

Probability and Computing: Homework 2

Out: 09/05 Due: Friday 09/21; put your solutions into the envelope taped to my door (HEC339) before 4:00pm or email them to me.

Take time to write clear and concise answers; confused and long-winded solutions will be penalized.

Problem 1 Let X and Y two arbitrary discrete random variables with finite expectations. Simplify

$$\mathbf{E}[\max\{X, Y\}] + \mathbf{E}[\min\{X, Y\}].$$

Problem 2 Consider the following scheme for generating a random $r \in \{1, \dots, n\}$ using coins C_1, C_2, \dots, C_n , where $\Pr[C_i \text{ comes up Heads}] = 1/i$. Flip coins $C_n, C_{n-1}, C_{n-2}, \dots$ in sequence until the first Heads appears; if this happens on coin C_i then output $r = i$. Prove that this schemes generates $r \in \{1, \dots, n\}$ uniformly at random. What is the expected number of coin tosses?

Problem 3 A monkey types on a 26-letter keyboard. Assume that each of the 26 letters is equally likely to be hit at each keystroke. What is the expected number of times the sequence “UCF” occurs in a random text of length 10^6 . (Hint: Let X be the number of occurrences. Write X as the sum of indicator variables and use linearity of expectations. This should be pretty simple!)

Problem 4 Suppose we toss a coin C with $\Pr[C \text{ comes up Heads}] = p$ until we observe k Heads. Let the random variable X denote the number of tosses. (Observe that we recover the geometric distribution (MU Section 2.4) as a special case for $k = 1$.)

(a) Prove that the distribution of X is

$$\Pr[X = t] = \binom{t-1}{k-1} p^k (1-p)^{t-k}.$$

This distribution is known as the *negative binomial* distribution.

(b) What is the expectation of X ? Hint: Use linearity of expectations and the formula for the expectation of a geometric random variable. This should also be pretty simple!

Problem 5 Let $b \in \{0, 1\}^n$ be an arbitrary binary string of length n . Its Hamming weight, denoted by $w(b)$, is the number of times 1 occurs in b . Set

$$B := \{b \in \{0, 1\}^{1000} : w(b) = 500\},$$

i.e., the set of binary strings of length 1000 and Hamming weight 500. Alice and Bob have a fair coin. They want to use it to generate $b \in B$ uniformly at random.

- (a) Bob suggests the following scheme: flip the coin 1000 times; if you get exactly 500 Heads, output the sequence; otherwise, try again. How many tosses do you expect to have to make using Bob's scheme? You may assume that $n = 1000$ is large enough that asymptotic results hold; so, instead of dealing with large factorials you should use Stirling's approximation: $n! \approx \left(\frac{n}{e}\right)^n \sqrt{2\pi n}$.
- (b) Alice claims that the following scheme is much more efficient: flip the coin until you have either 500 Heads or 500 Tails (one of these events must happen before 1000 tosses); output this sequence, padded at the end with Heads or Tails respectively to make the total length 1000. Clearly, this scheme requires at most 1000 tosses. But does this scheme really generate binary strings $b \in B$ uniformly at random? Justify your answer with a precise calculation.
- (c) Suggest your own scheme for sampling uniformly at random from B . What is the expected number of tosses required by your scheme? There is a scheme with expected number of tosses as low as 2000. However, you will already get full credit if your scheme is substantially better than Bob's scheme (and, of course, it has to be correct;).

Problem 6 Describe a method for using a fair coin to generate a permutation π of $\{1, \dots, n\}$ uniformly at random. What is the expected number of coin tosses of your scheme. It should be $O(n \log n)$.

Problem 7 (Bonus) A permutation π of $\{1, \dots, n\}$ can be represented as a set of cycles as follows. Let there be one vertex for each $i, i = 1, \dots, n$. If the permutation maps the number i to the number $\pi(i)$, then a directed arc is drawn from vertex i to vertex $\pi(i)$. This leads to a graph that is a set of disjoint cycles. Notice that some of the loops could be self-loops. What is the expected number of cycles in a permutation chosen uniformly at random?

Problem 8 (Bonus) Generalizing on the notion of a cut-set, we define an r -way cut in a graph as a set of edges whose removal breaks the graph into r or more connected components. Explain how the randomized min-cut algorithm (MU Section 1.4) can be used to find minimum r -way cuts, and bound its success probability.