CMSC 651, Spring 2018, University of Maryland
HW1, due as PDF to the address *cmsc651.umd@gmail.com* by 11:59PM on February 12, 2018

Notes: (i) Please work on this with your group (one writeup per group). Consulting other sources (including the Web) is not allowed. (ii) Write your solutions neatly and *include your names*; if you are able to make partial progress by making some additional assumptions, then **state these assumptions clearly and submit your partial solution**. (iii) **Please make the subject line of your email** "*651 HW1*" followed by your full name, and email a PDF to *cmsc651.umd@gmail.com*: PDF generated from Word or LaTeX strongly encouraged.

This HW. The "first-moment method" refers to only using the expectation of a random variable; Markov's inequality is a good example. Similarly, the second-moment method uses the mean and the variance, as typified by Chebyshev's inequality.

0. Read and understand Lectures 1-7 from the randomization lecture notes. Specifically, you need to be familiar with Markov's and Chebyshev's inequalities mentioned above.

1. We want to show here that independence does not always imply conditional independence. More specifically, for any given pair of *independent* events A, B with $0 < \Pr[A], \Pr[B] < 1$:

(a) construct an event C such that $\Pr[(A \wedge B) \mid C] < \Pr[A \mid C] \cdot \Pr[B \mid C]$; prove that this inequality holds. **(5 points)**

(b) construct an event D such that $\Pr[(A \wedge B) \mid D] > \Pr[A \mid D] \cdot \Pr[B \mid D]$; prove that this inequality holds. **(5 points)**

Hint: Some simple Boolean functions C, D of A and B will do.

2. We study lower tails here, using the first-moment method.

(i) Use the first-moment method to prove the following: there is a constant $c > 0$ such that for any undirected graph G – with number of edges denoted m as usual – the probability that the simple "random cut" algorithm (Lecture 4, Section 2.1) produces a cut of size at least $0.499m$, is at least c. **(5 points)**

(ii) How will you use the result of problem (i) if you want an algorithm to construct a cut of size at least $0.499m$ with probability at least 0.9? **(5 points)**

3. Use the second-moment method to show that the probability that the simple "random cut" algorithm produces a cut of size at least $0.499m$, is actually as high as $1 - O(1/m)$. The point of this problem is that we can get (significantly) better bounds as we go to higher moments, i.e., as we use more information about the underlying random variable. (**Hint:** Look at the co-variances carefully.) **(8 points)**

4. We study the second moment further here.

(i) For any random variable X with some mean μ and variance v, and for any $a > 0$, prove that each of $\Pr[X \le \mu - a]$ and $\Pr[X \ge \mu + a]$ is upper-bounded by $v/(v + a^2)$. (Use a suitable quadratic function, just like in the proof of Chebyshev's inequality.) Second, in the context of Chebyshev's inequality (where we want to upper-bound $\Pr[|X - \mu| \ge a]$), is it a good idea to combine these two to get an alternative bound? Why, or why not? **(5 + 2 points)**

(ii) A median for a random variable X is any value u such that $\Pr[X \le u] \ge 1/2$ and $\Pr[X \ge u] \ge 1/2$. Show that for any random variable X with some mean μ and standard deviation σ, its median u is such that $\mu - \sigma \le u \le \mu + \sigma$. **(5 points)**

5. Consider a monkey typing a random n-symbol string S: it types a symbol from the English alphabet $\{a, b, c, \ldots, z\}$ n times, uniformly at random and independently.

(i) Let X be the number of times the string "proof" appears as a substring (i.e., contiguous subsequence) of our random string S. It is easy to see that $\mathbf{E}[X] = (n-4) \cdot (1/26)^5$. Prove that $\mathbf{var}[X] \leq O(n)$. (Thus, Chebyshev can then be used to prove good tail bounds for X.) **(8 points)**

(ii) Let s be an *arbitrary* string of some length d. Let Y be the number of times the string s appears as a substring of our random string S. Prove that $\mathbf{var}[Y] \leq O(n \cdot (1/26)^d)$. **(7 points)**

6. Consider Section 2.1 in Lecture 14 of the randomization lecture notes. There, in the discussion of the fourth moment, we are essentially claiming that $\mathbf{E}[(X - n/2)^4] \leq O(n^2)$. Prove this. **(5 points)**

Why are we interested in such higher moments? An easy generalization of Chebyshev's inequality is the following k^{th} *central-moment inequality* for any positive integer k: if X is a random variable with mean μ and if $a > 0$, then the tail probability $\Pr[|X - \mu| \geq a]$, is at most $\mathbf{E}[(|X - \mu|)^k]/a^k$. (This becomes simpler if k is an *even* positive integer, in which case $(|X - \mu|)^k = (X - \mu)^k$.) Thus, understanding such higher moments is often very helpful. LaTeX