Probability and Statistics in Computing **Monsoon 2019**

HW 1: Preliminaries

Out: August 22, 2018 **Due:** September 5

(1) Please take time to write clear and concise solutions. You are STRONGLY *encouraged to submit LaTeXed solutions by email. (2) Collaboration is OK, but please write your answers yourself, and include in your answers the names of* EVERYONE *you collaborated with and* ALL *references other than class notes you consulted.*

1. (3 points) $\mathbf{E}\left[X^2\right]$ is assumed to be well defined for all random variables X appearing in this problem. Prove the following:

   (a) (0 points) $\mathbf{Var}\left[\sum_{i=1}^n X_i\right] = \sum_{i=1}^n \mathbf{Var}\left[X_i\right] + \sum_{1 \le i \ne j \le n} \mathbf{Cov}\left[X_i, X_j\right].$

   (b) (1 point) $\mathbf{Var}\left[X\right] = \frac{1}{2}\mathbf{E}\left[(X-Y)^2\right]$, where Y is independent of X and has distribution identical to X.

   (c) (1 point) If $f : \mathbb{R} \to \mathbb{R}$ is α-Lipschitz (i.e., $|f(x) - f(y)| \le \alpha\,|x - y|$ for all $x, y \in \mathbb{R}$) then
   $$\mathbf{Var}\left[f(X)\right] \le \alpha^2 \mathbf{Var}\left[X\right].$$

   (d) (1 points) **[Bunyakovsky-Cauchy-Schwarz inequality]** $\mathbf{E}\left[|XY|\right]^2 \le \mathbf{E}\left[X^2\right] \cdot \mathbf{E}\left[Y^2\right]$.

   Hint: It might help to consider the function $f(t) := \mathbf{E}\left[(|X| - t\,|Y|)^2\right]$.

2. (3 points) Let (X, Y) be jointly Gaussian random variables with mean $(0, 0)$ and co-variance matrix $\Sigma = \begin{pmatrix} 1 & 1 \\ 1 & 2 \end{pmatrix}$. Describe the distribution of the vector $(5X + 3Y, 3X + 4Y)$.

3. (4 points) In class, we saw that the Neyman-Pearson "threshold test" may not have the optimal power for a given significance level if that significance level cannot be achieved exactly by a threshold test. Now consider the null probability distribution $p_0 := (p_0(i))_{i=1}^n$ and the alternative probability distribution $p_1 := (p_1(i))_{i=1}^n$, both on the sample space $[n]$. Assume that the $p_o(i)$ and $p_1(i)$, $i \in [n]$ are rational numbers that can each be expressed as a ratio of integers with denominator M.

   Give an algorithm that on input a rational number α (also expressed as a ratio of integers with denominator M) and the probability vectors p_0 and p_1 as above, outputs in time polynomial in n and M a description of the test of optimal power with significance level at most α.

4. (4 points) Let X be a non-negative random variable with finite second moment $\mathbf{E}\left[X^2\right]$. We denote by $\mathbf{I}[Z]$ the indicator random variable for event Z, so that $\mathbf{E}\left[\mathbf{I}[Z]\right] = \mathbf{P}\left[Z\right]$ for any event Z. Let $\alpha \in (0, 1)$ be a fixed real number, and let $\mu = EX$.

   (a) (0 points) Show that $\mathbf{E}\left[X \cdot \mathbf{I}[X \le \alpha\mu]\right] \le \alpha\mu$.

   (b) (1 points) Show that $\mathbf{E}\left[X \cdot \mathbf{I}[X > \alpha\mu]\right]^2 \le \mathbf{E}\left[X^2\right] \mathbf{P}\left[X > \alpha\mu\right]$.

   (c) (1 point) **[Paley-Zygmund inequality]** Show therefore that
   $$\mathbf{P}\left[X > \alpha\mu\right] \ge (1 - \alpha)^2 \frac{\mathbf{E}\left[X\right]^2}{\mathbf{E}\left[X^2\right]}.$$

   (d) (2 points) Let $X_1, X_2, \ldots X_n$ be 4-wise independent uniformly distributed Rademacher variables (i.e., each X_i is uniformly distributed in $\{+1, -1\}$). Let $S = \sum_{i=1}^n X_i$. Show that for $\alpha \in (0, 1)$
   $$\mathbf{P}\left[|S| > \alpha\sqrt{n}\right] \ge (1 - \alpha^2)^2 \cdot \frac{n}{3n - 2} \ge \frac{(1 - \alpha^2)^2}{3}.$$
   Note that this is a kind of *anti-concentration* bound: we are putting a *lower bound* on the probability that S is far from its expectation. Such bounds are very important in many areas of mathematics, including in the study of random matrices.

   Note: $X_1, X_2, \ldots, X_n$ are said to be k-wise independent if for any subset S of $[n]$ of size at most k and any collection $(f_i)_{i \in S}$ of functions for which the expectations are well-defined, one has
   $$\mathbf{E}\left[\prod_{i \in S} f_i(X_i)\right] = \prod_{i \in S} \mathbf{E}\left[f_i(X_i)\right].$$

Note: *The next two problems will not be graded, but you are expected to be able to solve them.*

5. (0 points) Let X be a random variable supported on the interval $[a, b]$ and satisfying $\mathbf{E}[X] = 0$. Assume that $a < 0 < b$. Define $\phi(\lambda) := \log \mathbf{E}[\exp(\lambda X)]$ for $\lambda \geq 0$.

   (a) Show that for $x \in [a, b]$ and any real λ, $\exp(\lambda x) \leq \frac{x-a}{b-a} \cdot \exp(\lambda b) + \frac{b-x}{b-a} \cdot \exp(\lambda a)$. Hence obtain an explicit upper bound $f(\lambda)$ on $\phi(\lambda)$ in terms of a, b and λ.

   (b) **[Hoeffding lemma]** By analyzing the first two derivatives of f, or otherwise, show that

$$\phi(\lambda) \leq \frac{\lambda^2(b-a)^2}{8}.$$

6. (0 points) Let $Y_0, Y_1, \ldots Y_n$ be a sequence of random variables taking value in some set $\mathcal{Y}$. A sequence $X_1, X_2, \ldots, X_n$ of random variables is said to be a *martingale*[1] with respect to the sequence Y if there is a sequence of deterministic functions $f_1, f_2, \ldots, f_n$ such that $X_i = f_i(Y_0, Y_1, \ldots, Y_i)$, and further

$$\mathbf{E}[X_{i+1} \mid Y_0, \ldots, Y_i] = X_i \qquad \forall i \geq 1.$$

   Let us suppose that this martingale has the *bounded difference property*: there exists a deterministic sequence of constants $c_1, c_2, \ldots, c_n$ such that

$$|X_i - X_{i-1}| \leq c_i \quad \forall i \geq 1.$$

   (a) Show that for any $\lambda \geq 0$,

$$\log \mathbf{E}[\exp(\lambda(X_i - X_{i-1})) \mid Y_0, \ldots Y_{i-1}] \leq \frac{\lambda^2 c_i^2}{2}.$$

   (b) **[Hoeffding-Azuma inequality]** Show therefore that

$$\mathbf{P}[|X_n - \mathbf{E}[X_0]| \geq t] \leq 2\exp\left(-\frac{t^2}{2\sum_{i=1}^n c_i^2}\right).$$

[1] Martingales can be defined in more generality, but this form of the definition is usually sufficient for algorithmic applications.