36-755, Fall 2016
Homework 2

Due Sep 21 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. (Mill's ratio). Let $\Phi \colon \mathbb{R} \to [0, 1]$ the c.d.f. of the standard Gaussian distribution on $\mathbb{R}$ and ϕ its p.d.f..
   (a) Prove that, for all $x > 0$,
   $$\frac{x}{1+x^2}\phi(x) \leq 1 - \Phi(x) \leq \frac{1}{x}\phi(x)$$
   (b) Prove that, for all $x > 0$,
   $$1 - \Phi(x) \leq \frac{1}{2}\exp\left(-x^2/2\right).$$

4. Show that if $X \in SG(\sigma^2)$ than $X^2 \in SE(\nu^2, \alpha)$ where
$$\nu = \alpha = 16\sigma^2.$$

   Hint: For this problem you may find it helpful to use the following facts:
   (a) **The C_r inequality:** *If X and Y are random variables such that $\mathbb{E}|X|^r < \infty$ and $\mathbb{E}|Y|^r < \infty$, where $r \geq 1$, then*
   $$\mathbb{E}|X+Y|^r \leq 2^{r-1}\left(\mathbb{E}|X|^r + \mathbb{E}|Y|^r\right)$$
   (b) The following bound, proved in class
   $$\mathbb{E}|X|^r \leq (2\sigma^2)^{r/2} r\Gamma(r/2) \quad r \geq 1.$$
   Also, feel free to prove the claim by obtaining sharper (smaller) bounds on ν^2 and/or $1/\alpha$.

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

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

and

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

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

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

for any $t \geq 0$. 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 definition of sub-Gaussianity in this paper is different than the one given in class (though they can be shown to be equivakent). Make sure to keep track of the constants that depend on σ^2.

6. Exercise 2.14 in the book.

7. **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

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

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

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

   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'*...*