Toolkit for TCS **Feb-June 2021**

HW 1: Preliminaries and miscellaneous topics

Out: February 22, 2021 **Due:** March 8, 2021

(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. (5 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) (1 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) (2 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. (5 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) (3 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 $\alpha \in (0,1)$

   $$\mathbf{P}\left[|S| > \alpha\sqrt{n}\right] \ge (1-\alpha^2)^2\cdot\frac{n}{3n-2} \ge \frac{(1-\alpha^2)^2}{3}.$$

   Inequalities like the above are known as *small ball* bounds (and sometimes also as *anti-concentration* bounds). In contrast to concentration inequalities, they put an *upper bound* on the probability that S is in a "small ball" around its expectation. Such bounds are very important in many areas of mathematics, e.g. in the study of random matrices.

3. (5 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

   $$\mathbf{E}\left[X_{i+1} \mid Y_0, \ldots, Y_i\right] = X_i \qquad \forall i \ge 0.$$

   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

   $$|X_i - X_{i-1}| \le c_i \quad \forall i \ge 1.$$

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

(a) (3 points) Show that for any $\lambda \geq 0$,

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

(b) (2 points) **[Hoeffding-Azuma inequality]** Show therefore that

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

4. (5 points) Let X be a random variable satisfying $\mathbf{E}\left[\exp(\lambda\,|X|)\right] \leq \exp(\lambda^2 \nu/2)$ for a fixed positive ν and all real λ.

   (a) (2 points) Show that there exist absolute constants C_1 and C_2 such that for all positive $\lambda \leq \frac{C_1}{\nu}$, the random variable $Z := X^2$ satisfies

   $$\mathbf{E}\left[\exp(\lambda Z)\right] \leq \exp(C_2 \cdot \nu \cdot \lambda).$$

   (b) (3 point) Suppose that $X_1, X_2, \ldots, X_n$ are i.i.d. copies of X, and let $\mathbf{X}$ denote the n-dimensional vector $(X_1, X_2, \ldots, X_n)$. Use the above to show that for any $t > C_2\nu$,

   $$\mathbf{P}\left[\|\mathbf{X}\|_2 > \sqrt{nt}\right] \leq \exp\left(-\frac{C_1(t - C_2\nu)}{\nu} \cdot n\right).$$

5. (5 points) **[Count-Min sketch, Cormode and Muthukrishnan, 2005]** Consider the problem of estimating the frequency counts of individual elements in a data stream (in class, we looked at the AMS algorithm which estimates the sum of squares of these frequency counts). Let M be the number of different types of elements, as in the case of AMS. Let $\mathcal{H}$ be a family of hash functions mapping $[M]$ to $[k]$ ($k \geq 2$), such that if a function h is chosen uniformly at random from $\mathcal{H}$ then for all $i \neq j \in [M]$ and $a, b \in [k]$,

   $$\mathbf{P}_{h\sim\text{Uniform}(\mathcal{H})}\left[h(i) = a \wedge h(j) = b\right] = \frac{1}{k^2}.$$

   (Such a hash family is called 2-universal).

   Consider now the following algorithm for this problem. At the beginning of the algorithm, we sample independently s functions $h_1, h_2, \ldots, h_s$ from $\mathcal{H}$, and initialize all entries of an $s \times k$ array C to 0.

   Now, whenever a new element e arrives, we increment the entries $C(i, h_i(e))$, $1 \leq i \leq s$, by one. On being queried the frequency of item e at any point, we output $D_e := \frac{k}{k-1} \cdot \min\left\{C(j, h_j(e)) \mid 1 \leq j \leq s\right\}$.

   (a) (3 points) Suppose that at some given time, the number of times the element e has been seen is F_e, and the total number of elements seen so far is F. Show that at such a time,

   $$\mathbf{E}\left[C(j, h_j(e))\right] = \left(1 - \frac{1}{k}\right)F_e + \frac{F}{k} \quad \text{for } 1 \leq j \leq s.$$

   Note that the only randomness here is in the choice of the function h_j which is sampled uniformly at random from $\mathcal{H}$.

   (b) (1 point) Show that $D_e \geq F_e$. Show also that at any given time,

   $$\mathbf{P}\left[D_e \geq (1+\epsilon)F_e + \frac{(1+\epsilon)}{k-1}F\right] \leq (1+\epsilon)^{-s}.$$

   (c) (1 point) Suppose that the total number of items seen over the run of the algorithm is n. Let a positive $\epsilon < 1$ be fixed. Show that if we choose $k \geq 2 + 1/\epsilon$ and $s \geq (2/\epsilon)\log(Mn/\delta)$, then with probability at least $1 - \delta$, we have $0 \leq D_e - F_e \leq 2\epsilon F$ for all $e \in M$ and at all time-steps $t \in [n]$.