Homework 2
36-705
Due: Thursday Sept 6th by 3:00pm via Canvas

1. Let $X$ be a Gaussian RV with mean $\mu$ and variance $\sigma^2$. Show that its mgf for any $t$ is given by,

$$M_X(t) = \exp(t\mu + t^2\sigma^2/2).$$

Show further that a random variable $X$ is sub-Gaussian if and only if $-X$ is sub-Gaussian.

2. Suppose that we have a positive random variable $X \geq 0$, and suppose further that its mgf $M_X(t)$ exists for all $t \geq 0$. We will focus in this question on tail bounds that are not centered around the mean.

There are two ways to use Markov’s inequality: one involves using the moments (like Chebyshev) in which case we obtain a bound of the form:

$$P(X \geq u) \leq \frac{E[X^k]}{u^k}.$$  

An alternative is to use the mgf with a tuning parameter (the Chernoff technique), i.e.

$$P(X \geq u) \leq \inf_{t \geq 0} \frac{E[\exp(tX)]}{\exp(tu)}.$$  

Show that,

$$\inf_{k=0,1,2,...} \frac{E[X^k]}{u^k} \leq \inf_{t \geq 0} \frac{E[\exp(tX)]}{\exp(tu)}.$$  

This in turn implies that the best moment tail bound is never worse than the best Chernoff bound. We usually use the Chernoff method because it is typically much easier to find the best $t$ than to find the best $k$.

**Hint:** The following fact might be useful (if you use it, prove it): if $c \leq \frac{a_i}{b_i}$ for all $i$, then

$$c \leq \frac{\sum_{i=1}^{\infty} a_i}{\sum_{i=1}^{\infty} b_i}.$$  

3. In lecture we claimed that the $U$ statistic with

$$g(X_j, X_k) = \frac{1}{2}(X_j - X_k)^2,$$

leads to the sample variance, i.e.

$$\frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{\mu})^2 = \frac{1}{\binom{n}{2}} \sum_{j<k} g(X_j, X_k).$$

Prove this fact.
4. There are senses in which the Markov and Chebyshev inequalities are tight, i.e. they cannot be improved without using more information about the random variable.

**Markov:** Construct a positive random variable $X$ for which Markov’s inequality is tight, i.e. fix a value $t$, now construct a random variable $X \geq 0$ for which:

$$P(X \geq t) = \frac{\mathbb{E}[X]}{t}.$$ 

**Chebyshev:** Construct a random variable $X$ for which Chebyshev’s inequality is tight in the same sense as above.

**Hint:** The easiest examples for these types of random variables are discrete, i.e. think of point masses of different heights at different values of $X$, and try to find a location, height combination that works.

5. **Being sloppy with constants:** In class we discussed the Hoeffding bound for bounded RVs:

$$\mathbb{P} \left( \left| \frac{1}{n} \sum_{i=1}^{n} X_i - \mu \right| \geq t \right) \leq 2 \exp \left( -\frac{n t^2}{(b-a)^2} \right),$$

and Bernstein’s bound:

$$\mathbb{P} \left( \left| \frac{1}{n} \sum_{i=1}^{n} X_i - \mu \right| \geq t \right) \leq 2 \exp \left( -\frac{n t^2}{2(\sigma^2 + (b-a)t)} \right).$$

Ignore constant factors in everything that follows.

(a) Upper bound the variance $\sigma^2$ in terms of the bounds $(b-a)$.

(b) Use this to argue that ignoring constants, for any regime where the tail bound is “interesting” (i.e. non-trivial) the Bernstein bound is not worse than Hoeffding’s bound. You will have to explain carefully what regime you are focussing on and why.

(c) Show that for many interesting regimes Bernstein’s bound dominates Hoeffding’s bound, i.e. how small does $\sigma$ need to be so that the Bernstein bound is better/tighter than Hoeffding’s bound.