COMS 4772 Fall 2016 Homework 1

Due Friday, September 30

Instructions:

- Pick four of the following five problems to be graded. (If you do not designate which problems should be graded, we will pick arbitrarily for you.)
- The usual homework policies (`http://www.cs.columbia.edu/~djhsu/coms4772-f16/about.html`) are, of course, in effect.
- Using this LaTeX template will be helpful for grading purposes.

Problem 1 (25 points)**.** In this problem, "volume" refers to $(d-1)$-dimensional volume (or "surface area" in d-dimensions).

(a) Prove that there is a constant $C > 0$ (not depending on d) such that, for any set $T \subset S^{d-1}$ of $|T| = d^{100}$ unit vectors, the set

$$\bigcap_{\boldsymbol{u} \in T} \left\{ \boldsymbol{x} \in S^{d-1} : \left| \langle \boldsymbol{u}, \boldsymbol{x} \rangle \right| \leq C \sqrt{\frac{\ln d}{d}} \right\}$$

accounts for 99% of the volume of S^{d-1}. (Assume $d \geq 2$ so $\ln(d) > 0$.)

(b) Prove that there is a constant $c > 0$ (not depending on d) such that, for any $\boldsymbol{u} \in S^{d-1}$, the set

$$\left\{ \boldsymbol{x} \in S^{d-1} : \left| \langle \boldsymbol{u}, \boldsymbol{x} \rangle \right| > \frac{c}{\sqrt{d}} \right\}$$

accounts for 99% of the volume of S^{d-1}.

Solution. □

Problem 2 (25 points)**.** Let $B_1^d := \{\boldsymbol{x} \in \mathbb{R}^d : \sum_{i=1}^d |x_i| \leq 1\}$ denote the d-dimensional *cross polytope* (as explained in Ball's lecture notes).

(a) Prove that $B^d \subseteq \sqrt{d}B_1^d$.

(b) Use the fact $B^d \subseteq \sqrt{d}B_1^d$ to derive a bound on the volume of B^d of the form

$$\mathrm{vol}(B^d) \;\leq\; c \cdot \left(\frac{c'}{d}\right)^{d/2}$$

for some positive constants $c, c' > 0$. Explain each step in your derivation.

Hint: Stirling's approximation implies $\sqrt{2\pi}n^{n+1/2}e^{-n} \leq n! \leq n^{n+1/2}e^{1-n}$ for all $n \in \mathbb{N}$.

Solution. $\square$

Problem 3 (25 points)**.** Let X be an $[a, b]$-valued random variable (i.e., $\mathbb{P}(X \in [a, b]) = 1$) with $\mathbb{E}(X) = 0$. For simplicity, assume X has a probability density function f. In this problem, you will prove $\psi_X(\lambda) \leq (b-a)^2/8$ using a technique due to McAllester and Ortiz (2003).

(a) Consider the family of density functions $\{g_\lambda : \lambda \in \mathbb{R}\}$, where

$$g_\lambda(x) \;:=\; \frac{e^{\lambda x}}{\mathbb{E}e^{\lambda X}} f(x) \quad \text{for all } x \in \mathbb{R}\,.$$

Briefly verify that if $Y_\lambda \sim g_\lambda$, then

$$\begin{aligned} \mathbb{E}(Y_\lambda) &= \psi'_X(\lambda)\,, \\ \operatorname{var}(Y_\lambda) &= \psi''_X(\lambda)\,, \end{aligned}$$

where ψ'_X is the first-derivative of ψ_X, and ψ''_X is the second-derivative of ψ_X. (You don't need to write much at all for this part.)

(b) Prove that any $[a, b]$-valued random variable has variance at most $(b-a)^2/4$.

(c) The fundamental theorem of calculus implies

$$\psi_X(\lambda) \;=\; \int_0^\lambda \int_0^\mu \psi''_X(\gamma)\,\mathrm{d}\gamma\,\mathrm{d}\mu\,.$$

Use this identity and the results of parts (a) and (b) to prove that $\psi_X(\lambda) \leq (b-a)^2/8$.

Solution. □

Problem 4 (25 points)**.** Let $\boldsymbol{U}$ be a random unit vector with the uniform density on S^{d-1}, and let $X := \langle \boldsymbol{v}, \boldsymbol{U} \rangle$, where $\boldsymbol{v}$ is a fixed unit vector in S^{d-1}.

(a) Prove that $\psi_{X^2-\mathbb{E}(X^2)}(\lambda) \le \psi_{Z^2-1}(\lambda/d)$ for all $\lambda \in \mathbb{R}$, where $Z \sim \mathrm{N}(0,1)$.

Hint: You may use the fact that if $Y_d \sim \chi^2(d)$ and $\boldsymbol{U}$ are independent, then $\sqrt{Y_d}\boldsymbol{U} \sim \mathrm{N}(\boldsymbol{0}, \boldsymbol{I})$ (standard multivariate Gaussian in $\mathbb{R}^d$). Jensen's inequality may also be useful.

(b) Use the result of part (a) to derive a tail bound for $\left|X^2 - \mathbb{E}(X^2)\right|$. Explain each step in your derivation.

Solution. □

Problem 5 (25 points)**.** Let $\Phi \colon \mathbb{R} \to [0,1]$ denote the cumulative distribution function for $\mathrm{N}(0,1)$, i.e., $\Phi(t) = \mathbb{P}(Z \le t)$ where $Z \sim \mathrm{N}(0,1)$. Prove that for any $d \in \mathbb{N}$, if

1. $X_1, X_2, \ldots, X_d$ are independent random variables;
2. $\mathbb{E}X_i = 0$ and $\mathbb{E}X_i^2 = 1$ for all $i \in [d]$;

then for a $1 - o(1)$ fraction of unit vectors $\boldsymbol{u} \in S^{d-1}$, the random vector $\boldsymbol{X} = (X_1, X_2, \ldots, X_d)$ satisfies

$$\sup_{t \in \mathbb{R}} \left| \mathbb{P}\left(\langle \boldsymbol{u}, \boldsymbol{X} \rangle \le t\right) - \Phi(t) \right| \ \le \ O\left(\frac{\rho}{d^{0.49}}\right) ,$$

where $\rho := \max_{i \in [d]} \mathbb{E}|X_i|^3$.

You can use the following theorem (which you do not need to prove):

Theorem 1 (Berry-Esséen theorem)**.** *There is an absolute positive constant $C > 0$ such that the following holds. Let $Y_1, Y_2, \ldots, Y_n$ be independent random variables with $\mathbb{E}Y_i = 0$, $\sigma_i^2 := \mathbb{E}Y_i^2 < \infty$. Define $v_n := \sum_{i=1}^n \sigma_i^2$ and $\rho_n := \sum_{i=1}^n \mathbb{E}|Y_i|^3$. Then*

$$\sup_{t \in \mathbb{R}} \left| \mathbb{P}\left(\frac{Y_1 + Y_2 + \cdots + Y_n}{\sqrt{v_n}} \le t\right) - \Phi(t) \right| \ \le \ \frac{C\rho_n}{v_n^{3/2}} .$$

Solution. □

References

D. McAllester and L. Ortiz. Concentration inequalities for the missing mass and for histogram rule error. *Journal of Machine Learning Research*, 4(Oct):895–911, 2003.