Chris Piech
CS109

Handout #9
April 8, 2016

Problem Set #2

Due: 2:30pm on Monday, April 18th

For each problem, briefly explain/justify how you obtained your answer in order to obtain full credit. Your explanations will help us determine your understanding of the problem, whether or not you got the correct answer. Moreover, in the event of an incorrect answer, we can still try to give you partial credit based on the explanation you provide. It is fine for your answers to include summations, products, factorials, exponentials, or combinations, unless you are specifically asked for a single numeric answer.

1. Let E and F be events defined on the same sample space S. Prove that:

   $P(EF) \geq P(E) + P(F) - 1$

   (As mentioned in class, this formula is known as Bonferroni's Inequality.)

2. Say in Silicon Valley, 36% of engineers program in Java and 24% of the engineers who program in Java also program in C++. Furthermore, 33% of engineers program in C++.
   a. What is the probability that a randomly selected engineer programs in Java and C++?
   b. What is the conditional probability that a randomly selected engineer programs in Java given that he/she programs in C++?

3. After a long night of programming, you have built a powerful, but slightly buggy, email spam filter. When you don't encounter the bug, the filter works very well, always marking a spam email as SPAM and always marking a non-spam email as GOOD. Unfortunately, your code contains a bug that is encountered 20% of the time when the filter is run on an email. When the bug is encountered, the filter always marks the email as GOOD. As a result, emails that are actually spam will be erroneously marked as GOOD when the bug is encountered. Let p denote the probability that an email is actually non-spam, and let q denote the conditional probability that an email is non-spam given that it is marked as GOOD by the filter.
   a. Determine q in terms of p.
   b. Using your answer from part (a), explain mathematically whether q or p is greater. Also, provide an intuitive justification for your answer.

4. Say all computers either run operating system W or X. A computer running operating system W is twice as likely to get infected with a virus as a computer running operating system X. If 73% of all computers are running operating system W, what percentage of computers infected with a virus are running operating system W?

5. Two cards are randomly chosen without replacement from an ordinary deck of 52 cards. Let E be the event that both cards are Aces. Let F be the event that the Ace of Spades is one of the chosen cards, and let G be the event that at least one Ace is chosen. Compute:
   a. $P(E \mid F)$
   b. $P(E \mid G)$

6. Five servers are located in a computer cluster. After one year, each server independently is still working with probability p, and otherwise fails (with probability $1 - p$).
   a. What is the probability that *at least* 1 server is still working after one year?
   b. What is the probability that *exactly* 3 servers are still working after one year?
   c. What is the probability that *at least* 3 servers are still working after one year?

7. Consider a hash table with 5 buckets, where the probability of a string getting hashed to bucket i is given by p_i (where $\sum_{i=1}^{5} p_i = 1$). Now, 6 strings are hashed into the hash table.
   a. Determine the probability that *each of* the first 4 buckets has at least 1 string hashed to each of them. Explicitly expand your answer in terms of p_i's, so that it does not include any summations.
   b. Assuming $p_1 = 0.1$, $p_2 = 0.25$, $p_3 = 0.3$, $p_4 = 0.25$, $p_5 = 0.1$, explicitly compute your answer to part (a) as a numeric value.

8. Two emails are received at a mail server. Suppose that each email is spam with probability 0.8 and that whether each email message is spam in an independent event from the other.
   a. Suppose that you are told that at least one of the two emails is spam. Compute the conditional probability that both emails are spam.
   b. Suppose now that one of the emails is randomly (accidentally) forwarded from the server to your account, and you see that this email is spam. What is the probability that both emails originally received by the server are spam in this case? Explain your answer.

9. It is college admission season, and a high school student is anxiously waiting to receive mail telling her whether she has been accepted to Stanford. She estimates that the conditional probabilities of receiving notification on each day of the coming week, given that she is accepted or rejected, are as follows:

Day	P(mail | accepted)	P(mail | rejected)
Monday	0.10	0.05
Tuesday	0.25	0.10
Wednesday	0.25	0.15
Thursday	0.20	0.15
Friday	0.10	0.20

   She estimates that her probability of being accepted is 0.6.

   a. What is the probability she receives mail (with either decision) on Monday?
   b. What is the conditional probability the she receives mail on Tuesday, given that she does not receive mail on Monday?
   c. What is the conditional probability the she will be eventually accepted if she does not receive mail on Monday, Tuesday, and Wednesday?
   d. What is the conditional probability that she will be accepted if she receives no mail that entire week?

10. Suppose we want to write an algorithm `fairRandom` for randomly generating a 0 or a 1 with equal probability (= 0.5). Unfortunately, all we have available to us is a function:

```
int unknownRandom();
```

that randomly generates bits, where on each call a 1 is returned with some unknown probability p that need not be equal to 0.5 (and a 0 is returned with probability $1 - p$).

Consider the following algorithm for `fairRandom`:

```
int fairRandom() {
   int r1, r2;
   while (true) {
      r1 = unknownRandom();
      r2 = unknownRandom();
      if (r1 != r2) break;
   }
   return r2;
}
```

a. Show mathematically that `fairRandom` does indeed return a 0 or a 1 with equal probability.

b. Say we want to simplify the function, so we write the `simpleRandom` function below. Would the `simpleRandom` function also generate 0's and 1's with equal probability? Explain why or why not. Determine P(`simpleRandom` returns 1) in terms of p.

```
int simpleRandom() {
   int r1, r2;
   r1 = unknownRandom();
   while (true) {
      r2 = unknownRandom();
      if (r1 != r2) break;
      r1 = r2;
   }
   return r2;
}
```

Continued on back

11. The color of a person's eyes is determined by a pair of eye-color genes, as follows:
    - if *both* of the eye-color genes are blue-eyed genes, then the person will have blue eyes
    - if *one or more* of the genes is a brown-eyed gene, then the person will have brown eyes

    A newborn child independently receives one eye-color gene from each of its parents, and the gene it receives from a parent is equally likely to be either of the two eye-color genes of that parent. Suppose William and both of his parents have brown eyes, but William's sister (Claire) has blue eyes. (We assume that blue and brown are the only eye-color genes.)
    a. What is the probability that William possesses a blue-eyed gene?
    b. Suppose that William's wife has blue eyes. What is the probability that their first child will have blue eyes?
    c. Still assuming that William's wife has blue eyes, if their first child had brown eyes, what is the probability that their next child will also have brown eyes?

12. A bit string of length n is generated randomly such that each bit is generated independently with probability p that the bit is a 1 (and 0 otherwise). How large does n need to be (in terms of p) so that the probability that there is at least one 1 in the string is at least 0.7?

13. Your colleagues in a comp-bio lab have sequenced DNA from a large population in order to understand how a gene (G) influences two particular traits (T_1 and T_2). They find that $P(G) = 0.6$, $P(T_1|G) = 0.8$ and $P(T_2|G) = 0.9$. They also observe that if a subject does not have the gene, they express neither T_1 nor T_2. The probability of a patient having both T_1 and T_2 given that they have the gene is 0.72
    a. Are T_1 and T_2 conditionally independent given G?
    b. Are T_1 and T_2 conditionally independent given G^C?
    c. What is $P(T_1)$?
    d. What is $P(T_2)$?
    e. Are T_1 and T_2 independent?

14. Consider the following algorithm for betting in roulette:
    i. Bet \$10 on "red"
    ii. If "red" comes up on the wheel (with probability 18/38), then you win \$10 (and keep your \$10 bet) and you <u>immediately</u> quit (i.e., you do not do step (iii) below).
    iii. If "red" did not come up on the wheel (with probability 20/38), then you lose your initial \$10 bet. But, you then bet \$10 on "red" on each of the next <u>two</u> spins of the wheel. After those two spins, you quit (no matter what the outcome of the spins).

    Let X denote your winnings when you quit. (Note that you are only running this algorithm once).

    a. Determine $P(X > 0)$.
    b. Determine $E[X]$.