COMS 4995 (Randomized Algorithms): Exercise Set #4

For the week of September 23–27, 2019

Instructions:

(1) *Do not turn anything in.*

(2) The course staff is happy to discuss the solutions of these exercises with you in office hours or in the course discussion forum.

(3) While these exercises are certainly not trivial, you should be able to complete them on your own (perhaps after consulting with the course staff or a friend for hints).

Exercise 15

Suppose $X_1, \ldots, X_n$ are pairwise independent random variables. (I.e., for every $i, j \in \{1, 2, \ldots, n\}$ with $i < j$, and for all values x_i and x_j in the range of X_i and X_j, $\mathbf{Pr}[X_i = x_i \text{ and } X_j = x_j] = \mathbf{Pr}[X_i = x_i] \cdot \mathbf{Pr}[X_j = x_j]$.) Prove that

$$\mathbf{E}[X_i \cdot X_j] = \mathbf{E}[X_i] \cdot \mathbf{E}[X_j]$$

for every $1 \le i < j \le n$. (In other words, expectations of products factor as long as only two variables are involved.)

Exercise 16

Suppose $X_1, \ldots, X_n$ are pairwise independent random variables. Prove that variances add in this case:

$$\mathrm{Var}\left[\sum_{i=1}^{n} X_i\right] = \sum_{i=1}^{n} \mathrm{Var}[X_i]\,.$$

[Hint: use the definition of variance to expand the left-hand side and use the previous exercise to deal with the cross terms.]

Exercise 17

Prove the following generalization of Chebyshev's inequality: for every positive even integer k, every random variable X with $\mathbf{E}\big[X^\ell\big] < +\infty$ for $\ell = 1, 2, \ldots, k$, and every $\beta > 0$,

$$\mathbf{Pr}[|X - \mathbf{E}[X]\,| \ge \beta] \le \frac{\mathbf{E}\big[(X - \mathbf{E}[X])^k\big]}{\beta^k}.$$

(In other words, the higher the moments of a random variable you have control over, the better the control you have over the probability of a large deviation.)

Exercise 18

Let D_1, D_2 denote two probability distributions on a finite set Ω. Define the *statistical distance* between D_1 and D_2 as

$$\max_{S \subseteq \Omega} |\mathbf{Pr}_{D_1}[S] - \mathbf{Pr}_{D_2}[S]|\,.$$

Define the ℓ_1 *distance* as

$$\sum_{\omega \in \Omega} \left| \mathbf{Pr}_{D_1}[\omega] - \mathbf{Pr}_{D_2}[\omega] \right|.$$

Prove that the statistical distance is precisely half the ℓ_1 distance. (We used this fact in the proof of the Leftover Hash Lemma in Lecture #7.)

Exercise 19

Recall from Lecture #7 that the *collision probability* of a random variable X with range R is the probability that two independent copies of X take on the same value:

$$cp(X) = \sum_{i \in R} \mathbf{Pr}[X = i]^2.$$

Prove that a random variable X that satisfies

$$\max_{i \in R} \mathbf{Pr}[X = i] \leq \delta$$

has collision probability at most δ. Prove that the two quantities coincide when X is distributed uniformly over the range of X.

Exercise 20

Prove that for a vector $\vec{v} \in \mathbb{R}$,

$$\|\vec{v}\|_2 \leq \|\vec{v}\|_1 \leq \sqrt{d} \cdot \|\vec{v}\|_2.$$

For each of the two inequalities, give an example of a non-zero vector such that it holds with equality. (We used the second inequality in the proof of the Leftover Hash Lemma in Lecture #7.)

[Hint: For the second inequality, use Cauchy-Schwarz.]