

Due Tuesday, July 24, 4:59pm

You may work in groups of up to 4 people (no larger!). Please read the group collaboration policies on the course website at <http://inst.cs.berkeley.edu/~cs70/su12/> before beginning group work. You *must* write up the solution set entirely on your own. You must never look at any other students' solutions (not even a draft), nor share your own solutions (not even a draft).

Please put your answer to each problem on its own sheet of paper. Label each sheet of paper with your name, student ID, section number (101, 102, or 103), the assignment number, the problem number, and "CS70–Summer 2012." Turn in your homework in the boxes labeled "CS70" on the 2nd floor of Soda Hall. Submit each problem separately in its appropriate box (i.e., your answer to question i goes into CS 70 drop box i). Failure to follow these instructions may cost you points, or cause you to receive no credit at all.

Note that this homework is due at 4:59pm on Tuesday.

1. (8 pts.) Marbles

Box A contains 1 black and 3 white marbles, and box B contains 2 black and 4 white marbles. A box is selected at random, and a marble is drawn at random from the selected box.

- (a) What is the probability that the marble is black?
- (b) Given that the marble is white, what is the probability that it came from the box A?

2. (14 pts.) Find the Probabilities

Let A and B be two events. Suppose $\Pr[A] = 1/5$, $\Pr[A|B] = 2/5$ and $\Pr[B|A] = 1/2$.

- (a) What is $\Pr[B]$? Show your work.
- (b) What is $\Pr[A \cup B]$? Show your work.
- (c) Can A and B be independent events? Give a brief explanation.
- (d) Can A and B be disjoint events? Give a brief explanation.

3. (12 pts.) Serve on the jury

In the OJ Simpson murder trial, OJ Simpson was accused of murdering his ex-wife, Nicole Simpson. The prosecution introduced evidence showing that OJ had previously abused Nicole. One of Simpson's defense lawyers, Alan Dershowitz, made the following argument in OJ Simpson's defense. Dershowitz stated that 1 in 1,000 women abused by their husbands are later killed by their abuser, so the fact that OJ Simpson had previously abused his wife is not relevant and should be disregarded. Assume for this problem that Dershowitz's 1 in 1,000 statistic is accurate.

- (a) Are we entitled to conclude that there is only a 1/1000 probability that OJ Simpson murdered Nicole? Why or why not?

- (b) Suppose we select at random a woman who has been abused by her husband. Define the following events: M is the event that the woman is murdered at some point in her life; G is the event that the woman is murdered by her abuser at some point in her life. A plausible estimate is that 0.2% of abused women will be murdered by someone other than their abuser at some point in their life. Calculate the probability that the selected woman is murdered by her abuser, given that she is murdered.
- (c) Based upon your answer to part (b), do you agree or disagree with Dershowitz's argument? Based upon your calculation, would you consider it relevant that OJ Simpson previously abused Nicole? Would you judge it more accurate to use the 1 in 1,000 number or the number you calculated in part (b)? Why?

4. (10 pts.) Probability models

Suppose a document has n letters and you are told the probability that each letter is typed wrongly is p .

- (a) Explain why there is not enough information to calculate the probability that the entire document is typed correctly.
- (b) Find the smallest possible value for this probability in terms of p . For what values of p is this zero? For the remaining values of p (i.e. when the minimum probability that the document is typed correctly is nonzero), specify the probability assignment to the underlying sample space for which the smallest value is achieved.
- (c) Let $n = 2$. Suppose you are told additionally that q_1 is the conditional probability that the second letter is typed wrongly given that the first letter is typed wrongly, and q_2 is the conditional probability that the second letter is typed wrongly given that the first letter is typed correctly. Do we now have a complete probability assignment on all the sample points of the experiment? If so, give it.
- (d) What constraint(s) do q_1 and q_2 have to satisfy?
- (e) Under what values of q_1 and q_2 is the probability that the document is typed correctly (i) equal to, (ii) larger than, and (iii) smaller than the probability when the events that the two letters are typed incorrectly are independent?

5. (14 pts.) A Very Small Example of Hashing

Suppose we hash three objects randomly into a table with three (labelled) entries. We are interested in the lengths of the linked lists at the three table entries.

- (a) List all the outcomes in the sample space of the experiment. How many of them are there?
- (b) Let X be the length of the linked list at entry 1 of the table. Write down X explicitly as a function on the sample space mapping to the real line (either in a figure as in class or as a list). Compute and plot the distribution and expectation of X .
- (c) Let Y be the length of the *longest* linked list among all three. Write down Y explicitly as a function on the sample space mapping to the real line. Compute and plot the distribution and expectation of Y .
- (d) Is the expectation of X larger than, equal to or smaller than that of Y ?
- (e) Compute the distribution of X for the general case when m objects are hashed randomly into a table of size n , i.e. give an expression for the probability that X takes on each value in its range. (Computing the distribution of Y for the general case is not so easy, so we won't ask you to do it!)

6. (14 pts.) Random Variables and Their Distributions

A biased coin with probability p landing with a head is flipped 4 times.

- (a) Give the sample space and assign probabilities to the sample points.

- (b) Let X be the total number of heads in the four flips. Draw a Venn diagram showing the five events $X = i, i = 0, 1, 2, 3, 4$ as well as the sample space and the sample points. Is X a random variable?
- (c) Are the events $X = 1$ and $X = 2$ disjoint? Are they independent?
- (d) Let Y be the first flip when a head appears. Draw a Venn diagram showing the four events $Y = i, i = 1, 2, 3, 4$ as well as the sample space and the sample points. Is Y a random variable?
- (e) Are the events $X = 3$ and $Y = 4$ disjoint? Are they independent? How about the events $X = 2$ and $Y = 2$? Disjoint? Independent?
- (f) If X is a random variable, compute its distribution. If not, modify its definition appropriately to make it into a random variable and then compute its distribution.
- (g) If Y is a random variable, compute its distribution. If not, modify its definition appropriately to make it into a random variable and then compute its distribution.

7. (10 pts.) Chopping up DNA

- (a) In a certain biological experiment, a piece of DNA consisting of a linear sequence (or string) of 4001 nucleotides is subjected to bombardment by various enzymes. The effect of the bombardment is to randomly cut the string between pairs of adjacent nucleotides: each of the 4000 possible cuts occurs independently and with probability $\frac{1}{500}$. What is the expected number of pieces into which the string is cut? Justify your calculation.
[HINT: Use linearity of expectation! If you do it this way, you can avoid a huge amount of messy calculation. Remember to justify the steps in your argument; i.e., do not appeal to “common sense.”]
- (b) Suppose now that the cuts are no longer independent, but highly correlated, so that when a cut occurs in a particular place other cuts close by are much more likely. The probability of each individual cut remains $\frac{1}{500}$. Does the expected number of pieces increase, decrease, or stay the same? Justify your answer with a precise explanation.

8. (12 pts.) Conditional Independence

On a particular day, let A be the event that Alex has no homework, and let B be the event that the weather is nice. Assume $\Pr[A] = 1/3$ and $\Pr[B] = 2/3$, and that A and B are independent: $\Pr[A \cap B] = \Pr[A]\Pr[B]$.

- (a) If Alex has no homework and the weather is nice, Alex always plays volleyball, but otherwise (if Alex has homework or the weather is not nice) Alex never plays Volleyball. Let C be the event that Alex plays volleyball. Express C in terms of A and B .
- (b) Conditioned on the event that Alex plays volleyball, what is the probability that the weather is nice? Conditioned on the event that Alex does *not* play volleyball, what is the probability that the weather is nice?
- (c) Here’s a new definition: events D and E are *conditionally independent given F* if $\Pr[D \cap E|F] = \Pr[D|F]\Pr[E|F]$. Are A and B conditionally independent given C ? Are they conditionally independent given \bar{C} ?

9. (11 pts.) Machine Failures

Two faulty machines, M_1 and M_2 , are repeatedly run synchronously in parallel (i.e., both machines execute one run, then both execute a second run, and so on). On each run, M_1 fails with probability p_1 and M_2 with probability p_2 , all failure events being independent. Let the random variable X_1 denote the number of runs until the first failure of M_1 , and X_2 denote the number of runs until the first failure of M_2 .

Let $X = \min(X_1, X_2)$ denote the number of runs until the first failure of *either* machine. Compute the distribution of X . What is its expectation?