ECE 543: Statistical Learning Theory

Maxim Raginsky

February 12, 2021

Homework 1

Assigned February 11; due February 24, 2021

Note: natural logarithms are used throughout.

1. **A concentration bound for χ^2 random variables.** Let $X_1, \dots, X_n$ be i.i.d. Gaussian random variables with mean 0 and variance 1. The sum of their squares, $U = X_1^2 + \dots + X_n^2$ has the χ^2 *distribution with n degrees of freedom.* In this problem, you will prove the following tail bound for U:

$$\mathbf{P}\left(U - n \geq 2\sqrt{nt} + 2t\right) \leq e^{-t}, \qquad \text{for all } t > 0. \tag{1}$$

(a) Let Z be a real-valued random variable, such that the bound

$$\log \mathbf{E}[e^{sZ}] \leq \frac{vs^2}{2(1-cs)} \tag{2}$$

holds with some $v, c > 0$ for all $0 < s < 1/c$. Use the Chernoff bounding trick to prove that

$$\mathbf{P}\left(Z \geq \sqrt{2vt} + ct\right) \leq e^{-t}$$

for all $t > 0$.

(b) Now use the result of part (a) to prove (1).

Hint: Show that $Z = U - n$ satisfies (2) for suitable choices of v and c. You may also find the following inequality useful:

$$-s - \frac{1}{2}\log(1-2s) \leq \frac{s^2}{1-2s}, \quad 0 < s < 1/2.$$

2. **Convexity.** Let I be an interval of the real line. A function $f : I \to \mathbb{R}$ is called *convex* if

$$f\big(\lambda x + (1-\lambda)y\big) \leq \lambda f(x) + (1-\lambda) f(y)$$

for all $\lambda \in [0,1]$ and all $x, y \in I$. Equivalently, f is convex if the straight line segment joining the points $(x, f(x))$ and $(y, f(y))$ for any two $x, y \in I$ lies above the graph of f. Here are some useful facts about convex functions:

- **Second-order condition.** If I is an open interval and f is twice differentiable on I, then it is convex if and only if $f''(x) \geq 0$ for all $x \in I$.
- **Jensen's inequality.** If $f : I \to \mathbb{R}$ is convex, then for any random variable X with values in I,

$$f(\mathbf{E}[X]) \leq \mathbf{E}[f(X)].$$

In this problem, you will get to explore the world of convex functions.

(a) Given a pair $x, y \in I$, consider the function $F_{x,y} : [0,1] \to \mathbb{R}$, defined by

$$F_{x,y}(t) = f\big(x + t(y-x)\big).$$

Prove that f is convex if and only if $F_{x,y}$ is convex for all $x, y \in I$.

Note: the function f is not necessarily differentiable.

(b) Suppose that $g : [0,a] \to \mathbb{R}$ is convex and monotone increasing. Prove that the function $f(x) = g(|x|)$ is convex on the interval $[-a, a]$.

(c) Use convexity to prove the following inequality: for any $a > 0$,

$$e^{ax} \leq \cosh a + x \sinh a, \qquad -1 \leq x \leq 1.$$

(d) Let U be a real-valued random variable. Prove that its logarithmic moment-generating function $\psi(a) = \log \mathbf{E}[e^{aX}]$ is convex on the real line. (You may assume that interchanging derivative and expectation is permissible.)

3. **Improving the Hoeffding bound.** Let $X_1, \ldots, X_n$ be n independent Bernoulli(θ) random variables. Their sum $S = X_1 + \ldots + X_n$ is a Binomial(n, θ) random variable. The Hoeffding bound tells us that, for any $\alpha \in [\theta, 1]$,

$$\mathbf{P}\left(S \geq \alpha n\right) \leq e^{-2n(\alpha-\theta)^2}. \tag{3}$$

In this problem, you will obtain an improvement of (3).

(a) Use the Chernoff bounding trick to show that, for any $\alpha \in [\theta, 1]$,

$$\mathbf{P}\left(S \geq \alpha n\right) \leq e^{-nd(\alpha\|\theta)}, \tag{4}$$

where

$$d(\alpha\|\theta) := \alpha \log \frac{\alpha}{\theta} + (1-\alpha) \log \frac{1-\alpha}{1-\theta}$$

is the Kullback–Leibler divergence (or relative entropy) between Bernoulli(α) and Bernoulli(θ) random variables.

(b) Prove that the bound in (4) is, indeed, tighter than the Hoeffding bound (3).

4. **Generalizing Hoeffding's inequality.** In class, we have proved Hoeffding's inequality that gives an exponential bound on the deviation probability $\mathbf{P}[|X_1 + \ldots + X_n| \geq t]$ for a sum of independent random variables that are bounded and have zero mean. In this problem, you will develop a generalization of Hoeffding's inequality to sums of dependent random variables that satisfy a certain weak orthogonality condition.

(a) In preparation for the rest of the problem, derive the inequality

$$\cosh x \leq e^{x^2/2}, \qquad x \in \mathbb{R}$$

as a consequence of Hoeffding's lemma.

Hint: Find a suitable bounded random variable U, such that $\cosh x = \mathbf{E}[e^{xU}]$.

(b) We say that a collection $X_1, \ldots, X_n$ of random variables is a *multiplicative system* if, for any $1 \leq k \leq n$ and any set of k indices $1 \leq i_1 < i_2 < \ldots < i_k \leq n$,

$$\mathbf{E}[X_{i_1} X_{i_2} \ldots X_{i_k}] = 0.$$

Prove that if $X_1, \ldots, X_n$ are a multiplicative system, then

$$\mathbf{E}\left[\prod_{i=1}^{n}(a_i X_i + b_i)\right] = \prod_{i=1}^{n} b_i$$

for any choice of real constants $a_1, \ldots, a_n$ and $b_1, \ldots, b_n$.

(c) Let $U_1, \ldots, U_n$ be n possibly dependent random variables, and let Z be any real-valued random variable jointly distributed with them. For each i, define the martingale difference $X_i = \mathbf{E}[Z|U^i] - \mathbf{E}[Z|U^{i-1}]$ (where $\mathbf{E}[Z|U^0] \equiv \mathbf{E}Z$). Prove that $X_1, \ldots, X_n$ are a multiplicative system.

(d) Consider a multiplicative system $X_1, \ldots, X_n$, such that $-c_i \leq X_i \leq c_i$ for each i, where $c_i > 0$ are some finite constants. Prove that, for any $t > 0$,

$$\mathbf{E}\left[\exp\left(t \sum_{i=1}^{n} X_i\right)\right] \leq \prod_{i=1}^{n} \cosh(tc_i).$$

(e) Now for the final step: prove that if $X_1, \ldots, X_n$ are a multiplicative system of random variables satisfying the boundedness condition of part (c), then

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

5. **Bin packing.** This is a classical application of McDiarmid's inequality. Let $X_1, \ldots, X_n$ be i.i.d. random variables taking values in $[0, 1]$. Each X_i is the size of a package to be shipped. The packages are shipped in bin of size 1, so each bin can hold any set of packages whose

sizes sum to at most 1. Let $B_n = f(X_1, \ldots, X_n)$ be the minimal number of bins needed to ship the packages with sizes $X_1, \ldots, X_n$. Computing B_n is a hard combinatorial optimization problem; however, we can say something about its mean and tail behavior.

(a) Let μ be the common mean of the X_i's. Prove that $\mathbf{E} B_n \geq n\mu$.

(b) Prove that, for any $\varepsilon > 0$,

$$\mathbf{P}\left(\frac{B_n}{n} \leq \mu - \varepsilon\right) \leq \exp\left(-2n\varepsilon^2\right).$$