高级算法 (Fall 2019)/Problem Set 2

- 作业电子版于2019/11/5 23:59 之前提交到邮箱 njuadvalg@163.com
- 每道题目的解答都要有完整的解题过程。中英文不限。

Contents

- 1 Problem 1
- 2 Problem 2
- 3 Problem 3
- 4 Problem 4

Problem 1

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

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

and its *dual function*:

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

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

- Prove the following Chernoff bound:

  $$\Pr[X \geq t] \leq \exp(-\Psi_X^*(t)).$$

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

  $$\Psi'_X(\lambda) = t$$

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

- **Normal random variables.** Let $X \sim \mathrm{N}(\mu, \sigma)$ 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 $\Pr[X \geq t]$.

- **Poisson random variables.** Let $X \sim \mathrm{Pois}(\nu)$ be a Poisson random variable with parameter ν, that is, $\Pr[X = k] = \mathrm{e}^{-\nu}\nu^k/k!$ 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 $\Pr[X \geq t]$.

- **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** (https://en.wikipedia.org/wiki/Kullback–Leibler_divergence) between Y and X.

- **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 $\Pr[X \ge t]$ in terms of KL-divergence.

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

Problem 2

An n-dimensional hypercube Q_n is a graph with 2^n vertices, where each vertex is represented by an n-bit vector, and there is an edge between two vertices if and only if their bit-vectors differ in exactly one bit.

Given an n-dimensional hypercube with some non-empty subset of vertices S, which is called marked black. Let $f(u)$ denote the shortest distance from vertex u to any black vertex. Formally,

$$f(u) = \min_{v\in S} \mathrm{dist}_{Q_n}(u,v),$$

where $\mathrm{dist}_{Q_n}(u,v)$ denotes the length of the shortest path between u and v in graph Q_n .

Prove that if we choose u from all 2^n vertices uniformly at random, then, with high probability, $f(u)$ can not deviate from its expectation too much:

$$\Pr[|f(u) - \mathbb{E}[f(u)]| \ge t\sqrt{n\log n}] \le n^{-c}.$$

Give the relation between c and t.

Problem 3

Let $Y_1, Y_2, Y_3, \ldots, Y_n$ be a set of n random variables where each $Y_i \in \{0,1\}$. All variables Y_i follow some joint distribution over $\{0,1\}^n$ and they may NOT be mutually independent. Assume the following property holds for $(Y_i)_{1\le i\le n}$. For any $1 \le i \le n$ and it holds that

$$\forall c_1 \in \{0,1\}, c_2 \in \{0,1\}, \ldots, c_{i-1} \in \{0,1\}, \quad \Pr[Y_i = 1 \mid \forall 1 \le j < i, Y_j = c_j] \le p.$$

Let $X_1, X_2, X_3, \ldots, X_n$ be a set of n mutually independent random variables where each $X_i \in \{0,1\}$. Assume

$$\forall 1 \le i \le n: \quad \Pr[X_i = 1] = p.$$

Prove that for any $a \ge 0$, it holds that

$$\Pr\left[\sum_{i=1}^{n} Y_i \geq a\right] \leq \Pr\left[\sum_{i=1}^{n} X_i \geq a\right].$$

Prove that for any $t \geq 0$, it holds that

$$\Pr\left[\sum_{i=1}^{n} Y_i \geq np + t\right] \leq \exp\left(-\frac{2t^2}{n}\right).$$

Hint: Although random variables $Y_1, Y_2, Y_3, \ldots, Y_n$ may not be mutually independent, we can still bound the tail probability for $\sum_{i=1}^{n} Y_i$. This tool is called the stochastic dominance.

To prove the first inequality, you only need to construct a coupling $\mathcal{C}$ between $(X_i)_{1\leq i\leq n}$ and $(Y_i)_{1\leq i\leq n}$ such that

$$\Pr_{\mathcal{C}}[\forall 1 \leq i \leq n, Y_i \leq X_i] = 1.$$

This implies the random sequence $(Y_i)_{1\leq i\leq n}$ is stochastically dominated by the random sequence $(X_i)_{1\leq i\leq n}$.

Problem 4

Let U be a universal set. We use 2^U to denote the collection of all subsets of U. Let $\mathcal{F}$ be a family of hash functions, in which each hash function $h : 2^U \to \{0,1\}^m$ maps subsets of U to 0-1 vectors of length m. A locality sensitive hashing scheme is a distribution on a family $\mathcal{F}$ of hash functions operating on 2^U, such that for two subsets $A, B \in 2^U$,

$$(1) \qquad \Pr_{h\in\mathcal{F}}[h(A) = h(B)] = sim(A, B).$$

Here $sim : 2^U \times 2^U \to [0,1]$ is called the similarity function. Given a hash function family $\mathcal{F}$ that satisfies Equation (1), we will say that $\mathcal{F}$ is a locality sensitive hash function family corresponding to similarity function $sim(\cdot,\cdot)$.

- For any similarity function $sim(A, B)$ that admits a locality sensitive hash function family as defined in Equation (1), prove that the distance function $d(A, B) \triangleq 1 - sim(A, B)$ satisfies triangle inequality, formally,

$$\forall A, B, C \in 2^U: \quad d(A, B) + d(B, C) \geq d(A, C).$$

- Show that there is no locality sensitive hash function family corresponding to Dice's and the Overlap coefficient. Dice's coefficient is defined as:

$$sim_{Dice}(A, B) = \frac{2|A \cap B|}{|A| + |B|}.$$

Overlap coefficient is defined as:

$$sim_{Ovl}(A, B) = \frac{|A \cap B|}{\min(|A|, |B|)}.$$

Hint: use the triangle inequality result.

- Construct a collection of hash function $\mathcal{B}$ where $f:\{0,1\}^m\rightarrow\{0,1\}$ for each $f\in\mathcal{B}$, together with a probability distribution on $\mathcal{B}$ such that

$$\forall x,y\in\{0,1\}^m:\quad \Pr_{f\in\mathcal{B}}[f(x)=f(y)]=\begin{cases}1 & \text{if } x=y;\\ \frac{1}{2} & \text{if } x\neq y.\end{cases}$$

Then use the hash function family $\mathcal{B}$ to prove the following result. Given a locality sensitive hash function family $\mathcal{F}$ ($h:2^U\rightarrow\{0,1\}^m$ for each $h\in\mathcal{F}$) corresponding to a similarity function $sim(A,B)$, we can obtain a locality sensitive hash function $\mathcal{F}'$ ($h':2^U\rightarrow\{0,1\}$ for each $h'\in\mathcal{F}'$) corresponding to the similarity function $\frac{1+sim(A,B)}{2}$.