

The total probability of being able to recover the message is:

P(Case 1)+ P(Case 2) + P(Case 3) + P(Case 4) + P(Case 5)

= $0.9^{14} + 0.9^{12} * 0.1 * 13 + 0.9^{10} * 0.1^2 * 66 + 0.9^{13} * 0.1 + 0.9^{11} * 0.1^2 * 12$

= 0.88913

Solution 2

Let us draw a few pictures. (For me, that is always a good idea. A picture is worth a thousand words.)

These are examples of situations where 5 packets were erased among the first 14. The values (x, y, z)

are shown in the figure.

The examples show that more than four packets are erased among the first 14 packets if and only if $x+y+z \leq 9$.

Let's make sure that this is correct and that we did not miss some cases. Given (x,y,z), the first two erasures are packets $\{x+1, x+2\}$, the next two erasures are packets $\{x+2+y+1, x+2+y+2\}$, and the fifth erasure is packet $x+2+y+2+z+1 = x+y+z+5$. Thus, the fifth erasure is packet 14 or some previous packet if and only if

$$x+y+z+5 \leq 14, \text{ i.e., if and only if } x+y+z \leq 9.$$

Hence, the probability that more than four packets are erased among the first 14 packets is the probability that $x+y+z \leq 9$.

The probability of (x,y,z) is $q^x p \times q^y p \times q^z p$. Indeed, this expression is the probability that the first x packets are successful, than the next packet is erased, and so on. Hence, the probability that there are more than 4 erasures among the first 14 packets is given by

$$\sum_{x+y+z \leq 9} q^{x+y+z} p^3 = \sum_{n=0}^{9} a(n) q^n p^3$$

where $a(n) = \binom{n+2}{2}$ is the number of ways of splitting n among x,y,z. Hence, the required probability is

$$\sum_{n=0}^{9} \frac{(n+1)(n+2)}{2} q^n p^3.$$

We find, for $q = 0.9$, that this probability is 0.1109, which gives a success probability of 88.91%.

3. Clinical Trials

You are creating a test for a rare disease that only 1 in 1000 people has. If a person has the disease, there is a 95% chance that your test will be positive. But if the person does not have the disease, there is only an 85% chance the test will be negative.

a) Let D be the event that you have the disease, and H be the event that you are healthy. Let A be the event that the test comes out positive, and B be the event that it comes out negative. Write an expression for $\mathbb{P}(D|A)$ (the probability you have the disease given a positive test result) in terms of the probabilities given above. Plug in the given probabilities into your expression and calculate the numerical value of $\mathbb{P}(D|A)$.

Solution: We know from Bayes rule that

$$P(D|A) = \frac{P(D)P(A|D)}{P(A)}.$$

We know that $P(D) = 0.001$ and $P(A|D) = 0.95$, but it is unclear how to compute $P(A)$. However, we do know that the probability of a positive test result should be the probability of a positive test result and disease, plus the probability of a positive test result and healthy. This gives us

$$P(A) = P(A,D) + P(A,H) = P(D)P(A|D) + P(H)P(A|H).$$

which we have expanded above using Bayes rule. Now we have $P(A)$ in term of quantities we know, and can write

$$P(D|A) = \frac{P(D)P(A|D)}{P(D)P(A|D)+P(H)P(A|H)} = \frac{0.001\cdot 0.95}{0.001\cdot 0.95+0.999\cdot 0.15} = 0.62997\%,$$

so even given a positive test result, we get a very very small probability of disease.

b) Write an expression for $\mathbb{P}(H|B)$, the probability you are healthy given a negative test result. Evaluate the numerical result of this expression as well.

Solution: Similarly, we get

$$P(H|B) = \frac{P(H)P(B|H)}{P(B)} = \frac{P(H)P(B|H)}{P(H)P(B|H)+P(D)P(B|D)} = \frac{0.999\cdot 0.85}{0.999\cdot 0.85+0.001\cdot 0.05} = 99.9941\%.$$

c) For your test to gain approval, the chance of disease given a positive test result must be above 90%. What would the accuracy of the test have to be to ensure this result? (You may now assume the accuracy of the test is the same whether you have the disease or not. I.e. accuracy of test = $P(A|D) = P(\neg A|\neg D)$)

Solution: Solve for x:

$$P(D|A) = \frac{0.001\cdot x}{0.001\cdot x+0.999\cdot(1-x)} = 0.9 \quad \Rightarrow \quad x = 99.9889\%$$

4. Peaceful rooks

A friend of yours, Eithen Quinn, is fascinated by the following problem: placing m rooks on an $n\times n$ chessboard, so that they are in peaceful harmony (i.e. no two threaten each other). Each rook is a chess piece, and two rooks threaten each other if and only if they are in the same row or column. You remind your friend that this is so simple that a baby can accomplish the task. You forget however that babies cannot understand instructions, so when you give the m rooks to your baby niece, she simply puts them on random places on the chessboard. She however, never puts two rooks at the same place on the board.

1. Assuming your niece picks the places uniformly at random, what is the chance that she actually accomplishes the task and does not prove you wrong?

   Solution: After having placed i rooks in a peaceful position, i of the rows and i of the columns are taken. So for the next rook we have $n-i$ choices for the row and $n-i$ choices for the column in order to remain in a peaceful position. The total number of board cells left is n^2-i. So the chance that the next rook keeps the peace is $\frac{(n-i)^2}{n^2-i}$.
   The product over $i=0,\ldots,m-1$ gives us the final answer. So the answer is

$$\prod_{i=0}^{m-1}\frac{(n-i)^2}{n^2-i} = \frac{(n!)^2(n^2-m)!}{(n^2)!((n-m)!)^2}$$

2. If you were using checker pieces as a replacement for rooks (so that they can be stacked on top of each other), then what would be the probability that your niece's placements result in peace? Assume that two pieces stacked on top of each other threaten each other.

Solution: The only thing that changes from the previous part is that when placing the i-th piece, we no longer have $n^2 - i$ possibilities, but n^2 possibilities. So the answer changes to

$$\prod_{i=0}^{m-1} \frac{(n-i)^2}{n^2} = \frac{(n!)^2}{((n-m)!)^2 n^{2m}}$$

5. **Monty Hall Again** In the three-door Monty Hall problem, there are two stages to the decision, the initial pick followed by the decision to stick with it or switch to the only other remaining alternative after the host has shown an incorrect door. An extension of the basic problem to multiple stages goes as follow.
Suppose there are four doors, one of which is a winner. The host says: "You point to one of the doors, and then I will open one of the other non-winners. Then you decide whether to stick with your original pick or switch to one of the remaining doors. Then I will open another (other than the current pick) non-winner. You will then make your final decision by sticking with the door picked on the previous decision or by switching to the only other remaining door.

1. How many possible strategies are there?

   Solution: In the original Monty Hall problem, there was only two strategies - the sticking strategy where you would stick with your original pick after Monty opens the first door, or the switching strategy, where you would change your original pick after Monty opens the first door. Now, there are two stages where a decision needs to be made, and at each stage there are two choices. Either you can stick with your current door, or you can switch to another door (at stage one where there are two possible doors to choose from, choose one of them randomly). So, in all, there are $2 \cdot 2 = 4$ different strategies.

2. Find the best strategy and compute its probability of winning. You can do this using any method you want. Enumeration is a valid approach, but a less tedious method is to consider a simpler problem: what if you have a 3-door monty hall problem, except the probability of picking the right door at first is not $\frac{1}{3}$ but some general value p? What is the best strategy if $p = 0$? What is the best strategy if $p = 1$? Having done that, how would you reduce the 4-door problem to the above 3-door problem?

   Solution:

   Solution 1

   We calculate probability of winning given that we play with a specific strategy. We use RRR to denote picking the right door all 3 times. WRW then means picking the wrong door for the first time, right for the second time, and wrong for the third time. Thus, we win if the third letter is R.

   Note the notation of P(WWW ; S1), which means "probability of WWW under strategy S1", instead of P(WWW | S1), i.e. "probability of WWW conditioned on S1", because S1 is not random.

   S1: Stick and stick strategy

   Case 1 (RRR): Pick the right door at the beginning and stick with it, so we pick the right door all three times. **P(RRR ; S1) = P(pick the right door at the beginning) = 1/4**

   Case 2 (WWW): Pick one of the wrong doors at the beginning. **P(WWW ; S1) = 3/4**

   Notice that the sum of the two cases is 1, so we do not miss any case.

   The reason that we only need to consider 2 cases is because we stick to the same door throughout, so we either always picked the right door, or always picked the wrong door.

   P(win ; S1) = P(RRR;S1) = 1/4

   S2: Stick and switch strategy

   Case 1 (RRW): pick the right one at the beginning. **P(RRW ; S2) = 1/4**

Case 2 (WWR): Pick one of the wrong doors at the beginning. Then, the host will open another wrong one. We first stick with our door, so the host will open yet another wrong door. Thus, we are left with the right door to switch to. In short, if we pick a wrong one at the beginning, it is guaranteed that we will pick the right one at the end with this strategy. **P(WWR; S2) = P(pick a wrong door at the beginning) = 3/4**.

P(win ; S2) = P(WWR;S2) = 3/4

As an example, we will formally show the derivation of P(RRW; S2):

- P(RR ; S2) = P(R ; S2) = P(pick the right door at the beginning) = 1/4 since we are sticking the first round (so picking the right door at first = picking the right door the first 2 times), and the probability is just 1/4.

- P(RRW | RR; S2) = 1, because conditioned on us picking the right door, the only other door that remains closed must be wrong, so we are guaranteed to pick the wrong door if we switch.

- P(RRW; S2) = P(RR; S2)*P(RRW|RR; S2) = 1/4 * 1 = 1/4 by bayes rule.

For P(WWR; S2): - P(WW; S2) = P(W; S2) = 3/4, once again because we stick the first - P(RRW | RR; S2) = 1, because conditioned on us picking the right door, the only other door that remains closed must be right, so we are guaranteed to pick the right door if we switch.

- P(WWR; S2) = P(WWR|WW; S2) * P(WW; S2) = 1 * 3/4 = 3/4 by bayes rule.

Similar reasoning applies to the other cases.

S3: Switch and stick strategy

Case 1: **P(RWW ; S3) = 1/4**

Case 2: **P(WWW ; S3) = 3/4 * 1/2**. This comes from 3/4 probability of picking one of the wrong doors at first. Host opens another wrong door. Now there are 2 doors left, so the probability of picking the wrong door the second time is 1/2, then we stick with the wrong door.

Case 3: **P(WRR ; S3) = 3/4 * 1/2**. Follow the similar logic as case 2.

P(win ; S3) = P(WRR ; S3) = 3/8

S4: Switch and Switch strategy

Case 1: **P(RWR ; S4) = 1/4**, which is the probability of picking the right door at the first time. If we pick the right one the first time, host opens one of the wrong doors. We then switch to another wrong door. Host opens the last remaining wrong door. Finally, we can only switch to the original door, which is the right door.

Case 2: **P(WWR ; S4) = 3/4 * 1/2**. This come from 3/4 probability of picking one of the wrong doors at first. Host opens another wrong door. There are 2 doors left, so the probability of picking the wrong door the second time is 1/2. Since our second pick is wrong, the host must open the door we pick the first time, so we will switch to the only one door left, which is the right door.

Case 3: **P(WRW ; S4) = 3/4 * 1/2**. Follow the similar logic as case 2.

P(win ; S4) = P(RWR ; S4) + P(WWR ; S4) = 5/8

Thus, stick-and-switch strategy is the best.

Solution 2

Let Alice be a player and Bob be a host. There are four doors: 1, 2, 3, and 4. We let the first door Alice picks be door 1 without loss of generality.

Assume Alice always sticks to door 1. She has a probability $1/4$ of winning.

Assume Alice sticks to door 1 and then switches the second time. After Alice has picked door 1 twice, Bob will open two other doors that do not have a prize. Since the probability of the prize not behind door 1 is $3/4$, and we know that the other two doors do not have a prize for sure, the other door left thus has $3/4$ probability of having the prize behind it. Therefore, Alice has $3/4$ probability of winning.

Assume Alice switches to one of the two other doors that Bob did not open. The probability that she picks the door that has the prize is $(3/4) \times (1/2)$. Indeed, this happens if the prize is not behind door 1 (probability $3/4$) and if she also picks the one door out of two that has the prize (probability $1/2$).

If Alice sticks to her second choice, she then has a probability $(3/4) \times (1/2) = 3/8$ of winning.

Assume she switches again after Bob shows her a second door that has no prize. The claim is that she wins with probability $5/8$. To see this, note that she always wins if her first choice was the door with the prize. Indeed, say the prize was behind door 1 that Alice picks. Bob opens door 4, Alice switches to door 2, Bob opens door 3, Alice switches back to door 1. Also, we claim that if Alice had not picked the correct door, then she wins with probability $1/2$. Indeed, say the prize is not behind door 1 that Alice first picks. Bob opens one of the doors (2, 3, 4), say door 4. Alice picks door 2. With probability $1/2$, the prize is behind door 2 and Bob opens door 3, Alice switches to door 1 and loses. With probability $1/2$, the prize is behind door 3, Bob opens door 1, Alice switches to door 3 and wins. Thus, if she switches twice, Alice wins with probability $1/4 + (3/4)(1/2) = 5/8$.

Among these four strategies, stick and switch strategy has the highest probability of winning.