Stats300b Problem Set 5
Due: Thursday, February 15 at beginning of class

Exercises 6.1, 6.2, 7.1.

Answer to 6.1:

(a) We have \( \text{Var}(Y) = \inf_t \mathbb{E}[(Y - t)^2] \leq \mathbb{E}[(Y - \frac{b+a}{2})^2] \leq \max\{(b - \frac{b+a}{2})^2, (a - \frac{b+a}{2})^2\} = \frac{(b-a)^2}{4} \).

(b) Because we may pass derivatives through expectations, we have

\[
\varphi'(\lambda) = \frac{\mathbb{E}[X e^{\lambda X}]}{\mathbb{E}[e^{\lambda X}]} = \int x \frac{e^{\lambda x}}{\mathbb{E}[e^{\lambda X}]} dP(x) = \mathbb{E}_{Q_\lambda}[X].
\]

Taking second derivatives, we have

\[
\varphi''(X) = \frac{\mathbb{E}[X^2 e^{\lambda X}]}{\mathbb{E}[e^{\lambda X}]} - \frac{\mathbb{E}[X e^{\lambda X}]^2}{\mathbb{E}[e^{\lambda X}]^2} = \mathbb{E}_{Q_\lambda}[X^2] - \mathbb{E}_{Q_\lambda}[X]^2 = \text{Var}_{Q_\lambda}(X)
\]
as desired.

(c) We have \( \varphi'(0) = 0 \), and we use that \( X \) is bounded under all distributions \( Q_\lambda \) to find that \( \varphi''(\lambda) \leq \frac{(b-a)^2}{4} \). The integral equation

\[
\varphi(\lambda) = \varphi(0) + \varphi'(0) \lambda + \frac{1}{2} \varphi''(\tilde{\lambda}) \lambda^2
\]
for some \( \tilde{\lambda} \in [0, \lambda] \) by Taylor’s theorem coupled with \( \varphi(0) = \varphi'(0) = 0 \) then yields \( \varphi(\lambda) \leq \frac{(b-a)^2 \lambda^2}{8} \) as desired.

Answer to 6.2: The first inequality of the problem is known as the Marcinkiewicz-Zygmund inequality, and it is known to be sharp in the sense that there is another constant \( C'_k \) such that

\[
C'_k \mathbb{E} \left( \left( \sum_{i=1}^n X_i^2 \right)^{\frac{k}{2}} \right) \leq \mathbb{E}[|S_n|^k].
\]

Let us turn to the proof.

(a) Without loss of generality, let \( \sigma = 1 \). Let \( \varepsilon_i \in \{\pm 1\} \) be i.i.d. Rademacher variables and let \( S'_n = \sum_{i=1}^n X'_i \) for \( X'_i \) independent copies of \( X \). Then by a standard symmetrization argument, we have

\[
\mathbb{E}[|S_n|^k] = \mathbb{E}[|S_n - \mathbb{E}[S'_n]|^k] \leq \mathbb{E}[|S_n - S'_n|^k]
\]

\[
= \mathbb{E} \left[ \left( \sum_{i=1}^n \varepsilon_i (X_i - X'_i) \right)^k \right] = 2^k \mathbb{E} \left[ \left( \frac{1}{2} \sum_{i=1}^n \varepsilon_i X_i - \frac{1}{2} \sum_{i=1}^n \varepsilon_i X'_i \right)^k \right]
\]

\[
\leq 2^k \mathbb{E} \left[ \left( \sum_{i=1}^n \varepsilon_i X_i \right)^k \right],
\]

1
where the final inequality is a consequence of the convexity of $x \mapsto |x|^k$ for $k \geq 1$. Let $S_n^c = \sum_{i=1}^n \varepsilon_i X_i$ be the symmetrized version of $S_n$, and by conditioning on $X = (X_1, \ldots, X_n)$, which has $\ell_2$-norm $\|X\| = (\sum_{i=1}^n X_i^2)^{1/2}$, we have

$$2^{-k} \mathbb{E}[|S_n^c|^k] \leq \mathbb{E}[|S_n^c|^k|x] = \mathbb{E}\left[\mathbb{E}[|S_n^c|^k | X] \right] = \mathbb{E}\left[\int_0^\infty \mathbb{P}(|S_n^c|^k \geq t | X) dt \right],$$

where we have used that $\mathbb{E}[Z] = \int_0^\infty \mathbb{P}(Z \geq t) dt$ for all nonnegative random variables $Z$.

Conditional on $X$, we have that $S_n^c$ is $\|X\|^2$-sub-Gaussian, because $\varepsilon_i X_i \in [-X_i, X_i]$. Thus Hoeffding’s inequality implies

$$\mathbb{P}(|S_n^c| \geq t^{1/k} | X) \leq 2 \exp\left(-\frac{t^{2/k}}{2 \|X\|^2}\right).$$

For any constant $C$ we have by the change of variables $u = t^{2/k}/2C$, or $t = (2Cu)^{k/2}$ and

$$dt = k2^{\frac{k}{2}-1}C^\frac{1}{2}u^{\frac{k}{2}-1}du,$$

that

$$\int_0^\infty \exp\left(-\frac{t^{2/k}}{2C}\right) dt = k2^{\frac{k}{2}-1}C^\frac{1}{2} \int_0^\infty u^{\frac{k}{2}-1}e^{-u}du = k\Gamma(k/2)2^{\frac{k}{2}-1}C^\frac{1}{2}.$$

In particular, we find that

$$\mathbb{E}\left[\int_0^\infty \mathbb{P}(|S_n^c| \geq t^{1/k} | X) dt \right] \leq 2^k k\Gamma\left(\frac{k}{2}\right) \mathbb{E}\left[\sum_{i=1}^n X_i^2\right]^\frac{k}{2}.$$
Answer to 7.1.

(a) This is basically the same result we showed in class. We have $m_\theta(x, y)$ is Lipschitz in $\theta$, that is,

$$|m_\theta(x, y) - m_{\theta'}(x, y)| \leq |\langle \theta, x \rangle - \langle \theta', x \rangle| \leq \|\theta - \theta'\| \|x\|_s.$$ 

The covering numbers $N(\Theta, \| \cdot \|, \epsilon)$ are finite for all $\epsilon > 0$, so letting $\{\theta^i\}_{i=1}^N$ denote an $\epsilon$-cover of $\Theta$, we have that the pairs $[\ell_i, u_i]$ with $\ell_i(x, y) = m_{\theta^i}(x, y) - \epsilon \|x\|_s$ and $u_i(x, y) = m_{\theta^i}(x, y) + \epsilon \|x\|_s$ form a $2\epsilon$-bracketing of $\{m_\theta(\cdot, \cdot) \mid \theta \in \Theta\}$ in the $L_1(P)$ norm. Applying (for example) Hoeffding’s inequality, Theorem 19.4 gives the result.

(b) We have that

$$\log(1 + e^{-Mr}) \leq \log(1 + e^{-\|x\|_s \|\theta\|}) \leq \log(1 + e^{-yx^\top \theta})$$

so that $m_\theta(x, y) \leq \log(1 + e^{-\|x\|_s \|\theta\|}) \leq \log(1 + e^{Mr})$,

thus for any fixed $\theta$ and all $t \geq 0$, we have by Hoeffding’s inequality that

$$\mathbb{P}(|P_n m_\theta(X, Y) - M(\theta)| \geq t) \leq 2 \exp\left(-\frac{nt^2}{2M^2r^2}\right).$$

Now, using the result on volumes in class, we know that for any $\epsilon > 0$ the covering numbers of $\Theta = \{\theta \in \mathbb{R}^d : \|\theta\| \leq r\}$ satisfy $\log N(\Theta, \| \cdot \|, \epsilon) \leq d \log(1 + \frac{2r}{\epsilon})$. Let $\epsilon > 0$ and $\{\theta^i\}_{i=1}^N$, $N = N(\Theta, \| \cdot \|, \epsilon)$ be a minimal $\epsilon$-cover of $\Theta$. Then letting $\pi(\theta) = \arg\min_{i \leq N} \|\theta - \theta^i\|$, we have

$$\sup_{\theta \in \Theta} |P_n m_\theta - M(\theta)| \leq \sup_{\theta \in \Theta} |P_n m_\theta - P_n m_{\pi(\theta)}| + \max_i |P_n m_{\theta^i} - P m_{\theta^i}| + \sup_{\theta \in \Theta} |P m_{\pi(\theta)} - M(\theta)|$$

$$\leq 2M\epsilon + \max_{i \leq N} |P_n m_{\theta^i} - P m_{\theta^i}|,$$

where we have used that $m_\theta(x, y)$ is $\|x\|_s$-Lipschitz in $\theta$ and $\|x\|_s \leq M$. Thus we have for any $\epsilon > 0$ that for a minimal $\epsilon$-cover $\{\theta^i\}_{i=1}^N$ of $\Theta$, we have

$$\mathbb{P}(\sup_{\theta \in \Theta} |P_n m_\theta - P m_{\theta}| \geq 2M\epsilon + t) \leq \mathbb{P}\left(\max_{i \leq N(\epsilon)} |P_n m_{\theta^i} - P m_{\theta^i}| \geq t\right)$$

$$\leq N(\Theta, \| \cdot \|, \epsilon) \max_{i \leq N(\epsilon)} \mathbb{P}\left(|P_n m_{\theta^i} - P m_{\theta^i}| \geq t\right)$$

$$\leq 2 \exp\left(d \log\left(1 + \frac{2r}{\epsilon}\right) - \frac{nt^2}{2M^2r^2}\right),$$

where we have used a union bound, our covering number bound, and Hoeffding’s inequality (1).

If we choose

$$t^2 = \frac{2M^2r^2}{n} \left[\log \frac{2}{\delta} + d \log\left(1 + \frac{2r}{\epsilon}\right)\right],$$

we have

$$\mathbb{P}\left(\sup_{\theta \in \Theta} |P_n m_\theta - P m_{\theta}| \geq 2M\epsilon + t\right) \leq \delta,$$

and setting $\epsilon = r/n$ gives the desired result with

$$\epsilon_n(\delta) = \frac{2Mr}{n} + \sqrt{2Mr} \sqrt{\log \frac{2}{\delta} + 2 \log(1 + n)}.$$