36-755, Fall 2017
Homework 2

Due Wed Oct 4 by 5:00pm in Jisu's mailbox

1. Let $(X_1, \ldots, X_n)$ be zero-mean $SG(\sigma^2)$ random variables (not assumed independent). Give bounds on
$$\mathbb{E}\left[\max_i |X_i|\right]$$
and
$$\mathbb{P}\left(\max_i |X_i| \geq t\right), \quad t \geq 0.$$
You can use the corresponding result, proved in class, for $\max_i X_i$.

2. Let $(X_1, \ldots, X_n)$ be independent random variables with mean zero and let $(a_1, \ldots, a_n) \in \mathbb{R}^n$. Compute bounds for
$$\mathbb{P}\left(|\sum_{i=1}^n a_i X_i| \geq t\right)$$
under the assumption that the X_i's are in $SG(\sigma^2)$ and in $SE(\nu^2, \alpha)$. Compare the bounds. When does one dominate the other?

3. Orlicz norms. We have defined sub-gaussian and sub-exponential variables in terms of bounds on the moment generating functions. There exists an equivalent and more general way of expressing these properties using *Orlicz Norms* of random variables, which is more abstract but, at the same time, leads to simpler calculation. You will explore these concepts in this exercise. First, do the following problems in the book:

   (a) 2.18 and

   (b) 2.19.

   In this context, a random variables is said to be sub-gaussian if there exists a $K > 0$ such that
$$\mathbb{E}\left[e^{X^2/K^2}\right] \leq 2 \tag{1}$$
   and sub-exponential if there exists a constant $K' > 0$ such that
$$\mathbb{E}\left[e^{|X|/K'}\right] \leq 2. \tag{2}$$
   If X is sub-gaussian, its *sub-gaussian norm* is the smallest K satisfying (1), which correspond to $\|X\|_{\psi_2}$. Similarly, if X is sub-exponential, its *sub-exponential norm* is $\|X\|_{\psi_1}$, the smallest K' satisfying (2).

   (c) Prove that X is sub-gaussian if and only if X^2 is sub-exponential and
$$\|X^2\|_{\psi_1} = \|X\|_{\psi_2}^2$$

   (d) If X and Y are sub-gaussians, then XY is sub-exponential with
$$\|XY\|_{\psi_1} \leq \|X\|_{\psi_2} \|Y\|_{\psi_2}.$$

The last two properties would have made problems 2 and 8 in Homework 1 easier...

Remarks. (Please read) it is possible to show that the above definitions are equivalent to the ones given in class: see the Appendix of Chapter 2 of the textbook. In particular, if X is sub-exponential then

$$\mathbb{E}\left[e^{\lambda X}\right] \leq \exp \lambda^2 \|X\|_{\psi_1}^2, \qquad \forall |\lambda| \leq \frac{1}{\|X\|_{\psi_1}}.$$

From this, it is possible to derive the following, equivalent, versions of Hoeffding and Bernstein inequalities which you will also find in the literature.

- **Hoeffding inequality**. Let $X_1, \ldots, X_n$ be independent, mean-zero sub-gaussian variables. Then, there exists a universal constant $c > 0$ such that, for any $t \geq 0$,

$$\mathbb{P}\left(\left|\sum_{i=1}^n X_i\right| \geq t\right) \leq 2\exp\left(-\frac{ct^2}{\sum_{i=1}^n \|X_i\|_{\psi_2}^2}\right)$$

- **Bernestein inequality**. Let $X_1, \ldots, X_n$ be independent, mean-zero sub-exponential variables. Then, there exists a universal constant $c > 0$ such that, for any $t \geq 0$,

$$\mathbb{P}\left(\left|\sum_{i=1}^n X_i\right| \geq t\right) \leq 2\exp\left(-c\min\left\{-\frac{t^2}{\sum_{i=1}^n \|X_i\|_{\psi_1}^2}, \frac{t}{\sum_{i=1}^n \|X_i\|_{\psi_1}}\right\}\right)$$

In other words, mapping to the notation used in class, $\sigma = \|X\|_{\psi_2}$ and $\nu = \alpha = \|X\|_{\psi_1}$.

4. (Reading exercise. **Not to be graded for correctness, but only for effort**)
Suppose that $X_1, \ldots, X_n$ are zero-mean, independent random variables belonging to the class $SG(\sigma^2)$ and $A = (A_{i,j})$ a $n \times n$ matrix. Let

$$\|A\|_{\text{op}} = \max_{x\in\mathbb{R}^n, x\neq 0} \frac{\|Ax\|}{\|x\|}$$

and

$$\|A\|_{\text{HS}} = \sqrt{\sum_{i=1}^n \sum_{j=1}^n A_{i,j}^2}$$

be the operator and the Hiolbert-Schmidt (or Frobenius) norm of A. Notice that $\|A\|_{\text{op}}$ is also the largest absolute eigenvalue of A. The goal of this exercise is to derive an exponential inequality for the probability

$$\mathbb{P}\left(\left|X^\top AX - \mathbb{E}\left[X^\top AX\right]\right| \geq t\right), \forall t \geq 0.$$

Do so by reproducing the proof of Theorem 1.1 from the following reference, using the definition of sub-Gaussian and sub-Exponential variables given in class.

- Rudelson, M., and Vershynin, R. (2013). Hanson-Wright inequality and sub-gaussian concentration. Electron. Commun. Probab., 18(82), 1- 9.

Notice that the definitions of sub-gaussian and sub-exponential variables in this paper is different than the ones given in class and correspond to the ones in problem 3. Make sure to keep track of the constants that depend on σ^2.

5. **Robust statistics and the median-of-mean estimator**. Suppose we observe n i.i.d. random variables with distribution P and would like to construct a $1-\alpha$ confidence set for the expected value of P, where $\alpha \in (0,1)$.

   (a) If the common distribution P is in the class $SG(\sigma^2)$ provide such a confidence interval.

   (b) Now let's drop the assumption that P is a $SG(\sigma^2)$ distribution and in particular allow for very thick tails.

   How can we proceed?

   Here is a simple method. Assume that $\mathrm{Var}[X] = \sigma^2 < \infty$. For a fixed $\alpha \in [e^{1-n/2}, 1)$, set $b = \lceil \ln(1/\alpha) \rceil$ and note that $b \leq n/2$. Next, partition $[n] = \{1, \ldots, n\}$ into b blocks $B_1, \ldots, B_b$ each of size $|B_i| \geq \lfloor n/b \rfloor \geq 2$ and compute the sample mean in each block:

   $$\overline{X}_i = \frac{1}{|B_i|} \sum_{j \in B_i} X_j, \quad i = 1, \ldots, b.$$

   Finally define **the median-of-means** estimator as

   $$\hat{\mu} = \hat{\mu}(\alpha) = \mathrm{median}\left\{\overline{X}_1, \ldots, \overline{X}_b\right\},$$

   where, for any b-tuple of numbers $(x_1, \ldots, x_b)$,

   $$\mathrm{median}\{x_1, \ldots, x_b\} = x_{j^*},$$

   with

   $$|\{k \in [b] \colon x_k \leq x_{j^*}\}| \geq b/2 \quad \text{and} \quad |\{k \in [b] \colon x_k \geq x_{j^*}\}| \geq b/2,$$

   (if more than one such x_{j^*} satisfies the above inequalities, pick one of them at random).

   Show that the median-of-means estimator yields, up to constants, the same type of sub-Gaussian confidence interval obtained in the first part, but without requiring the assumption of sub-Gaussianity. That is, show that

   $$\mathbb{P}\left(|\hat{\mu} - \mu| \geq C\sqrt{\frac{\sigma^2 \log(1/\alpha)}{n}}\right) \leq \alpha,$$

   for some constant C, where $\sigma^2 = \mathrm{Var}[X]$. You may want to consult these paper:

   - M. Lerasle and R. I. Oliveira (2011). Robust empirical mean estimators.
     `https://arxiv.org/pdf/1112.3914v1.pdf`
   - Luc Devroye, Matthieu Lerasle, Gabor Lugosi and Roberto I. Oliveira (2016). Sub-Gaussian mean estimators.
     `https://arxiv.org/pdf/1509.05845v1.pdf`

   (c) The median-of-means estimator has an obvious drawback. What is it? *Hint: think of the situation when you want to use this estimator to compute confidence intervals at different levels α and α'...*

6. Let $(\mathcal{X}, d)$ be a metric space and, for $\delta > 0$, let $N(\delta)$ and $M(\delta)$ denote the δ-covering and δ-packing number, respectively. Show that

   $$M(2\delta) \leq N(\delta) \leq M(\delta).$$

7. **Efron-Stein inequality** In this exercise you will derive a nice result, known as the Efron-Stein inequality, that yieles useful bounds for the variance of functions of independent variables.

   Let $X_1, \ldots, X_n$ be independent random variables and $Z = f(X_1, \ldots, Z_n)$. We make no assumptions on the function $f \colon \mathbb{R}^n \to \mathbb{R}$ other than $\mathbb{E}[Z^2] < \infty$. For any $i = 1, \ldots, n$, let

$$\mathbb{E}_i[Z] = \mathbb{E}[Z | X_j, j \neq i].$$

   (a) Prove the Efron-Stein Inequality:

$$\mathbb{V}[Z] \leq \sum_{i=1}^{n} \mathbb{E}\left[(Z - \mathbb{E}_i[Z])^2\right]$$

   Hints: by Doob's representation, $Z - \mathbb{E}[Z] = \sum_{i=1}^{n} Y_i$, *for a martingale difference sequence* $(Y_1, \ldots, Y_n)$. *Then show that* $\mathbb{V}[Z] = \sum_{i=1}^{n} \mathbb{E}\left[Y_i^2\right]$. *Finally, show that, for all* i,

$$Y_i^2 \leq \mathbb{E}\left[(Z - \mathbb{E}_i[Z])^2 | X_1, \ldots, X_i\right].$$

   (b) For any $i = 1, \ldots, n$, let $Z_i = f_i(X_1, \ldots, X_{i-1}, X_{i+1}, \ldots, X_n)$, for an arbitrary function $g_i \colon \mathbb{R}^{n-1} \to \mathbb{R}$ of the $(n-1)$ variables X_j, $j \neq i$. Use the Efron-Stein inequality to show that

$$\mathbb{V}[Z] \leq \sum_{i=1}^{n} \mathbb{E}\left[(Z - Z_i)^2)\right].$$

   Hint: use conditionally the fact that, for any random variable X, $\mathbb{V} \leq \mathbb{E}\left[(X - c)^2\right]$, *for any* $c \in \mathbb{R}$.

   (c) Use the previous result to show that, if f satisfies the bounded difference property with constants $(c_1, \ldots, c_n)$, then

$$\mathbb{V}[Z] \leq \frac{1}{4} \sum_{i=1}^{n} c_i^2.$$

   (d) **Application to Kernel density estimation.** Let p be a Lebesgue density over the real line. Le $X_1, \ldots, X_n$ be an i.i.d. sample from the distribution P with density p. Let K be a non-negative function with $\int_{\mathbb{R}} K(t) dt = 1$ (a kernel). For a $h > 0$ (the bandwidth parameter) define the random function $\hat{p}_h$ given by

$$x \in \mathbb{R} \mapsto \hat{p}_h(x) = \frac{1}{nh} \sum_{i=1}^{n} K\left(\frac{x - X_i}{h}\right).$$

   This is the kernel density estimator of p with bandwidth h. Let

$$Z = \int_{\mathbb{R}} |\hat{p}_h(x) - p(x)| dx,$$

   be the L_1 distance between p and $\hat{p}_h$. Show that

$$\mathbb{V}[Z] \leq \frac{1}{n}.$$