CS225: Probability and Computing
Shanghai Jiao Tong University

Homework 2
due Saturday 16 July

Problem 1

Let X be a real-valued random variable with finite $\mathbb{E}[X]$ and finite $\mathbb{E}\left[e^{\lambda X}\right]$ for all $\lambda \geqslant 0$. We define the *log-moment-generating function* as

$$\Psi_X(\lambda) := \ln \mathbb{E}[e^{\lambda X}] \quad \text{for all } \lambda \geqslant 0,$$

and its *dual function*:

$$\Psi_X^*(t) := \sup_{\lambda \geqslant 0}(\lambda t - \Psi_X(\lambda)).$$

Assume that X is NOT almost surely constant. Then due to the convexity of $e^{\lambda X}$ with respect to λ, the function $\Psi_X(\lambda)$ is *strictly* convex over $\lambda \geqslant 0$.

(a) Prove the following *Chernoff bound*:

$$\Pr[X \geqslant t] \leqslant \exp(-\Psi_X^*(t)).$$

In particular if $\Psi_X(\lambda)$ is continuously differentiable, prove that the supreme in $\Psi_X^*(t)$ is achieved at the unique $\lambda \geqslant 0$ satisfying

$$\Psi_X'(\lambda) = t$$

where $\Psi_X'(\lambda)$ denotes the derivative of $\Psi_X(\lambda)$ with respect to λ.

(b) **Normal random variables.** Let $X \sim N(\mu, \sigma)$ be a Gaussian random variable with mean μ and standard deviation σ. What are the $\Psi_X(\lambda)$ and $\Psi_X^*(t)$? And give a tail inequality to upper bound the probability $\Pr[X \geqslant t]$.

(c) **Poisson random variables.** Let $X \sim \mathrm{Pois}(\nu)$ be a Poisson random variable with parameter ν, that is, $\Pr[X = k] = e^{-\nu}\nu^k/k!$ for all $k = 0, 1, 2, \ldots$. What are the $\Psi_X(\lambda)$ and $\Psi_X^*(t)$? And give a tail inequality to upper bound the probability $\Pr[X \geqslant t]$.

(d) **Bernoulli random variables.** Let $X \in \{0, 1\}$ be a single Bernoulli trial with probability of success p, that is, $\Pr[X = 1] = 1 - \Pr[X = 0] = p$. Show that for any $t \in (p, 1)$, we have $\Psi_X^*(t) = D(Y\|X)$ where $Y \in \{0, 1\}$ is a Bernoulli random variable with parameter t and $D(Y\|X) = (1-t)\ln\frac{1-t}{1-p} + t\ln\frac{t}{p}$ is the *Kullback-Leibler divergence* between Y and X.

(e) **Sum of independent random variables.** Let $X = \sum_{i=1}^{n} X_i$ be the sum of n *independently and identically distributed* random variables $X_1, X_2, \ldots, X_n$. Show that $\Psi_X(\lambda) = \sum_{i=1}^{n} \Psi_{X_i}(\lambda)$ and $\Psi_X^*(t) = n\Psi_{X_i}^*(\frac{t}{n})$. Also for binomial random variable $X \sim \mathrm{Bin}(n, p)$, give an upper bound to the tail inequality $\Pr[X \geqslant t]$ in terms of KL-divergence.

Give an upper bound to $\Pr[X \geqslant t]$ when every X_i follows the geometric distribution with a probability p of success.

Problem 2

Let X be a random variable with expectation 0 such that moment generating function $E[e^{(t|X|)}]$ is finite for some $t > 0$. We can use the following two kinds of tail inequalities for X.

Chernoff Bound:

$$\Pr[|X| \geqslant \delta] \leqslant \min_{t \geqslant 0} \frac{E[e^{t|X|}]}{e^{t\delta}}$$

kth-Moment Bound:

$$\Pr[|X| \geqslant \delta] \leqslant \frac{E[|X|^k]}{\delta^k}$$

(a) Show that for each δ, there exists a choice of k such that the kth-moment bound is stronger than the Chernoff bound. (Hint: the probabilistic method might help you)

(b) Why would we still prefer the Chernoff bound to the (seemingly) stronger kth-moment bound?

Problem 3

(Due to D.R. Karger and R. Motwani)

(a) Let S, T be two disjoint subsets of a universe U such that $|S| = |T| = n$. Suppose we select a random set $R \subseteq U$ by independently sampling each element of U with probability p. We say that the random sample R is *good* if the following two conditions hold: $R \cap S = \emptyset$ and $R \cap T \neq \emptyset$. Show that for $p = 1/n$, the probability that R is good is larger than some positive constant.

(b) Suppose now that random set R is chosen by sampling the elements of U with only *pairwise* independent. Show that for a suitable choice of the value of p, the probability that R is good is larger than some positive constant.

Problem 4

The *set-cover* problem is the following: given sets $S_1, \ldots, S_n$ over a universe U, find the smallest set $T \subseteq U$ such that for $1 \leqslant i \leqslant n, T \cap S_i \neq \emptyset$. An alternative formulation of this problem is the following: given a 0-1 matrix M, find a 0-1 column vector c such that the dot product of each row of M with c is positive while minimizing $\|c\|_1$. The matrix M has n rows, and the ith row is the incidence vector of the set S_i.

Given a matrix M, let $C(M)$ denote the size of the smallest set-cover for M. Let n be the number of rows in M. Show that we can adapt the technique of linear programming followed by randomized rounding to find a set-cover of size $O(\log n)$ times $C(M)$.

Problem 5

Given a binary string, define a run as a maximal sequence of contiguous 1s; for example, the following string

$$\underbrace{111}_{3}00\underbrace{11}_{2}00\underbrace{111111}_{6}0\underbrace{1}_{1}0\underbrace{11}_{2}$$

contains 5 runs, of length $3, 2, 6, 1, 2$.

Let S be a binary string of length n, generated uniformly at random. Let X_k be the number of runs in S of length k or more.

- Compute the exact value of $\mathbb{E}[X_k]$ as a function of n and k.
- Give the best concentration bound you can obtain for $|X_k - \mathbb{E}[X_k]|$.

Problem 6

For any two points $x, y \in \{0,1\}^n$, the *Hamming distance* $d_H(x, y)$ is defined as

$$d_H(x, y) := \sum_{i=1}^{n} |x_i - y_i|.$$

And the Hamming distance between a point $x \in \{0,1\}^n$ and a set $A \subseteq \{0,1\}^n$ is defined as $d_H(x, A) := \min_{y \in A} d_H(x, y)$. Given any subset $A \subseteq \{0,1\}^n$, $t > 0$, let

$$A_t := \{x \in \{0,1\}^n \mid \exists y \in A, d_H(x, y) \leqslant t\}.$$

Let $X = (X_1, \ldots, X_n) \in \{0,1\}^n$ be sampled such that $X_1, \ldots, X_n$ are mutually independent. Let $A \subseteq \{0,1\}^n$ be arbitrary. We denote that $\mu(A) := \Pr[X \in A]$.

(a) Show that

$$\mathbb{E}[d_H(X, A)] \leqslant \sqrt{2n \ln \frac{1}{\mu(A)}}.$$

(b) Prove a weaker version of Harper's Isoperimetric Inequality: For any constant $\epsilon > 0, \exists \delta > 0$, such that for $t = \epsilon n$, either $\mu(A) < 2^{-\delta n}$ or $\mu(A_t) > 1 - 2^{-\delta n}$.