Homework Assignment 2

Due Feb 28th midnight

SDS 384-11 Theoretical Statistics

1. (2+2+1) Consider a r.v. X such that for all $\lambda \in \mathbb{R}$

$$E[e^{\lambda X}] \leq e^{\frac{\lambda^2 \sigma^2}{2} + \lambda\mu} \tag{1}$$

Prove that:

(a) $E[X] = \mu$.

Solution.
Let $f(\lambda) = E[e^{\lambda X}]$ and let $g(\lambda) = e^{\lambda^2\sigma^2/2+\lambda\mu}$. We have $f(0) = g(0)$.

$$f'(0) = \lim_{h \to 0} \frac{f(h) - f(0)}{h} \leq \lim_{h \to 0} \frac{g(h) - g(0)}{h} = g'(0)$$

But we also have:

$$f'(0) = \lim_{h \to 0} \frac{f(0) - f(-h)}{h} \geq \lim_{h \to 0} \frac{g(0) - g(-h)}{h} = g'(0)$$

So $f'(0) = g'(0)$. So we have $E[X] = \mu$.

□

(b) $\text{var}(X) \leq \sigma^2$.

Solution.

Let us denote

$$\begin{aligned} M_c(\lambda) &= \exp(-\lambda\mu)M(\lambda) \\ &= \mathbb{E}[\exp(\lambda(X - \mu))] \end{aligned}$$

and similarly,

$$\begin{aligned} U_c(\lambda) &= \exp(-\lambda\mu)U(\lambda) \\ &= \exp\left(\frac{\lambda^2\sigma^2}{2}\right) \end{aligned}$$

Then, by construction, we have that $M_c(\lambda) \leq U_c(\lambda)$. Additionally, $M_c(0) = 1 = U_c(0)$, $M_c''(0) = \text{var}(X)$, and $U_c''(0) = \sigma^2$. Therefore, we have that

$$\begin{aligned}\text{var}(X) &= M_c''(0) \\ &= \lim_{\varepsilon \to 0} \frac{M_c(\varepsilon) + M_c(-\varepsilon) - 2M_c(0)}{\varepsilon^2} \\ &= \lim_{\varepsilon \to 0} \frac{M_c(\varepsilon) + M_c(-\varepsilon) - 2U_c(0)}{\varepsilon^2} \\ &\leq \lim_{\varepsilon \to 0} \frac{U_c(\varepsilon) + U_c(-\varepsilon) - 2U_c(0)}{\varepsilon^2} \\ &= U_c''(0) \\ &= \sigma^2\end{aligned}$$

which establishes the desired inequality.

□

(c) If the smallest value of σ satisfying the above equation is chosen, is it true that $\text{var}(X) = \sigma^2$? Prove or give a counter-example.

Solution.
We give a counterexample to establish that $\sigma^2 \neq \text{var}(X)$. Consider $X \sim \text{Bern}\,(p)$. Then, assuming that $\sigma^2 = p(1-p) = \text{var}(X)$, we have that

$$\begin{aligned}\mathbb{E}\left[\exp\left(\lambda\left(X - p\right)\right)\right] &= p\exp(\lambda(1-p)) + (1-p)\exp(-\lambda p) \\ &= \exp(\lambda(1-p))\left(p + (1-p)\exp(-\lambda)\right) \\ &\leq \exp\left(\frac{\lambda^2 p(1-p)}{2}\right) \qquad \text{by assumed subG bound} \\ \implies p + (1-p)\exp(-\lambda) &\leq \exp\left(\lambda(1-p)\left(\frac{\lambda p}{2} - 1\right)\right) \qquad (2)\end{aligned}$$

However, by choosing, for example, $\lambda = \frac{1}{4}$ and $p = \frac{1}{16}$, one can check that

$$p + (1-p)\exp(-\lambda) - \exp\left(\lambda(1-p)\left(\frac{\lambda p}{2} - 1\right)\right) \approx 0.0001 > 0$$

which is a *contradiction* of inequality (2). Therefore, we cannot always take $\sigma^2 = \text{var}(X)$.

□

2. (5pts) Given a symmetric positive semidefinite matrix $Q \in \mathbb{R}^{n\times n}$, consider $Z = \sum_{i,j} Q_{ij} X_i X_j$. When $X_i \sim N(0,1)$, prove the Hanson-Wright inequality.

$$P\left(Z \geq \text{trace}(Q) + t\right) \leq \exp\left(-\min\left\{c_1 t/\|Q\|_{op}, c_2 t^2/\|Q\|_F^2\right\}\right),$$

where $\|Q\|_{op}$ and $\|Q\|_F$ denote the operator and frobenius norms respectively. *Useful facts: Let $\lambda_1 \geq \lambda_2 \geq \ldots$ denote the eigenvalues of Q. Remember that $\|Q\|_{op} = \sup_{v:\|v\|=1} \|Qv\| = \lambda_1$. For a PSD matrix Q, trace$(Q) = \sum_i \lambda_i$, and $\|Q\|_F^2 = \sum_i \lambda_i^2$. Hint: The rotation-invariance of the Gaussian distribution and sub-exponential nature of χ^2-variables could be useful.*

3. (5+5+5) We will prove properties of subgaussian random variables here. Prove that:

(a) Moments of a mean zero subgaussian r.v. X with variance proxy σ^2 satisfy:

$$E[|X^k|] \leq k2^{k/2}\sigma^k\Gamma(k/2), \tag{3}$$

where Γ is the gamma function.

Solution.
We have that, by the Subgaussian assumption,

$$\begin{aligned}\mathbb{E}[|X|^k] &= \int_0^\infty \mathbb{P}(|X|^k > t)dt \\ &= \int_0^\infty \mathbb{P}(|X| > t^{1/k})dt \\ &\leq 2\int_0^\infty \exp\left(-\frac{t^2}{2\sigma^2}\right)dt\end{aligned}$$

Now, recalling that

$$\Gamma(z) = \int_0^\infty t^{z-1}\exp(-t)dt,$$

we may perform the change of variables $t = ax^b$ to obtain:

$$\begin{aligned}\Gamma(z) &= \int_0^\infty a^{z-1}x^{bz-b}\exp(-ax^b)abx^{b-1}dx \\ &= a^z b\int_0^\infty x^{bz-1}\exp(-ax^b)dx\end{aligned}$$

Thus,

$$\Gamma\left(\frac{1}{b}\right) = a^{\frac{1}{b}}b\int_0^\infty \exp(-ax^b)dx$$

Now, choosing $a = \frac{1}{2\sigma^2}$ and $b = \frac{2}{k}$, we combine our results to obtain:

$$\begin{aligned}\mathbb{E}[|X|^k] &\leq 2\int_0^\infty \exp\left(-\frac{t^2}{2\sigma^2}\right)dt \\ &= \frac{2}{a^{\frac{1}{b}}b}\Gamma\left(\frac{1}{b}\right) \\ &= \frac{2}{\left(\frac{1}{2\sigma^2}\right)^{\frac{k}{2}}\frac{2}{k}}\Gamma\left(\frac{k}{2}\right) \\ &= k2^{k/2}\sigma^k\Gamma\left(\frac{k}{2}\right)\end{aligned}$$

as desired. □

(b) If X is a mean 0 subgaussian r.v. with variance proxy σ^2, prove that, $X^2 - E[X^2]$ is a subexponential $(c_1\sigma^2, c_2\sigma^2)$ (we are using the (ν, b) parametrization of subexponentials we did in class, so ν^2 is the variance proxy). Here c_1, c_2 are positive constants. Many people will use $E[(X^2 - [EX]^2)^k] \leq E[X^{2k}]$. This is wrong. 1 pt off for this. For those who use the definition of sub-exponentiality from scratch, if they assume $t > 0$, take 1/2 point off. See the alternative solution here.

I am going to give two different solutions here. And point out common mistakes you may make.The first uses Bernstein's moment condition. In class we did a very easy bounded random variable example to show it is subexponential since it satisfies the bernstein m.c. Here is a far less trivial example of its use. The second solution gets to the answer through the definition of sub-exp r.v.s as we saw in class.

Solution.
Here, we wish to apply the Bernstein condition. Observe that

$$\begin{aligned}
&\left|\mathbb{E}(X^2 - \mathbb{E}X^2)^k\right| \\
&\leq \mathbb{E}\left|X^2 - \mathbb{E}X^2\right|^k && \text{Jensen's} \\
&= \mathbb{E}\left(\left|X^2 - \mathbb{E}X^2\right|^k 1\{X^2 \geq \mathbb{E}X^2\}\right) + \mathbb{E}\left(\left|X^2 - \mathbb{E}X^2\right|^k 1\{X^2 < \mathbb{E}X^2\}\right)
\end{aligned}$$

Now, observe that, almost surely,

$$\begin{aligned}
|X^2 - \mathbb{E}X^2|^k 1\{X^2 \geq \mathbb{E}X^2\} &\leq |X^2|^k 1\{X^2 \geq \mathbb{E}X^2\} && \text{since } \mathbb{E}X^2 \geq 0 \\
&\leq |X|^{2k} && \text{since } 1\{\cdot\} \leq 1 \text{ a.s.}
\end{aligned}$$

and similarly,

$$\begin{aligned}
|X^2 - \mathbb{E}X^2|^k 1\{X^2 < \mathbb{E}X^2\} &\leq |\mathbb{E}X^2|^k 1\{X^2 < \mathbb{E}X^2\} && \text{since } X^2 \geq 0 \text{ a.s.} \\
&\leq |\mathbb{E}X^2|^k && \text{since } 1\{\cdot\} \leq 1 \text{ a.s.} \\
&\leq \mathbb{E}|X|^{2k} && \text{by Jensen's, since } |\cdot|^k \text{ is convex}
\end{aligned}$$

Note the treatment above. Many of you may bound $E[(X^2 - [EX]^2)^k] \leq E[X^{2k}]$. This is incorrect, because $|X^2 - E[X^2]| \leq \max(X^2, E[X^2])$. I am gong to take a point off for this, just so that this sticks in our minds.
Finally, note that

$$\begin{aligned}
\operatorname{var}(X^2) &\leq \mathbb{E}X^4 \\
&\leq 4 \cdot 2^2 \sigma^4 \Gamma(2) \\
&= 2^4 \sigma^4 \\
&< 2^5 \sigma^4
\end{aligned}$$

Combining these bounds, we have that

$$\begin{aligned}\left|\mathbb{E}(X^2-\mathbb{E}X^2)^k\right| &\le 2\mathbb{E}|X|^{2k} \\ &\le 4k2^k\sigma^{2k}\underbrace{\Gamma(k)}_{=(k-1)!} \qquad \text{by the previous exercise} \\ &= \frac{1}{2}k!2^5\sigma^4\left(2\sigma^2\right)^{k-2}\end{aligned}$$

Therefore, by the Bernstein condition, we have that X^2 is subexponential with paratmeters $(\nu = 8\sigma^2, b = 4\sigma^2)$, as desired.

□

Solution.
Now we will prove the subexponentiality using the MGF. Note that we have $E[X^2] \le \sigma^2$.

$$\begin{aligned}E[\exp(\lambda(X^2-E[X^2]))] &\le \exp(-\lambda E[X^2])E[\exp(\lambda X^2)] \\ &= \exp(-\lambda E[X^2])\left(1+\lambda E[X^2]+\sum_{k\ge 2}\lambda^k\frac{E[X^{2k}]}{k!}\right) \\ &= \exp(-\lambda E[X^2])\left(1+\lambda E[X^2]+2\sum_{k\ge 2}2^k\sigma^{2k}|\lambda|^k\right) \\ (\text{For } |\lambda|<1/2\sigma^2, \text{ we have}) &= \underbrace{\exp(-\lambda E[X^2])\left(1+\lambda E[X^2]\right)}_{\text{This is } \le 1 \text{ since } \exp(x)\ge 1+x}+\frac{8\sigma^4\lambda^2\exp(-\lambda E[X^2])}{1-2\sigma^2|\lambda|} \\ (\text{For } |\lambda|<1/4\sigma^2, \text{ we have}) &\le 1+\underbrace{16\sigma^4\lambda^2\exp(|\lambda|\sigma^2)}_{E[X^2]\le\sigma^2} \le 1+\underbrace{16\sigma^4\lambda^2\exp(1/4)}_{|\lambda|\le 1/4\sigma^2} \\ &\le 1+2^5\sigma^4\lambda^2 \le \underbrace{\exp((4\sqrt{2}\sigma^2)^2\lambda^2)}_{\exp(x)\ge 1+x}\end{aligned}$$

So we have $X^2 - E[X^2]$ is sub exponential $(8\sigma^2, 4\sigma^2)$.

□

(c) Consider two independent mean zero subgaussian r.v.s X_1 and X_2 with variance proxies σ_1^2 and σ_2^2 respectively. Show that X_1X_2 is a subexponential r.v. with parameters $(d_1\sigma_1\sigma_2, d_2\sigma_1\sigma_2)$. Here d_1, d_2 are positive constants. Again, here also, I think most people will get the right answer if they end up proving it. So full score unless you see obvious mistakes.

Solution.

Observe that,

$$
\begin{aligned}
\mathbb{E}[(X_1X_2 - \mathbb{E}[X_1X_2])^k] &= \mathbb{E}[(X_1X_2 - \mathbb{E}[X_1]\mathbb{E}[X_2])^k] && \text{by independence} \\
&= \mathbb{E}[(X_1X_2)^k] && \text{mean 0} \\
&\le \mathbb{E}[|X_1X_2|^k] \\
&= \mathbb{E}[|X_1|^k]\mathbb{E}[|X_2|^k] && \text{independence} \\
&\le \left(k2^{k/2}\sigma_1^k\Gamma\left(\frac{k}{2}\right)\right)\left(k2^{k/2}\sigma_2^k\Gamma\left(\frac{k}{2}\right)\right) && \text{by part 1} \\
&= \left(k\Gamma\left(\frac{k}{2}\right)\right)^2 2^k(\sigma_1\sigma_2)^2
\end{aligned}
$$

Now, recall that, for k an odd integer,

$$
\begin{aligned}
\Gamma\left(\frac{k}{2}\right) &= \Gamma\left(\left\lfloor\frac{k}{2}\right\rfloor + \frac{1}{2}\right) \\
&= \sqrt{\pi}\frac{(2\left\lfloor\frac{k}{2}\right\rfloor)!}{4^{\lfloor k/2\rfloor}\lfloor k/2\rfloor!} \\
&= \sqrt{\pi}\frac{2\,(k-1)!}{4^{k/2}\lfloor k/2\rfloor!}
\end{aligned}
$$

Thus, we have that

$$
\begin{aligned}
\left(k\Gamma\left(\frac{k}{2}\right)\right)^2 &= \pi k^2\frac{((k-1)!)^2}{4^k\left(\lfloor k/2\rfloor\right)^2} && (4) \\
&\le k! \\
\iff \pi k! &\le 4^k\lfloor k/2\rfloor! && (5)
\end{aligned}
$$

Now, note that (4) is true for sufficiently large k. Similarly, when k is even,

$$\Gamma\left(\frac{k}{2}\right) = \left(\frac{k}{2} - 1\right)!$$

so we have that

$$
\begin{aligned}
\left(k\Gamma\left(\frac{k}{2}\right)\right)^2 &= k^2\left(\left(\frac{k}{2}-1\right)!\right)^2 \\
&\le k! && (6) \\
\iff k\left(\frac{k}{2}-1\right)! &\le \prod_{i=1}^{\frac{k}{2}-1}(k-i) \\
\iff 1 &\le \frac{k-1}{k}\prod_{i=2}^{\frac{k}{2}-1}\frac{k-i}{\frac{k}{2}+1-i} && (7)
\end{aligned}
$$

Observe that (6) is true for sufficiently large k. Therefore, there exists a universal constant C such that

$$\begin{aligned}\mathbb{E}[(X_1X_2 - \mathbb{E}[X_1X_2])^k] &= \left(k\Gamma\left(\frac{k}{2}\right)\right)^2 2^k(\sigma_1\sigma_2)^2 \\ &\le Ck!2^k(\sigma_1\sigma_2)^k \\ &\le \frac{1}{2}k!(\sigma_1\sigma_2)^2(\tilde{C}\sigma_1\sigma_2)^k\end{aligned}$$

For sufficiently large $\tilde{C}$. Therefore, since $\text{var}(X_1X_2) \le \sigma_1^2\sigma_2^2$, by Bernstein's theorem, X_1X_2 is subexponential with parameters $(\nu = \sqrt{2}\sigma_1\sigma_2, b = 2\tilde{C}\sigma_1\sigma_2)$. This establishes the desired result.

□

4. (1+4) Subgaussian and subexponential random variables have moments that are growing suitably so that we can have a bound on the MGF. Consider scalar random variables $X_1, \ldots, X_n$ that are IID samples from some distribution with mean μ. What if all we have is an upper bound on the variance, i.e. $E[(X_1 - \mu)^2] \le \sigma^2 < \infty$ - are there estimators for which we can obtain exponential tail bounds? This is what we will learn through this exercise. Assume $n = mk$ for some positive integers m, k. Divide the data into k disjoint chunks. For each chunk, compute the mean, call this m_i, $i = 1, \ldots, k$. Let your estimator be $\widehat{\mu}_n := \text{median}(\{m_i\}_{i=1}^k)$. We will show that, for some appropriately picked $k = k_\delta$,

$$P\left(|\widehat{\mu}_n - \mu| \ge c\sigma\sqrt{\frac{\log(1/\delta)}{n}}\right) \le \delta \tag{8}$$

where c is a constant.

(a) First show that, for $i \in \{1, \ldots, k\}$ $P\left(|m_i - \mu| \ge \frac{\sigma}{2\sqrt{m}}\right) \le 1/4$

Solution.
This is just Chebychev's inequality applied to every partition. □

(b) Now find a suitable k as a function of δ, such that Eq 8 holds. *Hint: Use the definition of a median to frame Eq 8 as a failure probability of a sum of k independent Bernoulli(p_i) RVs with $p_i \ge 1/4$.*

Solution.
Let $N = \left|\{i : |m_i - \mu| \ge c\sigma\sqrt{\frac{\log(1/\delta)}{n}}\}\right|$.

$$p := P\left(|m_i - \mu| \ge c\sigma\sqrt{\frac{\log(1/\delta)}{n}}\right)$$

Note $N \sim Bin(k,p)$.

$$P\left(|\widehat{\mu}_n - \mu| \geq c\sigma\sqrt{\frac{\log(1/\delta)}{n}}\right) \leq P(N \geq k/2)$$
$$\leq \exp(-2(k/2 - kp)^2/k)$$
$$\leq \exp(-k/8) =: \delta$$

Set $k = \lfloor 8\log(1/\delta) \rfloor$. Now subsituting $m = n/k$ gives the result.

$$P\left(|\widehat{\mu}_n - \mu| \geq 3\sigma\sqrt{\frac{\log(1/\delta)}{n}}\right) \leq P\left(|\widehat{\mu}_n - \mu| \geq .5\sigma\sqrt{\frac{\lfloor 8\log(1/\delta)\rfloor}{n}}\right) \leq \delta$$

$\square$