Liam Hardiman *April 21, 2020*

270C - Homework 1

2.2.7 Let $X_1, \ldots, X_N$ be independent random variables. Assume that $X_i \in [m_i, M_i]$ for every i. Then for any $t > 0$ we have

$$\Pr\left[\sum_{i=1}^N (X_i - E[X_i]) \geq t\right] \leq \exp\left(-\frac{2t^2}{\sum_{i=1}^N (M_i - m_i)^2}\right).$$

Is there an easy argument I'm missing here? Let $Y_i = X_i - E[X_i]$. Like in all of these Hoeffding-like proofs, we multiply by λ, exponentiate, and apply Markov to obtain

$$\Pr\left[\sum_{i=1}^N Y_i \geq t\right] \leq e^{-\lambda t} \prod_{i=1}^N E\left[e^{\lambda Y_i}\right]. \tag{1}$$

Let's bound that MGF. [I Googled around for this and found Hoeffding's lemma.] Since $Y_i \in [m_i, M_i]$ and $t \geq 0$, the convexity of $t \mapsto e^{\lambda t}$ gives

$$e^{\lambda Y_i} \leq \frac{Y_i - m_i}{M_i - m_i} e^{\lambda M_i} + \frac{M_i - Y_i}{M_i - m_i} e^{\lambda m_i}.$$

Taking the expectation and using the fact that the Y_i's are centered gives

$$E e^{\lambda Y_i} \leq \frac{M_i}{M_i - m_i} e^{\lambda m_i} - \frac{m_i}{M_i - m_i} e^{\lambda M_i}.$$

Set $\alpha = \frac{M_i}{M_i - m_i}$. Taking the logarithm of the above expression gives

$$\log(\alpha e^{\lambda M} + (1-\alpha) e^{\lambda m}) = \lambda m + \log(\alpha + (1-\alpha) e^{\lambda(M-m)}).$$

Regard this expression as a function φ of $u = \lambda(M - m)$. Some (tedious) calculus shows that this function is zero at zero, has zero derivative at zero, and has bounded (say by $K > 0$ independent of M_i and m_i) second derivative. By Taylor's theorem, there is some $\xi \in (0, u)$ such that

$$\varphi(u) = \varphi(0) + \varphi'(0)u + \frac{\varphi''(\xi)}{2} u^2 \leq \frac{1}{2} K \lambda^2 (M_i - m_i)^2.$$

We then have that

$$E e^{\lambda Y_i} \leq e^{K\lambda^2 (M_i - m_i)^2/2}.$$

Substituting this bound into (1) gives

$$\Pr\left[\sum_{i=1}^N Y_i \geq t\right] \leq \exp\left(-\lambda t + \frac{K\lambda^2}{2} \sum_{i=1}^N (M_i - m_i)^2\right).$$

Choosing $\lambda = \frac{t}{K \sum_{i=1}^N (M_i - m_i)^2}$ minimizes the above expression, which gives the desired conclusion, with some absolute constant K in place of 2.

$\square$

2.2.8 Imagine we have an algorithm for solving some decision problem. Suppose the algorithm makes a decision at random and returns the correct answer with probability $\frac{1}{2} + \delta$¡ for some $\delta > 0$. To improve the performance, we run the algorithm N times and take the majority vote. Show that for any $\epsilon \in (0,1)$, the answer is correct with probability at least $1 - \epsilon$, as long as

$$N \geq \frac{1}{2\delta^2} \log \frac{1}{\epsilon}.$$

Proof. If $X_1, X_2, \ldots, X_N$ are the indicators of wrong outputs, then we're wrong exactly when $\sum_{i=1}^N X_i > \frac{1}{2}N$. The expectation of this sum is $(\frac{1}{2} - \delta)N$, so by the bounded version of Hoeffding proved in exercise 2.2.7 we have

$$\begin{aligned}
\Pr\left[\sum_{i=1}^N X_i > \frac{1}{2}N\right] &= \Pr\left[\sum_{i=1}^N X_i - \left(\frac{1}{2} - \delta\right) N > N\delta\right] \\
&\leq e^{-2(N\delta)^2/N} \\
&= e^{2N\delta^2}.
\end{aligned}$$

This quantity is less than ϵ when $N > \frac{1}{2\delta^2} \log \frac{1}{\epsilon}$. □

2.2.9 Suppose we want to estimate the mean μ of a random variable X (assumed to take values in $[a,b]$ a.s.) from a sample $X_1, \ldots, X_N$ drawn independently from the distribution of X. We want an ϵ-accurate estimate.

(a) Show that a sample of size $N = O(\sigma^2/\epsilon^2)$ is sufficient to compute an ϵ-accurate estimate with probability at least $3/4$, where $\sigma^2 = \mathrm{Var}[X]$.

Proof. The sample mean $\hat{\mu} = \frac{1}{N}\sum_{i=1}^N X_i$ has mean μ. By Hoeffding we have

$$\Pr[|\hat{\mu} - \mu| > \epsilon] \leq 2e^{-2N\epsilon^2/(b-a)^2}.$$

Setting this less than $1/4$ and solving for N gives

$$N \geq \frac{(b-a)^2 \log 8}{\epsilon^2} \geq \frac{4 \log 8\sigma^2}{\epsilon^2} = O(\sigma^2/\epsilon^2).$$

□

(b) Show that a sample of size $N = O(\log(1/\delta)\sigma^2/\epsilon^2)$ is sufficient to compute an ϵ-accurate estimate with probability at least $1 - \delta$.

Proof. Like in part (a) we have

$$\Pr[|m\hat{u} - \mu| > \epsilon] \leq 2e^{-2N\epsilon^2/(b-a)^2}.$$

Setting this less than δ and solving for N gives

$$N > \frac{1}{\epsilon^2}(b-a)^2 \log \frac{2}{\delta} \geq \frac{4\sigma^2}{\epsilon^2} \log \frac{2}{\delta} = O\left(\frac{\sigma^2}{\epsilon^2} \log \frac{2}{\delta}\right).$$

□

2.2.10 Let $X_1, \dots, X_N$ be non-negative independent random variables with continuous distributions. Assume that the densities of X_i are uniformly bounded by 1.

(a) Show that the MGF of X_i satisfies

$$E\exp(-tX_i) \le \frac{1}{t} \quad \text{for all } t > 0.$$

Proof. Let f_i be the density of X_i. Since $f_i(x) \le 1$ for all x, we have

$$E\exp(-tX_i) = \int_0^\infty e^{-tx} f_i(x)\, dx \le \int_0^\infty e^{-tx}\, dx = \frac{1}{t}.$$

□

(b) Deduce that for any $\epsilon > 0$ we have

$$\Pr\left[\sum_{i=1}^N X_i \le \epsilon N\right] \le (e\epsilon)^N.$$

Proof. We have

$$\begin{aligned}\Pr\left[\sum_{i=1}^N X_i \le \epsilon N\right] &= \Pr\left[-t\sum_{i=1}^N (X_i/\epsilon) \ge -tN\right] \\ &= \Pr\left[\exp\left(-t\sum_{i=1}^N (X_i/\epsilon)\right) \ge e^{-tN}\right].\end{aligned}$$

Applying Markov's inequality gives

$$\begin{aligned}\Pr\left[\sum_{i=1}^N X_i \le \epsilon N\right] &\le e^{tN} E\left[\exp\left(-t\sum_{i=1}^N (X_i/\epsilon)\right)\right] \\ &= e^{tN}\prod_{i=1}^N E\exp(-(t/\epsilon)X_i) \\ &\le \frac{e^{tN}\epsilon^N}{t^N}.\end{aligned}$$

The right-hand side attains a minimum at $t = 1$, which finally gives

$$\Pr\left[\sum_{i=1}^N X_i \le \epsilon N\right] \le (e\epsilon)^N.$$

□

2.3.2 Let X_i be independent Bernoulli random variables with parameters p_i. Consider the sum $S_N = \sum_{i=1}^N X_i$ and denote its mean by $\mu = E[S_N]$. Then, for any $t < \mu$, we have

$$\Pr[S_N \le t] \le e^{-\mu}\left(\frac{e\mu}{t}\right)^t.$$

Proof. We compute.

$$
\begin{aligned}
\Pr[S_n \leq t] &= \Pr[-S_n \geq -t] \\
&= \Pr[\exp(-\lambda S_n) \geq e^{-\lambda t}] \\
&\leq e^{\lambda t} \prod_{i=1}^{N} E[e^{-\lambda X_i}] \\
&= e^{\lambda t} \prod_{i=1}^{N} [e^{-\lambda} p_i + (1 - p_i)] \\
&\leq e^{\lambda t} \exp\left[(e^{-\lambda} - 1) \sum_{i=1}^{N} p_i\right] \\
&= e^{-\lambda t} \exp\left[\mu\left(\frac{t}{\mu} - 1\right)\right].
\end{aligned}
$$

This holds for all $\lambda \geq 0$. Since $t < \mu$, setting $\lambda = \log(\mu/t)$ gives

$$
\Pr[S_n \leq t] \leq \left(\frac{\mu}{t}\right)^t e^{t-\mu} = e^{-\mu}\left(\frac{e\mu}{t}\right)^t.
$$

□

2.3.5 Show that in the setting of the previous problem, for $\delta \in (0, 1]$ we have

$$
\Pr[|S_N - \mu| \geq \delta\mu] \leq 2e^{-c\mu\delta^2},
$$

where $c > 0$ is an absolute constant.

Proof. By the previous exercise we have

$$
\begin{aligned}
\Pr[|S_N - \mu| \geq \delta\mu] &= \Pr[S_N \geq (1+\delta)\mu] + \Pr[S_N \leq (1-\delta)\mu] \\
&\leq e^{-\mu}\left(\frac{e\mu}{(1+\delta)\mu}\right)^{(1+\delta)\mu} + e^{-\mu}\left(\frac{e\mu}{(1-\delta)\mu}\right)^{(1-\delta)\mu} \\
&= \left(\frac{e^{\delta}}{(1+\delta)^{1+\delta}}\right)^{\mu} + \left(\frac{e^{-\delta}}{(1-\delta)^{1-\delta}}\right)^{\mu}
\end{aligned}
$$

First, we claim that

$$
\left(\frac{e^{\delta}}{(1+\delta)^{1+\delta}}\right)^{\mu} \leq e^{-\delta^2\mu/3}. \tag{2}
$$

To see this, we take the logarithm of the left-hand side:

$$
\mu[\delta - (1+\delta)\log(1+\delta)].
$$

Now an elementary calculus argument shows that $\log(1+\delta) \geq \frac{2\delta}{2+\delta}$ for $\delta \in (0, 1]$ (the difference is zero for $\delta = 0$ and its derivative is nonnegative). From this we deduce

$$
\mu[\delta - (1+\delta)\log(1+\delta)] \leq \mu\left[\delta - \frac{2\delta(1+\delta)}{2+\delta}\right] = -\frac{\mu\delta^2}{2+\delta},
$$

which establishes (2) after considering $\delta \in (0,1]$. Next, we claim that

$$\left(\frac{e^{-\delta}}{(1-\delta)^{1-\delta}}\right)^{\mu} \leq e^{-\delta^2\mu/2}. \tag{3}$$

Like in the preceding inequality, we take the logarithm of the left-hand side:

$$\mu[-\delta - (1-\delta)\log(1-\delta)].$$

By looking at the Taylor expansion, we see that $(1-\delta)\log(1-\delta) \geq -\delta + \frac{\delta^2}{2}$. Substituting this into the above quantity establishes (3). Inequalities (2) and (3) establish the desired claim. □

2.3.6 Let $X \sim Pois(\lambda)$. Show that for $t \in (0,\lambda]$ we have

$$\Pr[|X-\lambda| \geq t] \leq 2\exp\left(-\frac{ct^2}{\lambda}\right).$$

Proof. Let $X_{N,i}$ be a sequence of independent Bernoulli random variables with parameters $p_{N,i}$. Let $S_N = \sum_{i=1}^N X_{N,i}$. Furthermore, suppose that as $N \to \infty$

$$\max_{i \leq N} p_{N,i} \to 0, \quad \text{and} \quad E[S_N] \to \lambda.$$

Then by the Poisson limit theorem, $S_N \to X$ in distribution. Setting $\mu_N = E[S_N]$, we have

$$\begin{aligned}\Pr[|X-\lambda| \geq t] &= \lim_{N\to\infty} \Pr[|S_N - \mu_N| \geq t] \\ &\leq \lim_{N\to\infty} 2e^{-ct^2/\mu_N} \\ &= 2e^{-ct^2/\lambda}.\end{aligned}$$

□

2.4.2 Consider a random graph $G \sim G(n,p)$ with expected degree $d = O(\log n)$. Show that with high probability, all vertices have degrees $O(\log n)$.

Proof. Fix $c > 0$ to be determined later and fix $\epsilon > 0$. Chernoff tells us that for any vertex i we have

$$\Pr[d_i \geq cd] \leq e^{-d}\left(\frac{ed}{cd}\right)^{cd} = \left(\frac{e^{c-1}}{c^c}\right)^d.$$

A union bound over all vertices gives

$$\Pr[d_i \geq c_d \text{ for some } i \leq n] \leq n\left(\frac{e^{c-1}}{c^c}\right).$$

We take the logarithm and do some algebra.

$$\begin{aligned} n\left(\frac{e^{c-1}}{c^c}\right)^d < \epsilon &\iff \log n + d\log\frac{e^{c-1}}{c^c} < \log\epsilon \\ &\iff d < \frac{\log(n/\epsilon)}{\log(c^c/e^{c-1})}.\end{aligned}$$

For c sufficiently large (say 10), we have that $d = O(\log n)$. We have then shown that when $d = O(\log n)$, the probability that any vertex has degree larger than cd is less than ϵ. Consequently, with probability $1 - \epsilon$, each degree is $O(d) = O(\log n)$. $\square$

2.4.3 Consider a random graph $G \sim G(n, p)$ with expected degree $d = O(1)$. Show that with high probability, say 0.9, all vertices of G have degrees $O(\frac{\log n}{\log \log n})$.

Proof. Fix $c > 0$ to be determined later and fix $\epsilon > 0$. For ease of notation, let $f(n) = \frac{\log n}{\log \log n}$. Since $d = O(1)$, we can say $d_i \leq M$ for some fixed large M and for all i. Like in the previous exercise, Chernoff and a union bound give us

$$\Pr[d_i > cf(n) \text{ for some } i \leq n] \leq ne^{-d} \left(\frac{ed}{cf(n)} \right)^{cf(n)} \leq n \left(\frac{eM}{cf(n)} \right)^{cf(n)}.$$

Setting this less than ϵ and taking logarithms gives

$$n \left(\frac{eM}{cf(n)} \right)^{cf(n)} < \epsilon \iff \log n + cf(n)[1 + \log M - \log(cf(n))] < \log \epsilon$$

$$\iff cf(n)[\log(cf(n)) - 1 - \log M] > \log \frac{n}{\epsilon}.$$

Now the leading term of the left-hand side of the final inequality is

$$cf(n) \log(cf(n)) = c \frac{\log n}{\log \log n} \left(\log(c \log n) - \log \log \log n \right),$$

which is $\Omega(\log n)$ for an appropriate choice of c. For such c, we have that $\Pr[d_i > cf(n) \text{ for some } i \leq n] < \epsilon$, so with probability at least $1 - \epsilon$, $d_i = O(\frac{\log n}{\log \log n})$. $\square$

2.4.4 Consider a random graph $G \sim G(n, p)$ with expected degree $d = o(\log n)$. Show that with high probability, say 0.9, G has a vertex with degree $10d$.

Proof. Since $d = o(\log n)$, we must have $p = o(\frac{\log n}{n})$. If the degrees d_i were independent, we could write

$$\Pr[d_i \neq 10d \text{ for all } i \leq n] = \prod_{i=1}^{n} \Pr[d_i \neq 10d],$$

and then use Chernoff to bound the right-hand side. Since life isn't so simple, we have to be a bit more clever.

Maybe we can take an arbitrary subset and consider the outdegrees of its vertices: these will be independent. $\square$

2.5.5

(a) Show that if $X \sim \mathcal{N}(0, 1)$, the function $\lambda \mapsto E \exp(\lambda^2 X^2)$ is only finite in some bounded neighborhood of zero.

Proof. Let's compute that expectation.

$$E\exp(\lambda^2 X^2) = \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} e^{(\lambda^2 - 1/2)x^2}\ dx.$$

This integral is finite if and only if $\lambda \in (-1/\sqrt{2}, 1/\sqrt{2})$. $\square$

(b) Suppose that some random variable X satisfies $E\exp(\lambda^2 X^2) \le \exp(K\lambda^2)$ for all $\lambda \in \mathbb{R}$ and some constant K. Show that X is a bounded random variable, i.e. $\|X\|_\infty < \infty$.

Proof. Suppose X were not bounded. Then $\Pr[|X| \ge M] > 0$ for every M. We can bound that expectation below:

$$E\exp(\lambda^2 X^2) \ge E[\exp(\lambda^2 X^2) \cdot \mathbb{1}_{\{X \ge M\}}] \ge \exp(\lambda^2 M^2) \cdot \Pr[|X| \ge M].$$

$\square$

2.5.7 Define the norm $\|\cdot\|_{\psi_2}$ on the space of sub-gaussian random variables:

$$\|X\|_{\psi_2} = \inf\{t > 0 : E\exp(X^2/t^2) \le 2\}.$$

Show that $\|\cdot\|_{\psi_2}$ is indeed a norm.

Proof. Clearly $\|0\|_{\psi_2} = 0$. Conversely, if $\|X\|_\infty > 0$, then for some $\epsilon > 0$ and $\delta > 0$, $\Pr[|X| > \epsilon] > \delta$. We then have

$$E\exp(X^2/t^2) \ge \Pr[|X| > \epsilon]\exp(\epsilon^2/t^2) \ge \delta\exp(\epsilon^2/t^2).$$

This quantity can be made arbitrarily large for t arbitrarily small, so it cannot be less than 2 t arbitrarily small. We conclude that $\|X\|_{\psi_2} = 0$ if and only if $X = 0$.

Let c be any real number and suppose $\|X\|_{\psi_2} = r$. Then

$$\|cX\|_{\psi_2} = \inf\{t > 0 : E\exp(c^2X^2/t^2) \le 2\} = \inf\{t > 0 : E\exp(X^2/(t/c)^2) \le 2\} = |c|r,$$

so $\|\cdot\|_{\psi_2}$ is homogeneous.

For ease of notation, let $f(x) = e^{x^2}$. It suffices to show that for any two sub-gaussian random variables X and Y,

$$Ef\left(\frac{|X+Y|}{\|X\|_{\psi_2} + \|Y\|_{\psi_2}}\right) \le 2. \tag{4}$$

Since f is convex and increasing, we have for any s and t,

$$f\left(\frac{|X+Y|}{s+t}\right) \le f\left(\frac{|X|+|Y|}{s+t}\right) \le \frac{s}{s+t} f\left(\frac{|X|}{s}\right) + \frac{t}{s+t} f\left(\frac{|Y|}{t}\right).$$

Taking the expectation of both sides and setting $s = \|X\|_{\psi_2}$ and $t = \|Y\|_{\psi_2}$ gives

$$Ef\left(\frac{|X+Y|}{\|X\|_{\psi_2} + \|Y\|_{\psi_2}}\right) \le \frac{\|X\|_{\psi_2}}{\|X\|_{\psi_2} + \|Y\|_{\psi_2}} \cdot 2 + \frac{\|Y\|_{\psi_2}}{\|X\|_{\psi_2} + \|Y\|_{\psi_2}} \cdot 2 = 2.$$

Thus, $\|\cdot\|_{\psi_2}$ is a norm. $\square$

2.5.10 Let $X_1, X_2, \ldots$ be a sequence of sub-gaussian random variables, which are not necessarily independent. Show that

$$E \max_i \frac{|X_i|}{\sqrt{1+\log i}} \leq CK,$$

where $K = \max_i \|X_i\|_{\psi_2}$. Deduce that for every $N \geq 2$ we have

$$E \max_{i \leq N} |X_i| \leq CK\sqrt{\log N}.$$

Proof. We use the fact that for a nonnegative random variable Y, $E[Y] = \int_0^\infty \Pr[Y \geq t]\ dt$. We split this integral

$$\begin{aligned} E \max_{i\leq N} \frac{|X_i|}{\sqrt{1+\log i}} &= \int_0^\infty \Pr\left[\max_{i\leq N} \frac{|X_i|}{\sqrt{1+\log i}} \geq t\right]\ dt \\ &= \int_0^\alpha \Pr\left[\max_{i\leq N} \frac{|X_i|}{\sqrt{1+\log i}} \geq t\right]\ dt + \int_\alpha^\infty \Pr\left[\max_{i\leq N} \frac{|X_i|}{\sqrt{1+\log i}} \geq t\right]\ dt. \end{aligned}$$

The first integral is clearly bounded by α. For the second, we use a union bound.

$$\begin{aligned} E \max_{i\leq N} \frac{|X_i|}{\sqrt{1+\log i}} &\leq \alpha + \int_\alpha^\infty \sum_{i=1}^N \Pr\left[\frac{|X_i|}{\sqrt{1+\log i}} \geq t\right]\ dt \\ &= \alpha + \sum_{i=1}^N \int_\alpha^\infty \Pr[|X_i| \geq t\sqrt{1+\log i}]\ dt. \end{aligned}$$

Since the X_i's are sub-Gaussian, we can bound these tail probabilities.

$$\begin{aligned} E \max_{i\leq N} \frac{|X_i|}{\sqrt{1+\log i}} &\leq \alpha + \sum_{i=1}^N \int_\alpha^\infty 2\exp\left[-t^2(1+\log i)/K^2\right]\ dt \\ &= \alpha + \sum_{i=1}^N \int_\alpha^\infty 2\exp\left[-t^2/K^2\right] i^{-t^2}\ dt \\ &\leq \alpha + \sum_{i=1}^N \int_\alpha^\infty 2\exp\left[-t^2/K^2\right] i^{-\alpha^2}\ dt \\ &\leq \alpha + CK\sum_{i=1}^\infty i^{-\alpha^2}. \end{aligned}$$

Setting $\alpha = 2$ makes the sum converge and establishes the first claim.

□

2.5.11 Show that the bound exercise 2.5.10 is sharp. let $X_1, \ldots, X_N$ be independent $\mathcal{N}(0,1)$ random variables. Prove that

$$E \max_{i\leq N} X_i \geq c\sqrt{\log N}.$$

Proof. We have

$$E[\max_i X_i] = E[\max_i X_i \cdot \mathbb{1}_{\max_i X_i \geq 0}] + E[\max_i X_i \cdot \mathbb{1}_{\max_i X_i < 0}].$$

The probability that $\max_i X_i$ is negative is 2^{-N} and $\max_i X_i$ is monotone increasing in N, so by dominated convergence, the second term is $o(1)$. As for the first term, by independence we have

$$\begin{aligned}
E[\max_i X_i \cdot \mathbb{1}_{\max_i X_i \geq 0}] &= \int_0^\infty \Pr[X_i \cdot \mathbb{1}_{\max_i X_i \geq 0} \geq t]\ dt \\
&= \int_0^\infty 1 - \Pr[X_i < t,\ i \leq N]\ dt \\
&= \int_0^\infty 1 - \Phi(t)^N\ dt \\
&\geq \int_0^{\sqrt{c \log N}} 1 - \Phi(t)^N\ dt.
\end{aligned}$$

where $\Phi(t)$ is the CDF of the standard normal. I want an *upper* bound for $\Phi(t)$ here, but our usual tail bounds only give us *lower* bounds on it. Maybe a Mill's ratio thing? □

2.6.5 Let $X_1, \ldots, X_N$ be in independent sub-Gaussian random variables with zero means and unit variances, and let $a = (a_1, \ldots, a_N) \in \mathbb{R}^N$. Prove that for every $p \in [2, \infty)$ we have

$$\left(\sum_{i=1}^N a_i^2\right)^{1/2} \leq \Big\|\sum_{i=1}^N a_i X_i\Big\|_{L^p} \leq CK\sqrt{p}\left(\sum_{i=1}^N a_i^2\right)^{1/2}$$

where $K = \max_i \|X_i\|_{\psi_2}$ and C is an absolute constant.

Proof. First the lower bound. Since the X_i are independent, centered, and have unit variance, we have

$$\begin{aligned}
\left(\sum_{i=1}^N a_i^2\right)^{1/2} &= \left(E\left[\left(\sum_{i=1}^N a_i X_i\right)^2\right]\right)^{1/2} \\
&= \Big\|\sum_{i=1}^N a_i X_i\Big\|_{L^2} \\
&\leq \Big\|\sum_{i=1}^N a_i X_i\Big\|_{L^p},
\end{aligned}$$

for all $p \geq 2$. As for the upper bound, we've established that the sum of sub-Gaussians is sub-Gaussian, so $\sum_{i=1}^N a_i X_i$ is sub-Gaussian. By independence we also have

$$\begin{aligned}
\Big\|\sum_{i=1}^N a_i X_i\Big\|_{L^p} &\le C_1\sqrt{p}\cdot \|\sum_{i=1}^N a_i X_i\|_{\psi_2} \\
&\le C_2\sqrt{p}\left(\sum_{i=1}^N \|a_i X_i\|_{\psi_2}^2\right)^{1/2} \\
&\le C_2 K\sqrt{p}\left(\sum_{i=1}^N a_i^2\right)^{1/2}.
\end{aligned}$$

□

2.6.6 Show that in the setting of exercise 2.6.5, we have

$$c(K)\left(\sum_{i=1}^N a_i^2\right)^{1/2} \le \Big\|\sum_{i=1}^N a_i X_i\Big\|_{L^1} \le \left(\sum_{i=1}^N a_i^2\right)^{1/2}.$$

Here $K = \max_i \|X_i\|_{\psi_2}$ and $c(K) > 0$ is a quantity which may depend only on K.

Proof. First the lower bound. Let $Z = \sum_i a_i X_i$. First we claim that $\|Z\|_2 \le \|Z\|_1^{1/4}\|Z\|_3^{3/4}$. To see this, let $a, b > 0$ be such that $a + b = 1$ and let $p, q > 0$ be such that $\frac{1}{p} + \frac{1}{q} = 1$. We apply Hölder's inequality.

$$\begin{aligned}
\big\|Z\big\|_{L^2} &= \big\|Z^{a+b}\big\|_{L^2} \\
&\le \left(\big\|Z^{2a}\big\|_{L^p}\big\|Z^{2b}\big\|_{L^q}\right)^{1/2} \\
&= \big\|Z\big\|_{L^{2ap}}^a\big\|Z\big\|_{L^{2bq}}^b
\end{aligned}$$

Setting $a = 1/4$, $b = 3/4$, $p = q = 2$ establishes the claim. By independence, $\big\|Z\big\|_{L^2}^2 = \sum_i a_i^2$. Furthermore, by the previous exercise we have

$$\begin{aligned}
\left(\sum_{i=1}^N a_i^2\right)^{1/2} &\le \Big\|\sum_{i=1}^N a_i X_i\Big\|_{L^1}^{1/4}\cdot\Big\|\sum_{i=1}^N a_i X_i\Big\|_{L^3}^{3/4} \\
&\le \Big\|\sum_{i=1}^N a_i X_i\Big\|_{L^1}^{1/4}\cdot (CK\sqrt{3})^{3/4}\left(\sum_{i=1}^N a_i^2\right)^{3/8}.
\end{aligned}$$

Dividing through by the sum and raising each term to the power 4 establishes the desired lower bound.

$$(CK\sqrt{3})^{-3}\left(\sum_{i=1}^N a_i^2\right)^{1/2} \le \Big\|\sum_{i=1}^N a_i X_i\Big\|_{L^1}.$$

The upper bound trivially follows by Cauchy-Schwarz: $\big\|Z\big\|_{L^1} \le \Pr[\Omega]^{1/2}\cdot\big\|Z\big\|_{L^2} = \big\|Z\big\|_{L^2}$ for all measurable functions $Z : \Omega \to \mathbb{R}$. □

2.6.7 State and prove a version of Khintchine's inequality for $p \in (0, 2)$.

Solution. Set $Z = \sum_i a_i X_i$. The same argument used in 2.6.5 gives the upper bound for $p \in [1, 2)$. For $0 < p < 1$, we use Jensen's inequality and Cauchy-Schwarz:

$$\|Z\|_{L^p} = E[Z^p]^{1/p} \leq E[|Z|] \leq \Pr[\Omega] \cdot \|Z\|_{L^2} = \|Z\|_{L^2}.$$

We generalize the argument from the last exercise to get the lower bound. Let $a, b > 0$ be such that $a + b = 1$ and $q, r > 1$ be such that $\frac{1}{q} + \frac{1}{r} = 1$. Note that $\|Z\|_{L^2} = (\sum_i a_i^2)^{1/2}$ by independence. By the same Hölder inequality argument used in the previous exercise, we have

$$\|Z\|_{L^2} \leq \|Z\|_{L^{2aq}}^a \|Z\|_{L^{2br}}^b.$$

Applying Khintchine from 2.6.5 gives

$$\|Z\|_{L^2} \leq \|Z\|_{L^{2aq}}^a \cdot \left(CK\sqrt{2br}\right)^b \|Z\|_{L^2}^b.$$

Rearranging and using the fact that $1 - b = a$ gives

$$\left(CK\sqrt{2br}\right)^{-b} \|Z\|_{L^2}^a \leq \|Z\|_{L^{2aq}}^a \implies \left(CK\sqrt{2br}\right)^{-b/a} \|Z\|_{L^2} \leq \|Z\|_{L^{2aq}}.$$

Applying the relations $a + b = 1$ and $\frac{1}{q} + \frac{1}{r} = 1$ gives

$$\left(CK\sqrt{2(1-a)\frac{q}{q-1}}\right)^{1-1/a} \|Z\|_{L^2} \leq \|Z\|_{L^{2aq}}.$$

We need $2aq = p$, which gives

$$\left(CK\sqrt{2(1-a)\frac{p}{p-2a}}\right)^{1-1/a} \|Z\|_{L^2} \leq \|Z\|_{L^{2p}}.$$

This holds for all $a \in (0, 1)$ such that $p - 2a > 0$, so simply (or not so simply) choose such an a that maximizes the constant on the left. $\square$

2.8.5 Let X be a mean-zero random variable such that $|X| \leq K$. Prove the following bound on the MGF of X:

$$E\exp(\lambda X) \leq \exp(g(\lambda)E[X^2]) \quad \text{where} \quad g(\lambda) = \frac{\lambda^2/2}{1 - |\lambda|K/3},$$

provided that $|\lambda| < 3/K$.

Proof. First, we claim that for $|z| < 3$ we have the inequality

$$e^z \leq 1 + z + \frac{z^2/2}{1 - |z|/3}.$$

To see this, we look at the Taylor series for e^z.

$$\begin{aligned} e^z - 1 + z &= \sum_{n=2}^{\infty} \frac{z^n}{n!} \\ &= \frac{z^2}{2} \cdot \sum_{n=0}^{\infty} \frac{2}{(n+2)!} z^n. \end{aligned}$$

Now $\frac{2}{(n+2)!} \leq \frac{1}{3^n}$ for all $n \geq 0$, so we can bound the above sum by a geometric series for $|z| < 3$:

$$e^z - 1 + z \leq \frac{z^2}{2} \cdot \sum_{n=0}^{\infty} (z/3)^n = \frac{z^2/2}{1 - |z|/3}.$$

Now we substitute λX in for z and use the fact that X is centered and that $|\lambda| < 3/K$.

$$\begin{aligned} E[\exp(\lambda X)] &\leq E\left[1 + X + \frac{(\lambda X)^2/2}{1 - |\lambda X|/3}\right] \\ &\leq 1 + g(\lambda)E[X^2] \\ &\leq e^{g(\lambda)E[X^2]}. \end{aligned}$$

□

2.8.6 Use the result of exercise 2.8.5 to prove the following theorem. Let $X_1, \ldots, X_N$ be independent, mean zero random variables, such that $|X_i| \leq K$ for all i. Then, for every $t \geq 0$, we have

$$\Pr\left[\left|\sum_{i=1}^{N} X_i\right| \geq t\right] \leq 2\exp\left(-\frac{t^2/2}{\sigma^2 + Kt/3}\right),$$

where $\sigma^2 = \sum_{i=1}^{N} E[X_i^2]$.

Proof. We do the usual multiplying by λ and applying Markov trick. Applying the previous exercise gives

$$\Pr\left[\sum_{i=1}^{N} X_i \geq t\right] \leq \exp\left(\sigma^2 g(\lambda) - \lambda t\right).$$

Now we minimize that exponent.

$$\sigma^2 g(\lambda) - \lambda t \leq \frac{\sigma^2}{2}\lambda^2 - t\lambda.$$

This quadratic is minimized at $\lambda = t/\sigma^2$ and substituting this in gives the desired bound. □