Exercise 1 : Uniform vs pointwise bounds

(a) Let $i \in [m]$, $t \geq 0$

$$\lim_{t\to\infty} \mathbb{P}\left[|Z_i| \leq t\right]$$

$$= \lim_{t\to\infty} \int_{\mathbb{R}} f_{|Z_i|} \, d\mu(x)$$

$$= \lim_{t\to\infty} \sqrt{\frac{2}{\pi}} \int_{\mathbb{R}} \exp\left(\frac{-x^2}{2}\right) \mathbb{1}_{[0,\infty)}(x) \, \mathbb{1}_{[-\infty,t]}(x) \, d\mu(x)$$

$$= \lim_{t\to\infty} \sqrt{\frac{2}{\pi}} \int_0^\infty \exp\left(\frac{-x^2}{2}\right) \mathbb{1}_{[-\infty,t]}(x) \, d\mu(x)$$

$$= \lim_{t\to\infty} \sqrt{\frac{2}{\pi}} \int_0^t \exp\left(\frac{-x^2}{2}\right) d\mu(x)$$

$$= \lim_{t\to\infty} \sqrt{\frac{2}{\pi}} \sqrt{2\pi} \left(\Phi(t) - \frac{1}{2} \right) \qquad t \mapsto \Phi(t) \text{ continuous}$$

$$= \sqrt{\frac{2}{\pi}} \left(1 - \frac{1}{2}\right) \sqrt{2\pi} = 1$$

(b) $\lim_{m\to\infty} \mathbb{P}\left[\sup_{i\in[m]} |Z_i| \leq t\right]$ (take $t \geq 0$, as result is trivial for $t<0$)

$$= \lim_{m\to\infty} \mathbb{P}\left[\bigcap_{i\in[m]} |Z_i| \leq t\right]$$

$$= \lim_{m\to\infty} \prod_{i\in[m]} \mathbb{P}\left[|Z_i| \leq t\right]$$

$$= \lim_{m\to\infty} \prod_{i\in[m]} \sqrt{\frac{2}{\pi}} \int_0^t \exp\left\{-\frac{x^2}{2}\right\} dx$$

$$= \lim_{m\to\infty} \left(\frac{2}{\pi}\right)^{m/2} \left(\int_0^t \exp\left\{\frac{-x^2}{2\sigma^2}\right\} dx \right)^m$$

$$= \lim_{m\to\infty} \left(\frac{2}{\pi}\right)^{m/2} \left[(\phi(t) - \tfrac{1}{2}) \sqrt{2\pi} \right]^m$$

$$= \lim_{m\to\infty} \left(\frac{2}{\pi}\right)^{m/2} \left[\phi(t)\sqrt{2\pi} - \sqrt{\frac{\pi}{2}} \right]^m$$

$$= \lim_{m\to\infty} \left[\phi(t) \left(\frac{2}{\pi}\right)^{1/2} \sqrt{2\pi} - 1 \right]^m$$

$$= \lim_{m\to\infty} \left[2\phi(t) - 1 \right]^m \qquad 0 < 2\phi(t) < 2 \Rightarrow \quad -1 < 2\phi(t) - 1 < 1$$

$= 0$. as $t \mapsto \phi(t)$ continuous

(c) Define $F_t := \mathbb{Z} \to \mathbb{R}$ by $F_n(i) = \dfrac{Z_i}{\log n}$ $\forall i \in \mathbb{Z}, n \in \mathbb{N}$

$$\forall \varepsilon > 0, \lim_{n\to\infty} \mathbb{P}\left[|F_n(i)| > \varepsilon \right] = \lim_{n\to\infty} \mathbb{P}\left[|Z_i| > \varepsilon \log n \right]$$
$$= 1 - \lim \mathbb{P}\left[|Z_i| \leq \varepsilon \log n \right]$$
$$= 0 \text{ by (a)}$$

Hence $F_n(i) \xrightarrow{P} 0 \;\forall\, i \in \mathbb{Z}$, pointwise

But $\mathbb{P}\left[\sup_{i\in[n]} |F_i(j)| > \varepsilon \right] = 1 - \mathbb{P}\left[\sup_{i\in[n]} |F_i(j)| \leq \varepsilon \right]$

$$= 1 - \mathbb{P}\left[|Z_i| \leq \varepsilon \log n \right]^n \quad (\text{iid})$$
$$\leq 1 - \left(1 - \frac{1}{(\varepsilon \log n)^2}\right)^n \quad (\text{Chebyshev})$$
$$\to 1 \text{ as } n \to \infty. \quad \text{since } \frac{1}{(\varepsilon \log n)^2} \to$$

F_n don't converge to 0 in probability uniformly, so they do not converge at all uniformly since uniform convergence $\Rightarrow$ pointwise convergence.

2. Subgaussian parameter for bounded random variable.

(a) WTS: $\mathrm{Var}(X) \le \frac{(b-a)^2}{4}$. Let $\mu := \mathbb{E}X$

We first note the variance of X is finite.

$\mathbb{P}[a \le X \le b] = 1 \Rightarrow \mathbb{E}[X^2] := \sigma^2 < \infty$.

We have

$$0 \le \mathbb{E}[(b-X)(X-a)] = -E[X^2] + aEX + bEX - ab$$
$$= -EX^2 + EX(a+b) - ab$$

Hence $$\sigma^2 = \mathbb{E}X^2 - \mu^2 \le -ab + \mu(a+b) - \mu^2$$
$$= (b-\mu)(\mu-a)$$
$$\le \frac{1}{4}\left((b-\mu) + (\mu-a)\right)^2 \quad \text{using } ab \le \left(\frac{a+b}{2}\right)^2,$$
$$= \frac{1}{4}(b-a)^2$$

An alternative argument is shown in part (c).

(b) WTS: $k'(t) = \mathbb{E}_t[X]$, $k''(t) = \mathrm{Var}_t[X]$

Wob $\mathrm{supp}(X) \subseteq [a,b]$ so we can swap order of expecta° and differentia°

$$\frac{d}{dt} k(t) = \frac{1}{\mathbb{E}[e^{tX}]} \frac{d}{dt} \mathbb{E}[e^{tX}] = \frac{1}{\mathbb{E}[e^{tX}]} \frac{d}{dt} \int_{\mathbb{R}} e^{tx} \, d\mathbb{P}_X(x) = \frac{1}{\mathbb{E}[e^{tX}]} \int_{\mathbb{R}} x e^{tx} \, d\mathbb{P}_X(x)$$

$$\mathbb{E}_t[X] = \int_{\mathbb{R}} x \frac{\mathbb{E}[\mathbb{1}\{X=x\} e^{tX}]}{\mathbb{E}[e^{tX}]} \, d\mu(x) = \frac{1}{\mathbb{E}[e^{tX}]} \int_{\mathbb{R}} x \, \mathbb{E}[\mathbb{1}\{X=x\} e^{tX}] \, d\mu(x)$$
$$= \frac{1}{\mathbb{E}[e^{tX}]} \int_{\mathbb{R}} x e^{tx} \, d\mathbb{P}(x)$$

Hence $\frac{d}{dt} k(t) = \mathbb{E}_t [X]$

$$\frac{d^2(k(t))}{dt^2} = \frac{d}{dt} E_t [X]$$

$$= \frac{d}{dt} \frac{1}{\mathbb{E}[e^{tX}]} \int_{\mathbb{R}} x e^{tx} d\mathbb{P}(x) =$$

(1) $$= \frac{1}{(\mathbb{E}_{\mathbb{P}}[e^{tX}])^2} \left(\mathbb{E}_{\mathbb{P}}[xe^{tx}] \mathbb{E}_{\mathbb{P}}[xe^{tx}] + \underbrace{\mathbb{E}_{\mathbb{P}}[e^{tX}] \frac{d}{dt} \mathbb{E}_{\mathbb{P}}[xe^{tx}]}_{\mathbb{E}_{\mathbb{P}}[e^{tX}](\mathbb{E}_{\mathbb{P}}[x^2 e^{tx}])} \right)$$

$$\mathrm{Var}_t (X) = \mathbb{E}_t [(X - E_t [X])^2] = \frac{1}{\mathbb{E} e^{tX}} \int_{\mathbb{R}} (x - E_t [X])^2 \ \mathbb{E}[\mathbb{1}\{X = x\} e^{tX}] \, d\mu(x)$$

$$= \frac{1}{\mathbb{E}[e^{tX}]} \int_{\mathbb{R}} (x - E_t [X])^2 e^{tx} \, d\mathbb{P}(x)$$

$$= \frac{1}{\mathbb{E}[e^{tX}]} (E_{\mathbb{P}} [X^2 e^{tX}] - 2\mathbb{E}_{\mathbb{P}} [x E_t [X] e^{tx}] + (\mathbb{E}_t [X])^2$$

$$= \frac{1}{\mathbb{E}[e^{tX}]} (\mathbb{E}_{\mathbb{P}} [X^2 e^{tX}] - 2 E_t [X] E_{\mathbb{P}} [x e^{tx}] - E_t [X]^2 \mathbb{E}_{\mathbb{P}} [e^{tX}]$$

(2) $$= \frac{1}{(\mathbb{E}[e^{tX}])^2} (E_{\mathbb{P}} [x^2 e^{tX}] \mathbb{E}_{\mathbb{P}} [e^{tX}] - (\mathbb{E}_t [X])^2$$

Note $E_t [X] = \frac{1}{\mathbb{E}[e^{tx}]} \mathbb{E}_{\mathbb{P}} [x e^{tx}]$, we have $E_t [X] E_{\mathbb{P}} [xe^{tx}] \frac{1}{E[e^{tX}]} = (E_t [X])^2 \mathbb{E}_{\mathbb{P}} [e^{tX}]$

(1) = (2) so the result follows

(c). We rewrite:

$$\mathrm{Var}_t(X) = \int_a^b (x - \mathbb{E}_t[x])^2 \, \mathbb{P}_t[x] \, d\mu(x), \quad \mathbb{P}_t(x) = \frac{\mathbb{E}[\mathbb{1}_{\{X=x\}} e^{tX}]}{\mathbb{E}[e^{tX}]} = \frac{e^{tx}\,\mathbb{P}(X=x)}{\mathbb{E}[e^{tX}]}$$

But for $x \in [a,b]$, $\mathbb{E}_t[x] = \displaystyle\int_a^b x \, \frac{\mathbb{E}[\mathbb{1}_{\{X=x\}} e^{tX}]}{\mathbb{E}[e^{tX}]} \, d\mu(x)$

$$= \int_a^b x \, \frac{e^{tx}}{\mathbb{E}[e^{tX}]} \, d\mathbb{P}_X(x)$$

$$= \frac{1}{\mathbb{E}[e^{tX}]} \mathbb{E}[x e^{tX}].$$

$$\mathrm{Var}_t(X) = \frac{1}{\mathbb{E}[e^{tX}]} \int_a^b \left(x - \frac{\mathbb{E}[Xe^{tX}]}{\mathbb{E}[e^{tX}]}\right)^2 e^{tx} \, d\mathbb{P}_X(x)$$

$$= \frac{1}{\mathbb{E}[e^{tX}]} \left(\int_a^b x^2 e^{tx} \, d\mathbb{P}_X - 2 \frac{x e^{tx} \mathbb{E}[Xe^{tX}]}{\mathbb{E}[e^{tX}]} + \left(\frac{\mathbb{E}[Xe^{tX}]}{\mathbb{E}[e^{tX}]}\right)^2 \right) d\mathbb{P}_X(x)$$

$$= \frac{1}{\mathbb{E}[e^{tX}]} \left(\int_a^b x^2 e^{tx} \, d\mathbb{P}_X \right) - \left(\frac{\mathbb{E}[Xe^{tX}]}{\mathbb{E}[e^{tX}]} \right)^2$$

$$= \frac{1}{\mathbb{E}[e^{tX}]} \left(\mathbb{E}[X^2 e^{tX}] - \frac{\mathbb{E}[Xe^{tX}]^2}{\mathbb{E}[e^{tX}]} \right)$$

This is the variance under the law of X reweighted by $\dfrac{e^{tX}}{\mathbb{E}[e^{tX}]}$

From (a), we know the bound on variance does not depend on the choice of measure. In particular,

$$\mathrm{Var}_t[X] = \inf_{\alpha} \mathbb{E}_t[(Y-\alpha)^2] \le \mathbb{E}_t\left[\left(Y - \left(\frac{a+b}{2}\right)\right)^2\right]$$

$$\le \max\left[\left(b - \frac{b+a}{2}\right)^2, \left(a - \frac{b+a}{2}\right)^2\right]$$

$$= \frac{(b-a)^2}{4}$$

(d) Using the Taylor expansion of $k(t)$ centered at $t=0$, with
$k(0) = 0$, $k'(0) = \mathbb{E}_0[X] = \mathbb{E}[X]$

$$k(t) = k(0) + k'(0)t + \frac{1}{2}k''(\tilde{t})t^2, \quad \tilde{t} \in [0,t].$$

$$= t\mathbb{E}[X] + \frac{1}{2}\mathrm{Var}_{\tilde{t}}(X) \cdot t^2$$

$$\le t\mathbb{E}[X] + \frac{1}{2}\sup_{s\in\mathbb{R}} \mathrm{Var}_s(X) \cdot t^2$$

$$= t\mathbb{E}[X] + \frac{1}{2}\left(\frac{(a+b)^2}{4}\right) \cdot t^2$$

$$\Rightarrow \underbrace{\log(\mathbb{E}e^{tX}) - \log e^{t\mathbb{E}X}} \le \frac{1}{2}\frac{t^2(a+b)^2}{4}$$

$$= \log(\mathbb{E}e^{tX} / e^{t\mathbb{E}X})$$

$$= \log(\mathbb{E}[\exp(t(X - \mathbb{E}X))])$$

$$\Rightarrow \mathbb{E}[\exp(t(X-\mathbb{E}X))] \le \exp\left(\frac{1}{8}t^2(a+b)^2\right).$$

The result follows.

3. The many faces of subgaussian and subexponential random variables

(a) Suppose X is subgaussian

$$\mathbb{E}\left[\exp(t(X-\mathbb{E}X))\right] \le \exp\left(\tfrac{1}{2}t^2\sigma^2\right)$$

$$\Rightarrow \mathbb{E}\left[\exp(t(cX-\mathbb{E}cX))\right]$$
$$= \mathbb{E}\left[\exp(ct(X-\mathbb{E}X))\right]$$
$$\le \exp\left(\tfrac{1}{2}(ct)^2\sigma^2\right)$$
$$= \exp\left(\tfrac{1}{2}t^2(c\sigma)^2\right).$$

Hence cX is $\sqrt{c^2\sigma^2} = |c|\sigma$ - subgaussian.

(b) WTS: if X is σ-subgaussian, then
$\mathbb{P}\left[|X-\mathbb{E}X| \ge t\right] \le 2\exp\left(-\frac{1}{2\sigma^2}t^2\right) \ \forall t \ge 0.$

Let $\lambda > 0$ be a parameter to be chosen later, $t \ge 0$.

$$\mathbb{P}\left[X-\mathbb{E}X \ge t\right] = \mathbb{P}\left[e^{\lambda(X-\mathbb{E}X)} \ge e^{\lambda t}\right]$$
$$\le e^{-\lambda t}\,\mathbb{E}e^{\lambda(X-\mathbb{E}X)}$$
$$\le e^{-\lambda t}\exp\left\{\tfrac{1}{2}\lambda^2\sigma^2\right\}$$
$$= \exp\left\{\tfrac{1}{2}\lambda^2\sigma^2 - \lambda t\right\}$$

Note: $\frac{1}{2}\lambda^2\sigma^2 - \lambda t$ is convex in λ, $x \mapsto e^x$ is convex and non-decreasing so the composition is convex, and we optimize over λ by choosing

$\frac{d}{d\lambda} \exp(\frac{1}{2}\lambda^2\sigma^2 - \lambda t) = 0 \Rightarrow \lambda\sigma^2 - t = 0 \Rightarrow \lambda = t/\sigma^2$

This gives $\mathbb{P}[X - \mathbb{E}X \geq t] \leq \exp\{\frac{1}{2}\frac{t^2}{\sigma^2} - \frac{t^2}{\sigma^2}\} = \exp\{-\frac{1}{2}t^2\sigma^2\}$

Similarly, $\mathbb{P}[\mathbb{E}X - X \geq t] = \mathbb{P}[e^{\lambda(\mathbb{E}X - X)} \geq e^{\lambda t}] \;\forall \lambda > 0$

$$\begin{aligned} &\leq e^{-\lambda t}\mathbb{E}[e^{\lambda(\mathbb{E}X - X)}] \\ &= e^{-\lambda t}\exp\{\tfrac{1}{2}\lambda^2\sigma^2\} \quad (*) \\ &= \exp\{\tfrac{1}{2}\lambda^2\sigma^2 - \lambda t\} \end{aligned}$$

Wok X subgaussian $\Rightarrow \mathbb{E}[t(\mathbb{E}X - X) \geq t] \leq \exp\{\frac{1}{2}t^2\sigma^2\}$ by symmetry in t.

by an identical argument as above, for $\lambda \leq 0$,

$$\mathbb{P}[(X - \mathbb{E}[X]) \leq -t] \leq \mathbb{E}[e^{\lambda(X - \mathbb{E}X)}]e^{\lambda t} \leq e^{\frac{\sigma^2\lambda^2}{2} + \lambda t}$$

this holds $\forall \lambda \leq 0$, so we optimize by taking $\lambda = \frac{-t}{\sigma^2} \leq 0$

$$\Rightarrow \mathbb{P}[(X - \mathbb{E}X) \leq -t] \leq e^{-\frac{\sigma^2 t^2}{2}}$$

we obtain using a union bound

$$\mathbb{P}[|\mathbb{E}X - X| \geq t] \leq 2 \cdot \exp\{-\tfrac{1}{2}t^2\sigma^2\}.$$

(c) Suppose $\mathbb{P}\left[|X-\mathbb{E}X| \geq t\right] \leq 2\exp\left(\frac{-1}{2\sigma^2}t^2\right) \ \forall t \geq 0.$

$$\mathbb{E}\left[|X-\mathbb{E}X|^p\right] = \int_0^\infty \mathbb{P}\left[|X-\mathbb{E}X|^p \geq u\right] du$$

$$= \int_0^\infty \mathbb{P}\left[|X-\mathbb{E}X| \geq t\right] p t^{p-1} dt \quad (u = t^p)$$

$$\leq \int_0^\infty 2\exp\left\{\frac{-1}{2\sigma^2}t^2\right\} p t^{p-1} dt \quad \text{(by assumption)}$$

$$= 2p\sigma^2 \int_0^\infty e^{-w} t^{p-2} dw \qquad w = \frac{1}{2\sigma^2}t^2, \quad dw = \frac{1}{\sigma^2} t\, dt$$

$$= 2p\sigma^2 \int_0^\infty e^{-w} (2\sigma^2)^{p/2-1} w^{p/2-1} dw$$

$$= p(2\sigma^2)^{p/2} \int_0^\infty e^{-w} w^{p/2-1} dw$$

$$= p(2\sigma^2)^{p/2}\,\Gamma(p/2)$$

$$\leq p(2\sigma^2)^{p/2}\left(3(p/2)^{p/2}\right) \quad \left(\Gamma(x) \leq 3x^x \ \forall x \geq 1/2\right)$$

Hence $\left(\mathbb{E}\left[|X-\mathbb{E}X|^p\right]\right)^{1/p} \leq p^{1/p}(2\sigma^2)^{1/2}\, 3^{1/p}\, (p/2)^{1/2} = (3p)^{1/p}\, p^{1/2} \cdot \sigma$

Note: if $p=1$, $\quad \|X-\mathbb{E}X\|_{L_p} \leq 3 \cdot 1^{1/2}\sigma$

if $p \geq 2$, $\quad \|X-\mathbb{E}X\|_{L_p} \leq p^{1/p} \cdot 3^{1/p} p^{1/2}\sigma \leq 3e\sqrt{p}\,\sigma$

with $p \geq 1 \Rightarrow p \geq \log p \Rightarrow \frac{\log p}{p} \leq 1 \Rightarrow p^{\frac{1}{p}} \leq e$

Hence $\exists\, c' \in \mathbb{R}$ s.t. $\forall p \geq 1$, $\|X-\mathbb{E}X\|_{L_p} \leq c'\sqrt{p}\,\sigma$, $c' = 3e$

We now show that for $K = \sqrt{3}\,\sigma e$ s.t. $\mathbb{E}\exp(\lambda^2 X^2) \le \exp(K^2\lambda^2)$ $\forall\ \lambda$ s.t. $|\lambda| \le \frac{1}{K}$

Using the Taylor series expansion of $x \mapsto \exp x$

$$\mathbb{E}\exp\left(\lambda^2 (X-\mathbb{E}X)^2\right) = \mathbb{E}\left[1 + \sum_{p=1}^{\infty} \frac{[\lambda^2 (X-\mathbb{E}X)^2]^p}{p!}\right]$$

$$= 1 + \sum_{p=1}^{\infty} \frac{\lambda^{2p}\,\mathbb{E}[(X-\mathbb{E}X)^{2p}]}{p!}$$

$$= 1 + \sum_{p=1}^{\infty} \frac{\lambda^{2p}\,\|X-\mathbb{E}X\|_{2p}}{p!}$$

$$\le 1 + \sum_{p=1}^{\infty} \frac{(\lambda c'\sqrt{2p}\,\sigma)^{2p}}{p!} = 1 + \sum_{p=1}^{\infty} \frac{(\lambda^2 c'^2 (2p)\,\sigma^2)^p}{p!} \quad \text{by previous bound on } \|X-\mathbb{E}X\|_{2p}$$

$$\le \sum_{p=1}^{\infty} \frac{(c'^2 (2p)\sigma^2 \cdot e\lambda^2)^p}{p^p} \qquad \text{Stirling's approximation: } p! \ge (p/e)^p$$

$$= \sum_{p=1}^{\infty} (c'^2 2\sigma^2 e\lambda^2)^p = (1 - c'^2\sigma^2 e\lambda^2) \quad \text{if } |\lambda^2 c'^2 \sigma^2 e| < 1$$

Since $c' = 3e$, we need $\lambda^2 < (3^2 e^3 \sigma^2)^{-1}$

Using $\frac{1}{1-x} \le e^{2x}$ for $x \in [0, 1/2]$,

Hence $\mathbb{E}\left[\exp(\lambda^2(X-\mathbb{E}X)^2)\right] \le \exp\left(2\lambda^2(3e\sigma)^2 e\right)$ $\forall\ |\lambda| < (3e^{3/2}\sigma)^{-1} =: K^{-1}$

We now show that $\exists\ \check{K} > 0$ s.t. $\mathbb{E}\left[\exp(\lambda(X-\mathbb{E}X))\right] \le \exp(\check{K}^2\lambda^2)$ $\forall\ \lambda \in \mathbb{R}$.

Suppose $K = 1$. Then for $|\lambda| \le 1$,

$$\mathbb{E}e^{\lambda(X-\mathbb{E}X)} \le \mathbb{E}\left[\lambda(X-\mathbb{E}X) + e^{\lambda^2(X-\mathbb{E}X)^2}\right] \qquad \text{using } e^x \le x + e^{x^2}\ \forall x \in \mathbb{R}$$

$$= \mathbb{E}\left[e^{\lambda^2(X-\mathbb{E}X)^2}\right] \le \exp\{2\lambda^2 3^2 e^3 \sigma^2\} \qquad \text{by step 2.}$$

Hence with $\hat{k}^2 = \sqrt{2} \cdot 3e^{3/2}\sigma$, $|\lambda| \leq 1$, $\mathbb{E}\, e^{\lambda(X-\mathbb{E}X)} \leq \exp(\lambda^2 \hat{k}^2)$

Now consider $|\lambda| > 1$. Then using $2\lambda x \leq \lambda^2 + x^2 \quad \forall \lambda, x$,

$$\mathbb{E}\, e^{\lambda(X-\mathbb{E}X)} \leq e^{\lambda^2/2}\, e^{(X-\mathbb{E}X)^2/2}$$
$$\leq e^{\lambda^2/2} \exp\{\lambda^2 3^2 e^3 \sigma^2\} \quad \text{by step 2}$$
$$\leq \exp\left\{\lambda^2\left(\tfrac{1}{2} + 3^2 e^3 \sigma^2\right)\right\}$$

Hence with $\hat{k}^2 = \sqrt{\tfrac{1}{2} + 3^2 e^3 \sigma^2}$, $|\lambda| > 1$, $\mathbb{E}\, e^{\lambda(X-\mathbb{E}X)} \leq \exp(\lambda^2 \hat{k}^2)$

With σ^2 given, by the above,

$$\boxed{\begin{gathered}\forall \lambda \in \mathbb{R},\ k^* := \max\left\{\sqrt{\tfrac{1}{2} + 3^2 e^3 \sigma^2},\ 3e^{3/2}\sigma\right\} \\ \mathbb{E}\exp(\lambda(X-\mathbb{E}X)) \leq \exp(k^{*2}\lambda^2).\end{gathered}}$$

(d) Wlog, let $\mathbb{E}X = 0$. We assume $\mathbb{E}[e^{tX}] \leq e^{t^2\sigma^2/2}$

WTS $\mathbb{E}\left[e^{t(X^2 - \mathbb{E}X^2)}\right] \leq e^{t^2(C_4\sigma^2)^2} \quad \forall |t| \leq 1/|C_2\sigma|^2$

Using a similar "moment" argument as in (c),

$$\mathbb{E}\left[e^{t(X^2 - E[X^2])}\right] \leq 1 + \sum_{k=2}^{\infty} \frac{|t|^k}{k!} \mathbb{E}\left[|X^2 - E[X^2]|^k\right]$$

$$\leq 1 + \sum_{k=2}^{\infty} \frac{|t|^k}{k!} 2^k \mathbb{E}\left[\left(\frac{|X^2| + |\mathbb{E}[X^2]|}{2}\right)^k\right]$$

$$\leq 1 + \sum_{k=2}^{\infty} \frac{|t|^k}{k!} 2^{k-1} \left(\mathbb{E}[X^{2k}] + E[X^2]^k\right)$$

$$\leq 1 + \sum_{k=2}^{\infty} \frac{|t|^k}{k!} 2^k \mathbb{E}[X^{2k}]$$ Jensen's inequality applied to $Y = X^2$

$$= 1 + \sum_{k=2}^{\infty} \frac{|t|^k}{k!} 2^k \|X\|_{L_{2k}}^{2k}$$, $\|X\|_{L_k} \leq 3e\sqrt{k}\,\sigma$ by (c)

$$\leq 1 + \sum_{k=2}^{\infty} \frac{(\sqrt{k})^{2k}}{k!} \cdot |t\, 2c'^2\sigma^2|^k$$

$$\leq 1 + \sum_{k=2}^{\infty} |e t 2c'^2\sigma^2|^k$$ $\left(k! \geq \left(\frac{k}{e}\right)^k\right.$ Stirling$\left.\right)$

$$\leq 1 + \frac{|et2c'^2\sigma^2|^2}{1 - |et2c'^2\sigma^2|}$$ if t s.t. $|et2c'^2\sigma^2| < 1/2$ so the sum converges

$$\leq 1 + 2|et2c'^2\sigma^2|^2$$

$$\leq \exp\left\{(\sigma^2 t C_2)^2 \frac{1}{2}\right\}, \quad C_2 = ec'^2 \cdot 4$$

This is what we wanted to show

4. The bounded difference inequality

(a) Denote X_i the i^{th} coordinate of X. X_i's are independent.
$f : \mathcal{X} \to \mathbb{R}$ is s.t. if X, X' differ in only one coordinate,

$$|f(X) - f(X')| \le L_i \quad \forall i \in [n].$$

Define $Y_i = \mathbb{E}[f(X) | X_1, \dots X_i]$, $V_i = Y_{i+1} - Y_i | \mathcal{F}_i$, Then $Y_0 = \mathbb{E}[f(X)]$, $Y_n = f(X)$

First we verify that this is a martingale adapted to $\mathcal{F}_i$, the (X_i)-adapted σ-algebra sequence

- $\mathbb{E}[\mathbb{E}[f(X) | X_1, \dots X_i]]$

$$= \int_{\mathbb{R}^i} \left(\int_{\mathbb{R}^{n-i}} f(X_1, \dots X_i, x_{i+1}, \dots x_n) \, d\mathbb{P}_{[X]_{k=1}^i} \right) d\mathbb{P}_{[X]_{k=i+1}^n}$$

$$= \int_{\mathbb{R}^n} f(x_1, \dots x_n) \, d\mathbb{P}_X = \mathbb{E}[f(X)]$$

But $\mathbb{E}[f(X)] < \infty$ since otherwise, $\exists x \in \mathcal{X}$ s.t.
$f(x) = \infty$ and hence $\exists x' \in \mathcal{X}$, $\forall M > 0$,

$$\begin{cases} |f(x) - f(x')| > ||f(x)| - |f(x')|| \ge M \text{ if } f(x') \text{ finite} \\ |f(x) - f(x')| \text{ is undefined if } f(x) = \infty \ \forall x \in \mathcal{X} \text{ and} \end{cases}$$

hence f can not satisfy the bounded difference property.

Further, $\mathbb{E}[Y_{i+1} | X_1, \dots X_i]$

$$= \mathbb{E}[\mathbb{E}[f(X) | X_1, \dots X_{i+1}] | X_1, \dots X_i]$$

$$= \int_{\mathbb{R}^{n-i}} \int_{\mathbb{R}^{n-i-1}} f(X_1, \dots X_i, x_{i+1}, \dots x_n) \, d\mathbb{P}_{[X]_{k=i+2}^n} \, d\mathbb{P}_{[X]_{k=i+1}^n} \quad , \quad \mathbb{P}_{[X]_{k=j}^n} \text{ is density of } (x_j, \dots x_n)$$

$$= \int_{\mathbb{R}^{n-i}} f(X_1, \dots X_i, x_{i+1}, \dots x_n) \, d\mathbb{P}_{[X]_{k=i+1}^n} \int_{\mathbb{R}^{n-i-1}} d\mathbb{P}_{[X]_{k=i+2}^n} \quad \text{(allowed by DCT)}$$

$$= \mathbb{E}[f(X) \mid X_1, \dots X_i]$$

$$= Y_i$$

Let us show that $Y_{i+1} - Y_i \mid X_1, \dots X_i$ is $\frac{L_{i+1}}{2}$ - subgaussian.
It is sufficient to show that $|Y_{i+1} - Y_i| \mid \mathcal{F}_i \leq \frac{L_{i+1}}{2} \; \forall i \in [n]$.

Define $U_i := \sup_{x \in \mathcal{X}} \mathbb{E}[f(X) \mid X_1, \dots X_i, x] - \mathbb{E}[f(X) \mid X_1, \dots X_i]$

$M_i = \inf_{x \in \mathcal{X}} \mathbb{E}[f(X) \mid X_1, \dots X_i, x] - \mathbb{E}[f(X) \mid X_1, \dots X_i]$

with $U_i \geq Y_{i+1} - Y_i \mid \mathcal{F}_i$, $\quad M_i \leq Y_{i+1} - Y_i \mid \mathcal{F}_i$

Note $(*)$ $\quad Y_{i+1} - Y_i \mid \mathcal{F}_i = \mathbb{E}_{X_{i+2} \dots X_n}[f(X_1, \dots X_i, X_{i+1}, X_{i+2}, \dots) \mid X_1, \dots X_i, X_{i+1}]$

$$- \mathbb{E}_{X_{i+1} \dots X_n}[f(X_1, \dots X_{i-1}, X_i, X_{i+1}, \dots X_n) \mid X_1, \dots X_i] \;\Big|\; \mathcal{F}$$

given $X_1, \dots X_i = x_1, \dots x_i$ $\quad = \mathbb{E}_{X_{i+2}, \dots X_n}[f(x_1, \dots x_i, X_{i+1}, X_{i+2}, \dots X_n) \mid X_{i+1}]$

$$- \mathbb{E}_{X_{i+1}}\left[\mathbb{E}_{X_{i+2}, \dots X_n}[f(x_1, \dots x_i, X_{i+1}, \dots X_n)]\right] \;\Big|\; \mathcal{F}$$

$$(*) = \mathbb{E}_{X_{i+2}, \dots X_n} \left[f(x_1, \dots x_i, X_{i+1}, X_{i+2}, \dots X_n) - \mathbb{E}_{X_{i+1}} [f(x_1, \dots x_i, X_{i+1}, \dots X_n] \mid X_{i+1} \right] \mid \mathcal{F}$$

$$V_i - M_i \leq \sup_{\alpha, \beta \in \mathcal{X}} \mathbb{E}_{X_{i+2}, \dots X_n} \left[f(x_1, \dots x_i, \alpha, X_{i+2}, \dots X_n) - f(x_1, \dots x_i, \beta, X_{i+2}, \dots X_n) \right] \mid \mathcal{F}$$

$$= \mathbb{E}_{X_{i+2}, \dots X_n} \left[\sup_{\alpha, \beta \in \mathcal{X}} f(x_1, \dots x_i, \alpha, X_{i+2}, \dots X_n) - f(x_1, \dots x_i, \beta, X_{i+2}, X_n) \right] \mid \tilde{\mathcal{F}}$$

$$\leq \mathbb{E}_{X_{i+2}, \dots X_n} \sup_{\alpha, \beta \in \mathcal{X}} \left| f(x_1, \dots x_i, \alpha, X_{i+2}, \dots X_n) - f(x_1, \dots x_i, \beta, X_{i+2}, X_n) \right] \mid \tilde{\mathcal{F}}$$

$$\leq \mathbb{E}_{X_{i+2} \dots X_n} L_i = L_i$$

Since the conditional differences have range at most L_i, they are $L_i/2$ sub-gaussian by exercise 2.

Recall $V_i := \mathbb{E}[f(X) \mid X_1, \dots X_{i+1}] - \mathbb{E}[f(X) \mid X_1, \dots X_i]$

Then $(f - \mathbb{E} f)(X) = \sum_{i=1}^{n} V_{i-1}$

$$\mathbb{P}[(f - \mathbb{E} f)(X) \geq t] = \mathbb{P}\left[\sum_{i=1}^{n} V_{i-1} \geq t\right]$$

$$= \mathbb{P}\left[s \sum_{i=1}^{n} V_{i-1} \geq st\right] \qquad s > 0$$

$$= \mathbb{P}\left[\exp\left(s \sum_{i=1}^{n} V_{i-1}\right) \geq \exp(st)\right]$$

$$\leq \exp(-st)\, \mathbb{E}\left[\exp\left(s \sum_{i=1}^{n} V_{i-1}\right)\right]$$

This holds $\forall s > 0$, so

$$\mathbb{P}\left[(f - \mathbb{E}f)(X) \geq t \right] \leq \inf_{s>0} \exp(-st)\, \mathbb{E}\left[\prod_{i=1}^{n} e^{sV_{i-1}} \right]$$

We can bound the MGF of the sum of (dependent) variables, since MGFs uniquely define distributions, the distribution itself will have bounded support and hence subgaussian.

By exercise 2, since $|V_i| \leq \frac{L_{i+1}}{2}$, V_i is $\frac{L_{i+1}}{2}$ subgaussian,

$$\mathbb{E}\left[\prod_{i=1}^{n} e^{sV_{i-1}} \right] = \mathbb{E}\,\mathbb{E}\left(\prod_{i=1}^{n-1} e^{sV_{i-1}} e^{sV_n} \,\middle|\, X_1, \dots X_{n-1} \right)$$

$$= \mathbb{E}\left(\prod_{i=1}^{n-2} e^{sV_{i-1}}\, \mathbb{E}\left[e^{sV_n} \,\middle|\, X_1, \dots X_{n-1}\right] \right)$$

$$\leq \mathbb{E}\left[\prod_{i=1}^{n-2} e^{sV_{i-1}} \right] \exp\left(s^2 \frac{L_{i+1}^2}{8} \right) \quad \text{by exercise 2}$$

$$\leq \exp\left(\frac{s^2}{8} \sum_{i=0}^{n-1} L_{i+1}^2 \right) = \exp\left(\frac{s^2}{8} \|L\|_2^2 \right) \text{ by induction}$$

Hence $\mathbb{P}\left[(f - \mathbb{E}f)(X) \geq t \right] \leq \inf_{s>0} \exp\left(-st + \frac{s^2}{8} \|L\|_2^2\right)$

$-st + \frac{s^2}{8}\|L\|_2^2$ is convex in s, exp is convex and non-decreasing, $\exp\left(-st \ \frac{s^2}{8}\|L\|_2^2\right)$ is convex and we optimize by choosing

$s = 4t / \|L\|_2^2$, and substituting

$$\mathbb{P}\left[|f - \mathbb{E}f|(x) \geq t \right] \leq \inf_{s>0} \exp\left(\frac{-4t^2}{\|L\|_2^2} + \frac{16t^2}{8\|L\|_2^2} \right) = \exp\left(-2t^2/\|L\|_2^2\right)$$

This is what we wanted to show.

(b) let $f : \mathcal{X} \to \mathbb{R}$ be L-Lipschitz.

Then $\forall x, y \in \mathcal{X}$, $|f(x) - f(y)| \leq L \|x - y\|_{\mathcal{X}}$

Since $x, y \in \mathcal{X} \subseteq [-B, B]^n$, Suppose x, y differ in only one coordinate, say $i \in [n]$. We have

$$\|x - y\|_{\mathcal{X}} \leq |B - (-B)| = 2BL$$

Hence $|f(x) - f(y)| \leq 2BL$ and hence f satisfies the bounded difference property with with parameter $\{L_i\}_{i=1}^n$, $|L_i| \leq 2BL$ $\forall i \in [n]$.

(c) Consider $f : \mathcal{X} \to \mathbb{R}$ defined by $f(x) = \|x\|_2$.
Then $\forall\, x, y \in [-B, B]^n$,

$$|f(x) - f(y)| = |\, \|x\|_2 - \|y\|_2 \,| \leq \|x - y\|_2$$

i.e. f is 1-Lipschitz. By (b) f has bounded difference property with parameters $\{L_i\}_{i=1}^n$ s.t. $L_i \leq 2B \ \forall i \in [n]$.

By (a), $\mathbb{P}\left[|f(x) - \mathbb{E}[f(x)]| \geq t \right] \leq 2\exp\left(-2 \frac{t^2}{\|L\|_2^2}\right)$

Note $\|L\|_2^2 \leq \sqrt{n} \cdot 2B$, so

$$\mathbb{P}\left[|f(x) - \mathbb{E}[f(x)]| \geq t \right] \leq 2\exp\left(\frac{-t^2}{2nB^2}\right)$$

By exercise 3 (c), $f(x)$ is $\underbrace{\sqrt{n} B}_{\sigma} \cdot C$ - sub-gaussian for C a universal constant, and hence $f(x)$ is $O(B\sqrt{n})$-subgaussian.

6. Ahlswede - Winter inequality

(a) Let $X \in \mathbb{R}^{d \times d}$ symmetric

X is orthogonally diagonalizable with eigenvalues as entries of D, i.e.

$X = UDU^T$, $D = \mathrm{diag}(\lambda_1(X), \dots \lambda_d(X))$

$\forall k > 0$, $X^k = UD^kU^T$,

$\exp(X) = \sum_{k=0}^{\infty} \frac{1}{k!} X^k = U \sum_{k=0}^{\infty} \frac{1}{k!} D^k \, U^T = U\tilde{D}U^T$,

$\tilde{D} = \mathrm{diag}(\exp(\lambda_1(X)), \exp(\lambda_2(X)), \dots \exp(\lambda_d(X)))$.

If the eigenvalues of X are $\lambda_1(X) < \dots < \lambda_d(X)$, then

eigenvalues of $\exp(X)$ are $\exp(\lambda_1(X)) < \dots < \exp(\lambda_d(X))$

because $\lambda \longmapsto \exp \lambda$ is increasing

Hence $\lambda_i(\exp(X)) = \exp(\lambda_i(X))$.

(b) Let $\theta > 0$, $t > 0$,

$$\mathbb{P}[\lambda_{max}(M) > t] = \mathbb{P}[\theta \lambda_{max}(M) > \theta t] \quad \forall t > 0$$

$$\leq \mathbb{E}[\theta \lambda_{max}(M)] e^{-\theta t} \quad \text{(Markov)}$$

$$\leq \mathbb{E}[\theta \exp(\lambda_{max}(M))] e^{-\theta t}$$

so $\exp M$ symmetric

$$= \mathbb{E}[\theta \lambda_{max}(\exp(M))] e^{-\theta t} \quad M \text{ symmetric, (a)}$$

$$\leq \mathbb{E}\left[\theta \cdot \sum_{i=1}^{d} \lambda_i (\exp(M))\right] e^{-\theta t} \quad \exp(M) \text{ has only positive evals}$$

$$= \mathbb{E}\left[\theta \sum_{i=1}^{d} \lambda_i (\exp(UDU^T))\right] e^{-\theta t}$$

$$= \mathbb{E}\left[\theta \sum_{i=1}^{d} \exp(\lambda_i (UDU^T))\right] e^{-\theta t} \quad \exp M \text{ symmetric}$$

$$= \mathbb{E}\left[\theta \sum_{i=1}^{d} \exp\{\lambda_i\}\right] e^{-\theta t}, \quad \lambda_i(UDU^T) = U \lambda_i(D) U = \lambda_i$$

$$= \mathbb{E}[tr(\exp(\theta M))] e^{-\theta t}$$

Since $\exp(\lambda_i) = \lambda_i(\exp(D)) = \lambda_i \exp(M)$,

$$tr(\exp M) = tr \exp(D) = \sum_{i=1}^{n} \lambda_i \exp(D) = \sum_{i=1}^{n} e^{\lambda_i}$$

(c) θM_i symmetric, so θM_i self-adjoint.
Sums of self adjoint matrices are self adjoint
thus $\exp(\sum_{i=1}^{k} \theta M_i)$ is PSD $\forall k$ because it's symmetric with positive eigenvalues. But for PSD matrix B, we have
$tr(AB) \leq \|A\|_2 \, tr(B)$ (*)

$$\mathbb{E}[tr(\exp \theta M)] = \mathbb{E}\left[tr\left(\exp \sum_{i=1}^{n} \theta M_i \right)\right]$$

$$= \mathbb{E}\left[tr\left(\prod_{i=1}^{n} \exp(\theta M_i)\right)\right] \qquad tr(A+B) = tr A + tr B$$

$$\leq \mathbb{E}\left[\|\exp \theta M_1\|_2 \, tr\left(\prod_{i=2}^{n} \exp(\theta M_i)\right)\right] \quad (*)$$

$$= \mathbb{E}\left[\lambda_{max}(\exp(\theta M_1)) \, tr\left(\prod_{i=1} \exp(\theta M_i)\right)\right]$$

$$\leq \mathbb{E}\left[\prod_{i=1}^{n} \lambda_{max}(\exp(\theta M_i)) \, tr(I_d) \right] \text{ by induction}$$

$$= d \prod_{i=1}^{n} \mathbb{E}\left[\lambda_{max}(\exp(\theta M_i))\right] \text{ by independence}$$

$$= d\left(\lambda_{max}\left(\mathbb{E}[\exp(\theta M_i)]\right)\right)^n$$

The last line uses $\mathbb{E}[\lambda_i(M_i)] = \lambda_i(E[M_i])$.
Indeed $\mathbb{E}\lambda_i(UDU^T) = \mathbb{E}[\lambda_i(D)] = \lambda_i \mathbb{E}[D] = \lambda_i \mathbb{E}[M_i]$.

(d) The matrices $A = \begin{bmatrix} 2 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$, $B = \begin{bmatrix} 2 & 0 & -2 \\ 0 & 1 & 0 \\ -2 & 0 & 1 \end{bmatrix}$, $C = \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 2 \end{bmatrix}$

are self-adjoint in $\mathbb{R}^{3\times 3}$. As can be shown with

the provided code, $tr(e^{A+B+C}) > tr(e^A e^B e^C)$

(e) $\exp(X) - I - X - X^2 = \sum_{k=3}^{\infty} \frac{X^a}{k!} - \frac{1}{2} X^2$

Note: $X^a = (UDU^T)^a = UD^a U^T$, U orthogonal, $D = diag(\lambda_1 \dots \lambda_d(X))$

Since $\|X\|_2 \leq 1$, $|\lambda_{max}(X)| < 1$ and hence $\sum_{k=3}^{\infty} \frac{X^a}{k!}$ converges

We rewrite $\exp(X) - I - X - X^2 = U \tilde{D} U^T$,

$\tilde{D} = diag\left(\sum_{k=3}^{\infty} \lambda_i^a(X) \frac{1}{k!} - \frac{1}{2} \lambda_i^2(X) \right)$

now since $g : y \mapsto \exp(y) - 1 - y - y^2$ negative on $[-1,1]$,

so $\sum_{k=3}^{\infty} \lambda_i^a(X) \frac{1}{k!} - \frac{1}{2} \lambda_i^2(X) \leq 0 \quad \forall i$

$\Rightarrow \exp(X) \preccurlyeq I + X + X^2$

Using $\theta \le 1$, $\|M_i\| \le 1 \Rightarrow \|\theta M_i\| \le 1$

$$\Rightarrow I + \theta M_i + \theta^2 M_i^2 \succcurlyeq \exp(\theta M_i) \quad (*)$$

$$\Rightarrow \mathbb{E}[I + \theta M_i + \theta^2 M_i^2] \succcurlyeq \mathbb{E}[\exp(\theta M_i)]$$ by (c), (*) contains PSD matrices

$$\Rightarrow I + \theta^2 \mathbb{E}[M_i^2] \succcurlyeq \mathbb{E}[\exp(\theta M_i)]$$

$$\Rightarrow \lambda_{max}(I + \theta^2 \mathbb{E}[M_i^2]) \geq \lambda_{max} \mathbb{E}[\exp(\theta M_i)]$$

$$\Rightarrow 1 + \theta^2 \lambda_{max} \mathbb{E}[M_i^2] \geq \lambda_{max} \mathbb{E}[\exp(\theta M_i)]$$

$$\Rightarrow 1 + \theta^2 \sigma^2 \geq \lambda_{max} \mathbb{E}[\exp(\theta M_i)]$$

$$\Rightarrow e^{\sigma^2\theta^2} \geq \lambda_{max} \mathbb{E}[\exp(\theta M_i)]$$ since $1 + x \le e^x \ \forall x \in \mathbb{R}$

f) For $0 \le \theta \le 1$, $t > 0$

$$\mathbb{P}[\lambda_{max}(M) > t] \le e^{-\theta t}\, \mathbb{E}[\mathrm{tr}(\exp(\theta M))] \quad (b)$$

$$\le e^{-\theta t}\, d.\left(\lambda_{max} \mathbb{E}[e^{\theta M_i}]\right)^n \quad (c)$$

$$\le e^{-\theta t}\, d\, e^{n\theta^2\sigma^2} \quad (e)$$

$$(*) \quad = d e^{n\theta^2\sigma^2 - \theta t} \qquad t = n\sigma^2$$

- If $t \le 2\sigma^2 n$. Take $\theta = \frac{t}{2\sigma^2 n} \le 1 \rightarrow (*) \le d\exp(-t^2/4n\sigma^2)$
- If $t > 2\sigma^2 n$ Note that $\exp(n\theta^2\sigma^2 - \theta t) \le \exp\left(\frac{t}{2}\theta^2 - t\theta\right)$

  Take $\theta = 1 \rightarrow (*) \le d\exp(-t/2)$

Hence $\mathbb{P}[\lambda_{max}(M) > t] \le d.\max\left(\exp(-t^2/4n\sigma^2), \exp(-t/2)\right)$

g) Define $M_i = v_i v_i^T - \Sigma = v_i v_i^T - E[v_i v_i^T]$

$\sigma^2 = \| E[M_i^2] \| = \| Var(v_i v_i^T) \|$

$$X \text{ PSD} \Rightarrow E[XX^T] - Var(X) = \overbrace{E[X]E[X]^T}^{\text{PSD}}$$

We have $\sigma^2 = \| E[M_i^2] \|$

$$\leq \| E[(v_i v_i^T)^2] \|$$

$$= \| E[\underbrace{\|v_i\|_2}_{\leq 1} (\underbrace{v_i v_i^T}_{\text{PSD}})] \|$$

$$\leq \| E[v_i v_i^T] \| = \| \Sigma \| \quad \text{a.s.}$$

Using $E[M_i] = 0$, $\|M_i\| = \|v_i v_i^T - \Sigma\| = \| v_i v_i^T - E[v_i v_i^T] \| \overset{\text{a.s.}}{\leq} tr(v_i v_i^T - E[v_i v_i^T]) \leq 1$ as $\|v\| \leq 1$ and by using a similar argument to 3(a)

Using Ahlswede-Winter inequality, for $t > 0$,

$$\mathbb{P}\left[\lambda_{max} \left(\sum_{i=1}^n v_i v_i^T - n\Sigma \right) > t \right] \leq d \max \left(e^{-t^2/4n\sigma^2}, e^{-t/2} \right)$$

Take $t = n \varepsilon \|\Sigma\|$ $= d \max (exp(-\varepsilon^2 \|\Sigma\|^2 n / 4\sigma^2, exp(-\varepsilon \|\Sigma\| n/2))$

We have $\sigma^2 \leq \|\Sigma\|$, let $n \geq \frac{c}{\varepsilon^2 \|\Sigma\|} \log d$

$$\mathbb{P}\left[\lambda_{max} \left(\frac{1}{n} \sum_{i=1}^n v_i v_i^T - \Sigma \right) \geq \varepsilon \|\Sigma\| \right]$$

$$\leq d \max \left(exp(-\varepsilon^2 \|\Sigma\| n / 4, exp(-\varepsilon \|\Sigma\| n / 2) \right)$$

$$\leq d \max\left(\exp(-c\log d/4),\ \exp(-c\log d/2\varepsilon)\right)$$

$$\leq \max\left(d^{1-\frac{c}{4}},\ d^{1-\frac{c}{2\varepsilon}}\right)$$

$$\leq d^{1-\frac{c}{4}} \quad \text{as } \varepsilon \in (0,1)$$

For c sufficiently large, $d^{1-c/4} \ll 1$, which is what we wanted to show

In [ ]:
```
import numpy as np
import matplotlib.pyplot as plt
from scipy.linalg import eigh
from scipy.stats import semicircular
```

In [3]:
```
def marchenko_pastur_density(x, alpha):
    """Compute Marchenko-Pastur density for given x and alpha."""
    b = (1 + np.sqrt(alpha))**2
    a = (1 - np.sqrt(alpha))**2
    return (1 / (2 * np.pi * alpha * x)) * np.sqrt((b - x) * (x - a)) * (

# Part (a): Generate covariance matrices and plot eigenvalue histograms
def plot_eigenvalue_histograms(d, n_values):
    plt.figure(figsize=(10, 6))
    for n in n_values:
        X = np.random.randn(d, n) / np.sqrt(n)  # Standard normal scaled
        Sigma = X @ X.T
        eigvals = eigh(Sigma, eigvals_only=True)

        alpha = d / n
        x_range = np.linspace(0, (1 + np.sqrt(alpha))**2, 1000)
        density = marchenko_pastur_density(x_range, alpha)

        plt.hist(eigvals, bins=50, density=True, alpha=0.5, label=f'eigen
        plt.plot(x_range, density, '--', label=f'MP Density (n={n})')

    plt.xlabel('Eigenvalue')
    plt.ylabel('Density')
    plt.title('Eigenvalue Histogram vs Marchenko-Pastur Density')
    plt.legend()
    plt.show()

# Part (b): Compute ||Σ_n - I||_op for different n
def plot_operator_norm_difference(d, n_values):
    errors = []
    for n in n_values:
        X = np.random.randn(d, n) / np.sqrt(n)
        Sigma = X @ X.T
        error = np.linalg.norm(Sigma - np.eye(d), ord=2)  # Operator norm
        errors.append(error)

    plt.figure(figsize=(8, 5))
    plt.plot(n_values, errors, marker='o', linestyle='-')
    plt.xlabel('n')
    plt.ylabel(r'$\hat{e}(n) = ||\hat{\Sigma}_n - I||_{op}$')
    plt.title('Operator Norm Difference vs n')
    plt.grid()
    plt.show()

# Part (c): Compute ||Σ_n - D_k||_op for different k and n
def plot_operator_norm_dk(d, k_values, n_values):
    plt.figure(figsize=(8, 5))
    for k in k_values:
        errors = []
        D_k = np.diag([1]*k + [(1/2) * (d-k) / (d-k)] * (d-k))
        for n in n_values:
            X = np.random.multivariate_normal(mean=np.zeros(d), cov=D_k,
            Sigma = X @ X.T
```

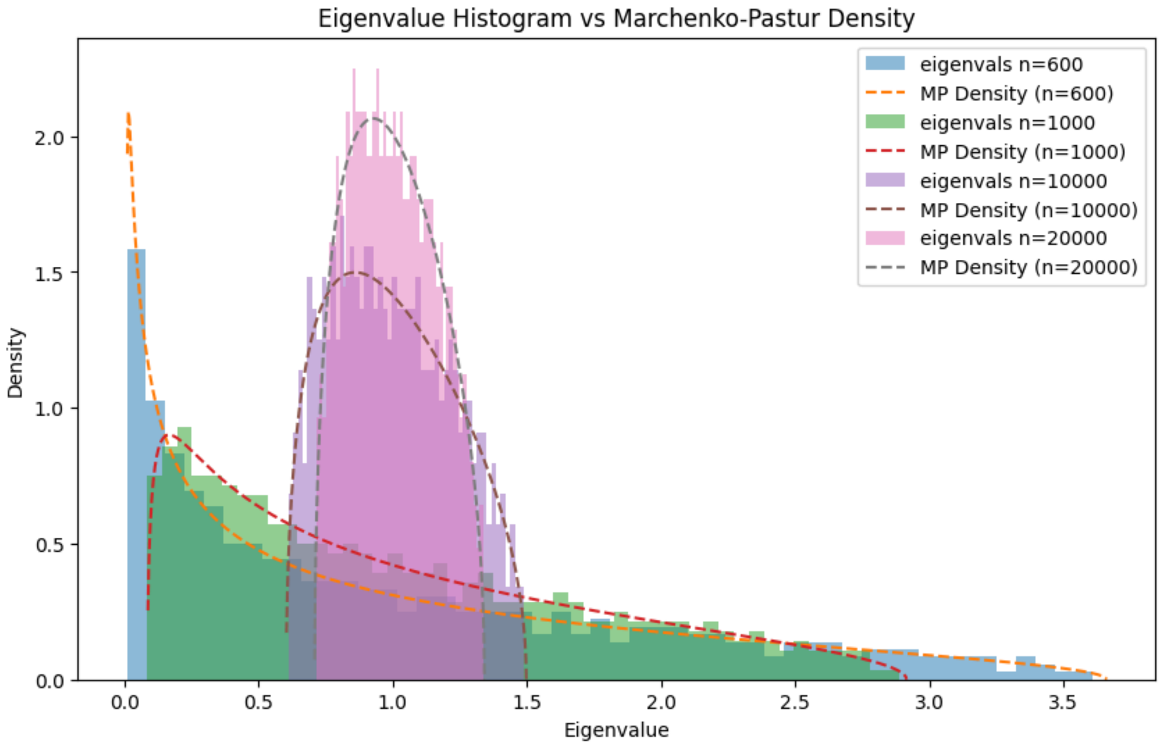
```
            error = np.linalg.norm(Sigma - D_k, ord=2)
            errors.append(error)
        plt.plot(n_values, errors, marker='o', linestyle='-', label=f'k={

    plt.xlabel('n')
    plt.ylabel(r'$\hat{e}_k(n) = ||\hat{\Sigma}_n - D_k||_{op}$')
    plt.title('Operator Norm Difference for Structured Covariance')
    plt.legend()
    plt.grid()
    plt.show()
```

Part (a)

In [5]:

```
# Run all parts
d = 500
n_values_a = [600, 1000, 10000, 20000]
plot_eigenvalue_histograms(d, n_values_a)
```

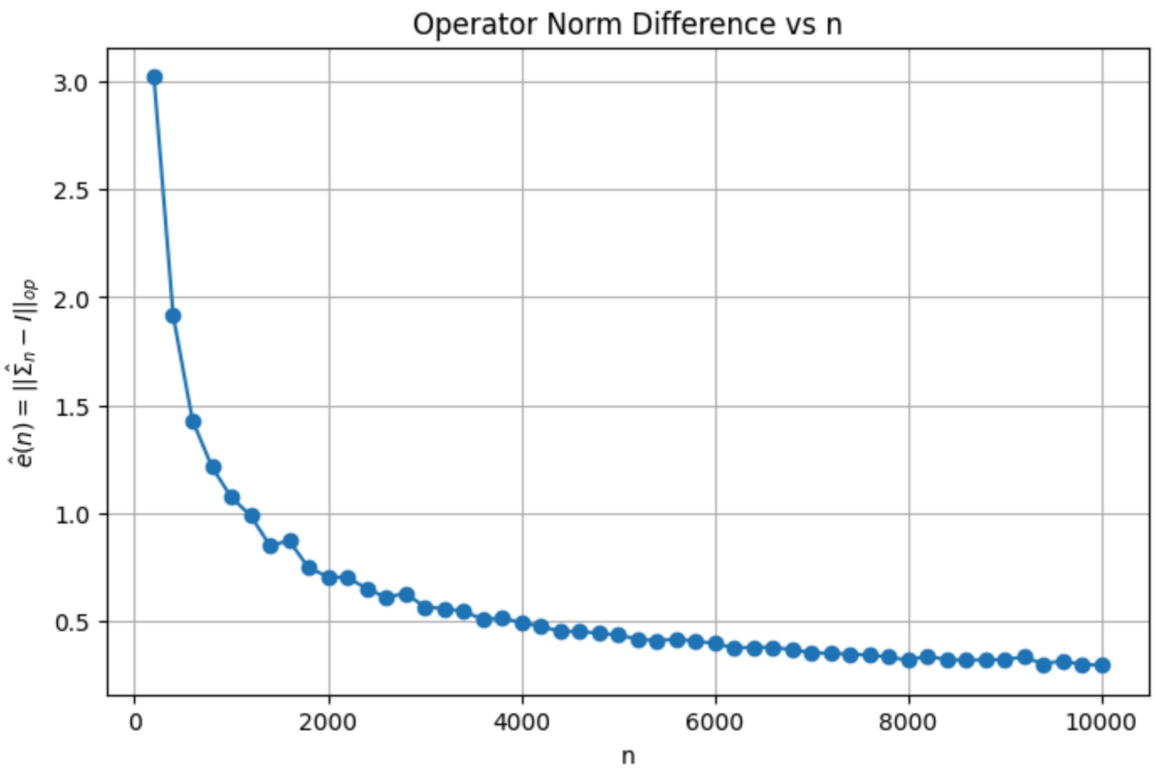
```
/var/folders/g2/8c32hb914ys7s2s9v6tmlcrm0000gn/T/ipykernel_19816/244745414
0.py:5: RuntimeWarning: divide by zero encountered in divide
  return (1 / (2 * np.pi * alpha * x)) * np.sqrt((b - x) * (x - a)) * (x >
= a) * (x <= b)
/var/folders/g2/8c32hb914ys7s2s9v6tmlcrm0000gn/T/ipykernel_19816/244745414
0.py:5: RuntimeWarning: invalid value encountered in sqrt
  return (1 / (2 * np.pi * alpha * x)) * np.sqrt((b - x) * (x - a)) * (x >
= a) * (x <= b)
```



Part (b)

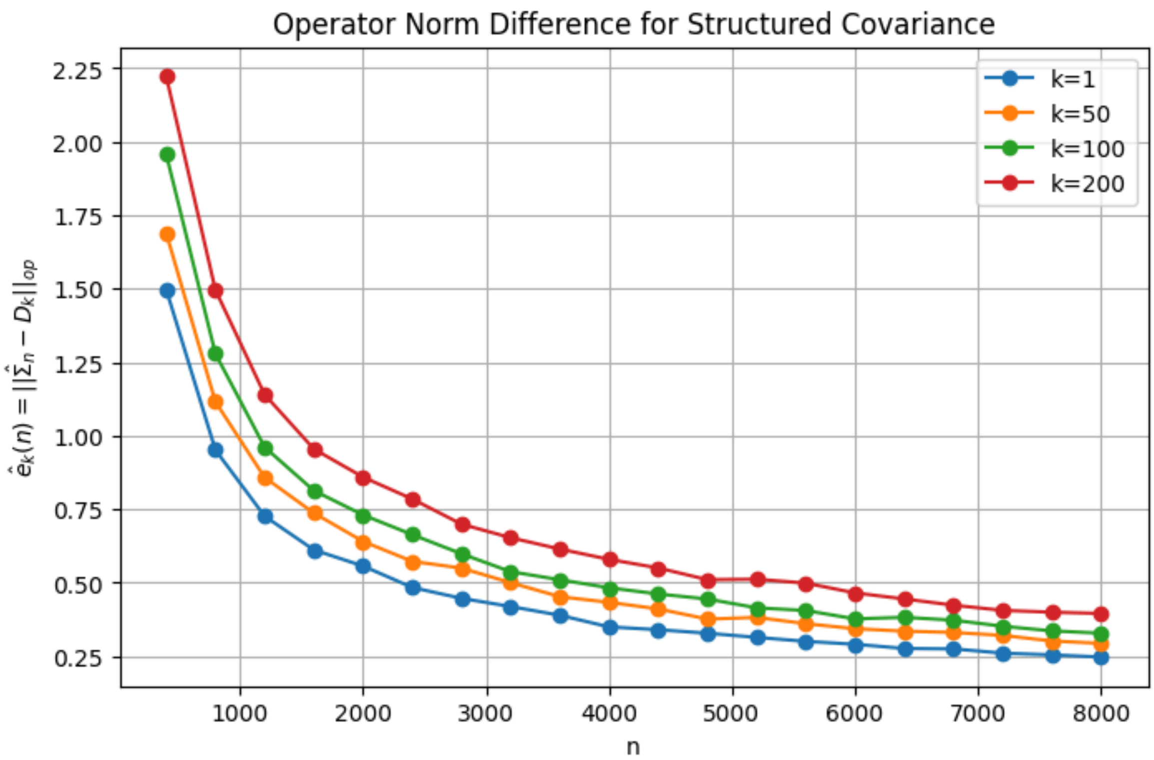
In [6]:

```
d = 200
n_values_b = np.arange(200, 10200, 200)
plot_operator_norm_difference(d, n_values_b)
```



Part (c)

In [7]:
```
d = 400
k_values_c = [1, 50, 100, 200]
n_values_c = np.arange(400, 8400, 400)
plot_operator_norm_dk(d, k_values_c, n_values_c)
```



We observe that while the rates remain indistinguishable, a lower k results in a better constant. Intuitively, reducing k decreases the effective number of dimensions, leading to an improved constant for $\sqrt{d/n}$.

Whatever this concept of "effective dimension" may be, it cannot depend solely on the trace and operator norm, as those remain constant. If we assume that the relevant definition is given by:

$$d_{\text{eff}}(\Sigma) = \frac{d}{\|\lambda(\Sigma)\|_1} f(\lambda(\Sigma))$$

the experiment suggests that f is strictly convex over $\mathbb{R}_+^d$. Possible choices for f include norms such as $\|x\|_k$ or functions like $\log \sum_{i=1}^n e^{\theta x_i}$, both of which converge to a limit for large k or θ. Specifically, they approach:

$$\max x \sum_{i=1}^d 1\{x_i = \max x\}$$

which corresponds to the operator norm in the absence of ties.

Additionally, we note that attempting to apply Corollary 39.4 from the lecture notes does not provide further insights for the same reason.

6(g)

In [8]:
```
import scipy.linalg

A = np.array([[2, 0, 0],
              [0, 1, 0],
              [0, 0, 1]])

B = np.array([[2, 0, -2],
              [0, 1, 0],
              [-2, 0, 1]])

C = np.array([[1, 0, 1],
              [0, 1, 0],
              [1, 0, 2]])

lhs_matrix = A + B + C
T1 = np.trace(scipy.linalg.expm(lhs_matrix))
T2 = np.trace(scipy.linalg.expm(A) @ scipy.linalg.expm(B) @ scipy.linalg.

print("Tr(exp(A+B+C)) =", T1)
print("Tr(exp(A) exp(B) exp(C)) =", T2)
print("Difference (T1 - T2) =", T1 - T2)
```

```
Tr(exp(A+B+C)) = 324.8616225402519
Tr(exp(A) exp(B) exp(C)) = 265.9650767893353
Difference (T1 - T2) = 58.89654575091663
```

In [ ]: