

ECE 543: Statistical Learning Theory

Maxim Raginsky

Homework 1

Assigned January 25; due February 1

Required reading: lecture notes of Hajek and Raginsky, Chapters 1 and 2

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 :

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

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

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

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

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

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

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

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

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

$$\mathbf{P}(S \geq \alpha n) \leq e^{-nd(\alpha \parallel \theta)}, \quad (4)$$

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).

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

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

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 $f : I \rightarrow \mathbb{R}$ 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 $F_{x,y} : [0, 1] \rightarrow \mathbb{R}$, defined by

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

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

(b) Suppose that $g : [0, a] \rightarrow \mathbb{R}$ 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 $a > 0$,

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

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

4. **Generalizing Hoeffding's inequality.** In class, we have proved Hoeffding's inequality that gives an exponential bound on the deviation probability $\mathbf{P}[|X_1 + \dots + X_n| \geq t]$ 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

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

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, \dots, X_n of random variables is a *multiplicative system* if, for any $1 \leq k \leq n$ and any set of k indices $1 \leq i_1 < i_2 < \dots < i_k \leq n$,

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

Prove that if X_1, \dots, 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, \dots, a_n and b_1, \dots, b_n .

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

(d) Consider a multiplicative system X_1, \dots, 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, \dots, X_n are a multiplicative system of random variables satisfying the boundedness condition of part (d), 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. **Random graphs.** An Erdős–Rényi random graph $G(n, p)$ is a graph with n vertices, such that, for every pair of vertices v, v' , there is an edge between them with probability p , independently of all others. A coloring of the graph is an assignment of a color to each vertex, such that any two vertices connected by an edge have distinct colors. The *chromatic number* χ is the minimal number of colors needed to color the graph. Show that

$$\mathbf{P}[|\chi - \mathbf{E}\chi| \geq t\sqrt{n}] \leq 2e^{-t^2}.$$