EE 126 Spring 2006 Midterm #1
Thursday, February 23, 11:10-12:30pm

SOLUTIONS

Problem 1: (14 points)

Since there is no direct flight from San Diego (S) to New York (N), every time Alice wants to go to the New York, she has to stop in either Chicago (C) or Denver (D). Due to bad weather conditions, both the flights from S to C *and* the flights from C to N have independently a delay of 1 hour with probability p. Similarly, at Denver airport, both incoming and outgoing flights are independently subject to a 2 hour delay with probability q. On any given occasion, Alice chooses randomly between the Chicago or Denver routes with equal probability.

(a) (2pt) What is the average total delay (across both legs of the overall trip) that she experiences in going from S to N?

(b) (3pt) Suppose Alice arrives at N with a delay of two hours. What is the probability that she flew through C?

(c) (3pt) Suppose that Alice wants to maximize the probability that she arrives in New York with a total delay < 2 hours. Under what conditions on p and q is going via Chicago a better choice than going via Denver?

(d) (3pt) Suppose now that Alice always flies through C. On average, how many trips does she make before experiencing a 2 hour delay?

(e) (3pt) Suppose now that the flight between S and D is known to be delayed, but Alice still randomly flies either via C or D with equal probability. With what delay should she expect to arrive at N?

Solution:

The problem can be modeled as a network having four nodes (S,D,C,S), where SD, SC, CD and CN are linked (See Fig. 1). We can define four independent random variable indicating the delay on each of the link:

- X_{SC} is 0 wp $1-p$ and is 1 wp p.
- X_{SD} is 0 wp $1-q$ and is 2 wp q.
- X_{CN} is 0 wp $1-p$ and is 1 wp p.
- X_{DN} is 0 wp $1-q$ and is 1 wp q.

Also, let us define D as the event that Alice flies through Denver, and C as the event that Alice flies through Chicago.

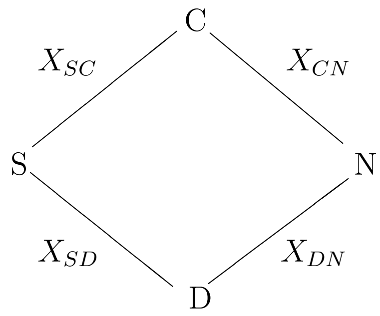

Figure 1: Network modeling Problem 1

(a) There are two possible ways to go to N from S, so using Total Probability law we have

$$\begin{aligned}
\mathbb{E}(\text{delay}) &= \mathbb{E}(\text{delay}|C)P(C) + \mathbb{E}(\text{delay}|D)P(D) \\
&= \mathbb{E}(X_{SC} + X_{CN})P(C) + \mathbb{E}(X_{SD} + X_{DN})P(D) \\
&= \frac{1}{2}(\mathbb{E}(X_{SC}) + \mathbb{E}(X_{CN}) + \mathbb{E}(X_{SD}) + \mathbb{E}(X_{DN})) \\
&= p + 2q
\end{aligned}$$

(b) Using Total Probability law and Bayes' rule, and since $P(D) = P(C) = \frac{1}{2}$ we have

$$\begin{aligned}
P(C|\text{delay} = 2) &= \frac{P(\text{delay} = 2|C)P(C)}{P(\text{delay} = 2|D)P(D) + P(\text{delay} = 2|C)P(C)} \\
&= \frac{P(X_{SC} + X_{CN} = 2)P(D)}{P(X_{SD} + X_{DN} = 2)P(D) + P(X_{SC} + X_{CN} = 2)P(C)} \\
&= \frac{P(X_{SC} + X_{CN} = 2)}{P(X_{SD} + X_{DN} = 2) + P(X_{SC} + X_{CN} = 2)} \\
&= \frac{p^2}{2q(1-q) + p^2}
\end{aligned}$$

(c) Flying via Denver

$$P(\text{delay} < 2|D) = P(X_{SD} + X_{DN} < 2) = (1-q)^2$$

and flying via Chicago

$$\begin{aligned}
P(\text{delay} < 2|C) &= P(X_{SC} + X_{CN} < 2) \\
&= 1 - P(X_{SC} + X_{CN} = 2) = 1 - p^2
\end{aligned}$$

Alice should fly via Chicago when $(1-q)^2 > 1-p^2$. This is the case when $q < 1-\sqrt{1-p^2}$ or, equivalently, $p > \sqrt{2q(1-q)}$.

(d) A delay of two hours happens with probability p^2 on each trip. We are asked for the mean of a geometric random variable with parameter p^2. Thus, the average number of trips is $\frac{1}{p^2}$.

(e) From the independence of the four random variables

$$\begin{aligned}
\mathbb{E}(\text{delay}|X_{SD}=2) &= \mathbb{E}(\text{delay}|X_{SD}=2,D)P(D|X_{SD}=2) \\
&\quad +\mathbb{E}(\text{delay}|X_{SD}=2,C)P(C|X_{SD}=2) \\
&= \frac{1}{2}(2+\mathbb{E}(X_{DN})+\mathbb{E}(X_{SC}+\mathbb{E}(X_{CN})) \\
&= \frac{1}{2}(2+2q+2p) = 1+q+p
\end{aligned}$$

Problem 2: (13 points)

We transmit a bit of information which is 0 with probability $1-p$ and 1 with p. Because of noise on the channel, each transmitted bit is received correctly with probability $1-\epsilon$.

(a) (2pt) Suppose we observe a "1" at the output. Find the conditional probability p_1 that the transmitted bit is a "1".

(b) (4pt) Suppose that we transmit the same information bit n times over the channel. Calculate the probability that the information bit is a "1" given that you have observed n "1"s at the output. What happens when n grows? Does it make sense intuitively?

(c) (3pt) For this part of the problem, we suppose that we transmit the symbol "1" a total of n times over the channel. At the output of the channel, suppose that we observe the symbol "1" three times in the n received bits, and that we observe a "1" at the n-th transmission. Given these facts, what is the probability that the k-th received bit is a "1"?

(d) (4pt) Now let's go back to the situation in part (a)— that is, some unknown bit is transmitted over the channel, and the received bit is a "1". Suppose in addition that the same information bit is transmitted a second time, and you again receive another "1". We want to find a recursive formula to update p_1 to get p_2, the conditional probability that the transmitted bit is a "1" given that we have observed two "1"s at the output of the channel. Show that the update can be written as

$$p_2 = \frac{(1-\epsilon)p_1}{(1-\epsilon)p_1 + \epsilon(1-p_1)}$$

Solution:

(a) Let A be the event that 1 is transmitted, A^c be the event that 0 is transmitted, B^n be the event the n-th bit we received is 1. Then using Total Probability law and Bayes' rule:

$$p_1 = P(A|B_1) = \frac{P(B_1|A)P(A)}{P(B_1|A)P(A) + P(B_1|A^c)P(A^c)} = \frac{(1-\epsilon)p}{(1-\epsilon)p + \epsilon(1-p)}$$

(b) As in the previous part, and assuming that the uses of the channel are independent, we have

$$\begin{aligned} P(A|B_1,\ldots,B_n) &= \frac{P(B_1,\ldots,B_n|A)P(A)}{P(B_1,\ldots,B_n|A)P(A) + P(B_1,\ldots,B_n|A^c)P(A^c)} \\ &= \frac{(1-\epsilon)^n p}{(1-\epsilon)^n p + \epsilon^n(1-p)} \\ &= \frac{p}{p + (\frac{\epsilon}{1-\epsilon})^n (1-p)} \end{aligned}$$

As $n \to \infty$, we have

$$p_\infty = \begin{cases} 1 & \epsilon < \frac{1}{2} \\ p & \epsilon = \frac{1}{2} \\ 0 & \epsilon > \frac{1}{2} \end{cases}$$

These results meet the intuitions. When $\epsilon < \frac{1}{2}$, it means the observations are positively correlated with the bit transmitted, thus knowing the observations B_n $n = 1, 2, \ldots$ will increase the conditional probability of A given the observations, until it hits 1. When $\epsilon = \frac{1}{2}$, it means the observations are independent to the bit transmitted, thus given a bunch of observations B_n $n = 1, 2, \ldots$ won't change the conditional probability of A given the observations, which is the same as the prior probability p. When $\epsilon > \frac{1}{2}$, it means the observations are negatively correlated with the bit transmitted, thus knowing the observations B_n $n = 1, 2, \ldots$ will decrease the conditional probability of A given the observations, until it hits 0.

(c) For $j < n$

$$\begin{aligned} P(B_j = 1 | \sum_{i=1}^{n} B_i = 3, B_n = 1, A) &= \frac{P(\sum_{i=1}^{n} B_i = 3 | B_j = 1, B_n = 1, A) P(B_1 = 1 | B_n = 1, A)}{P(\sum_{i=1}^{n} B_i | B_n = 1, A)} \\ &= \frac{\binom{n-1}{1}(1-\epsilon)\epsilon^{n-3}(1-\epsilon)}{\binom{n-2}{2}(1-\epsilon)^2\epsilon^{n-3}} \\ &= \frac{2}{n-1} \end{aligned}$$

while if $j = n$ $P(B_n = 1 | \sum_{i=1}^{n} B_i = 3, B_n = 1, A) = 1$.

(d) Assuming the uses of the channel are independent, we have

$$\begin{aligned} p_2 = P(A|B_2, B_1) &= \frac{P(B_2|A, B_1)P(A|B_1)}{P(B_2|B_1)} \\ &= \frac{\frac{P(B_2, B_1|A)P(A)}{P(B_1|A)P(A)} P(A|B_1)}{P(B_2|B_1)} \\ &= \frac{\frac{(1-\epsilon)^2}{(1-\epsilon)} p_1}{P(B_2|B_1)} \end{aligned}$$

And since

$$\begin{aligned} P(B_2|B_1) &= \frac{P(B_1, B_2)}{P(B_1)} \\ &= \frac{P(B_1, B_2|A)P(A) + P(B_1, B_2|A^C)P(A^C)}{P(B_1|A)P(A) + P(B_1|A^c)P(A^c)} \\ &= \frac{(1-\epsilon)^2 p + \epsilon^2(1-p)}{(1-\epsilon)p + \epsilon(1-p)} \\ &= (1-\epsilon)\frac{(1-\epsilon)p}{(1-\epsilon)p + \epsilon(1-p)} + \epsilon\frac{\epsilon(1-p)}{(1-\epsilon)p + \epsilon(1-p)} \\ &= (1-\epsilon)p_1 + (1-p_1)\epsilon \end{aligned}$$

We have

$$p_2 = \frac{(1-\epsilon)p_1}{(1-\epsilon)p_1 + (1-p_1)\epsilon}$$

One might have noticed that, if we let $p_0 = p$, the functions that we use to upgrade p_0 to p_1 (part (a)), and p_1 to p_2 are the same. Intuitively, for the first observation, p_0 is the prior probability of A before the observation, and p_1 is the posterior probability of A after the observation. Similarly, for the second observation, p_1 is the "prior" probability, and p_2 is the "posterior" probability. Thus the ways to upgrade two prior probabilities to the two posterior probabilities should be the same.

Problem 3: (13 points)

You play the lottery by choosing a set of 6 numbers from $\{1, 2, \ldots, 49\}$ without replacement. Let X be a random variable representing the number of matches between your set and the winning set. (The order of numbers in your set and the winning set does not matter.) You win the grand prize if all 6 numbers match (i.e., if $X = 6$).

(a) (3pt) Compute the PMF p_X of X. What is the probability of winning the grand prize?

(b) (3pt) Suppose that before playing the lottery, you (illegally) wiretap the phone of the lottery, and learn that 2 of the winning numbers are between 1 and 20; another 2 are between 21 and 40, and the remaining 2 are between 41 and 49. If you use this information wisely in choosing your six numbers, how does your probability of winning the grand prize improve?

(c) (3pt) Now suppose *instead* that you determine by illegal wiretapping that the maximum number in the winning sequence is $r \geq 6$. If you use this information wisely in choosing your 6 numbers, how does your probability of winning the grand prize improve?

(d) (4pt) Use a counting argument to establish the identity

$$\binom{n}{k} = \sum_{r=k}^{n} \binom{r-1}{k-1}.$$

(*Hint*: Part (c) of this problem may be useful.)

Solution:

(a) Among the 49 numbers used at the lottery, only 6 correspond to the winning sequence. If the number of matches between our set and the winning set is k, then we must have selected (without replacement and without ordering) exactly k elements from the winning set of size 6 and $6 - k$ elements from the remaining set of $49 - 6$ available numbers. Then, we have that

$$P(X = k) = \frac{\binom{6}{k}\binom{49-6}{6-k}}{\binom{49}{6}}$$

so the probability of winning the grand prize is

$$P(X = 6) = \frac{\binom{6}{6}}{\binom{49}{6}} = \frac{1}{\binom{49}{6}}$$

(b) In order to use wisely the information given to us, we select (without replacement and without ordering) two numbers from the set $\{1, \ldots, 20\}$, two numbers from the set $\{21, \ldots, 40\}$, and two numbers from the set $\{41, \ldots, 49\}$. Among the $\binom{20}{2}\binom{20}{2}\binom{9}{2}$ ways of selecting the six numbers, there is only one that corresponds to the right sequence. So, the probability of winning is

$$\frac{1}{\binom{20}{2}\binom{20}{2}\binom{9}{2}}$$

Which corresponds to an improvement of (roughly) a factor 10 with respect to case when no information is available.

(c) Since we know that the maximum number in the winning sequence is some number r, we select r in our sequence of numbers. Next, we use the information that r is the *maximum* number, and we select the remaining 5 numbers in the set $\{1, \ldots, r-1\}$. The probability of winning the grand prize becomes

$$\frac{1}{\binom{r-1}{5}}$$

(d) We can imagine to number in increasing order the n elements from which we are sampling k elements. Let us denote with r the maximum number that we select. For a fixed r we can use the result in part c) for determining in how many ways we can select the remaining $r-1$ elements: $\binom{r-1}{k-1}$. Summing over all the possible values of r, we have the result.