DSC 242: High-Dimensional Probability and Statistics 02 NOV 24

Homework #2 : Gabriel Riegner

Problem 1: Deduce Chebyshev's inequality by squaring both sides of the bound $|X-\mu| \geq 1$ and applying Markov's inequality.

Markov's inequality applied to the r.v. $|X-\mu|^k$ yields:

$$\mathbb{P}[\,|X-\mu| \geq t\,] \leq \frac{\mathbb{E}[\,|X-\mu|^k\,]}{t^k}$$

For the case where $K=1$ and $t=2$ we have:

$$\mathbb{P}[\,|X-\mu| \geq 1\,] \leq \frac{\mathbb{E}[\,|X-\mu|^2\,]}{1^2}$$

Since $\mathbb{E}[\,|X-\mu|^2\,]$ is the definition of variance, we arrive at Chebyshev's Ineq

$$\mathbb{P}[\,|X-\mu| \geq 1\,] \leq \mathbb{V}[X]$$

Problem 2: Let $a \leq X \leq b$ a.s. Show that X is $SG(b-a)$.

A r.v. X with mean $\mu = \mathbb{E}[X]$ is sub-Gaussian if there is a σ s.t.:

$$\mathbb{E}[e^{\lambda(X-\mu)}] \leq e^{\sigma^2\lambda^2/2} \quad \text{for all } \lambda \in \mathbb{R}$$

Take Z to be a normalized X s.t. $Z \in [-1,+1]$:

$$Z = \frac{X-\mu}{b-a}$$

To bound the MGF of Z, ie $\mathbb{E}[e^{\lambda Z}]$, we take the worst scenario where Z is a Rademaker r.v. that takes $\{-1,+1\}$ equiprobably.

$$\begin{aligned}
\mathbb{E}[e^{\lambda Z}] &= \frac{1}{2}(e^{\lambda} + e^{-\lambda}) = \frac{1}{2}\left(\sum_{k=0}^{\infty}\frac{\lambda^k}{k!} + \sum_{k=0}^{\infty}\frac{(-\lambda)^k}{k!}\right) \\
&= \sum_{k=0}^{\infty}\frac{\lambda^{2k}}{(2k)!} \\
&\leq 1 + \sum_{k=1}^{\infty}\frac{\lambda^{2k}}{(2k)!} = e^{\lambda^2/2}
\end{aligned}$$

$$\therefore \quad \mathbb{E}[e^{\lambda Z}] \leq e^{\lambda^2/2}$$

Rewrite this bound in terms of X, then substitute λ with $\lambda(b-a)$:

$$\mathbb{E}\left[e^{\lambda\left(\frac{X-\mu}{b-a}\right)}\right] \leq e^{\lambda^2/2} \Rightarrow \mathbb{E}[e^{\lambda(X-\mu)}] \leq e^{\frac{\lambda^2(b-a)^2}{2}}$$

$\therefore$ X is Sub-Gaussian with $\sigma = b-a$

Problem 3: Let $X \sim SG(\sigma)$. Use Chernoff's bound to show: for any $t > 0$ it holds that $\mathbb{P}[X > t] \leq \exp\left(-\frac{t^2}{2\sigma^2}\right)$ and $\mathbb{P}[X < -t] \leq \exp\left(-\frac{t^2}{2\sigma^2}\right)$

Define the sub-Gaussian r.v. (assuming X is zero mean)

$$\mathbb{E}[e^{\lambda X}] \leq e^{\sigma^2 \lambda^2 / 2} \quad \text{for all } \lambda \in \mathbb{R}$$

Substitute LHS of inequality:

$$\mathbb{P}[X > t] = \mathbb{P}[e^{\lambda X} > e^{\lambda t}]$$

Use Markov's inequality to bound the probability:

$$\mathbb{P}[e^{\lambda X} > e^{\lambda t}] \leq \mathbb{E}[e^{\lambda X}] / e^{\lambda X}$$

$$\mathbb{P}[X > t] \leq e^{\sigma^2 \lambda^2} / e^{\lambda t} = e^{\sigma^2 \lambda^2 / 2 - e^{\lambda t}}$$

optimize λ for tightest bound (ie Chernoff's bound):

$$\frac{d}{d\lambda}\left(\frac{\lambda^2 \sigma^2}{2} - \lambda t\right) = \lambda \sigma^2 - t = 0$$

$$\lambda = t / \sigma^2$$

Substitute $\lambda = t/\sigma^2$ into the definition of Sub-Gaussian:

$$\mathbb{P}[X > t] \leq e^{\sigma^2 \frac{(t/\sigma^2)}{2}} / e^{(t/\sigma^2) t}$$

$$= e^{t^2/2\sigma^2 - t^2/\sigma^2} = e^{-t^2/2\sigma^2}$$

$\therefore$ the probability that X exceeds t decays exponentially with t^2, with a rate depending on $2\sigma^2$. And by the symmetry of the sub-Gaussian definition, the r.v. $-X$ is also sub-Gaussian, so the same lower and upper deviation inequalities hold:

$$\mathbb{P}[X < -t] \leq e^{-t^2/2\sigma^2}$$

Problem 4: obtain a sharper version of Hoeffding's Inq, using Chernoff's Inq. for a Bernoulli sequence.

$X_i \sim \text{Bernoulli}(p_i)$ for $i = 1, \dots, n$. We would like to derive a bound on:

$$\mathbb{P}\left[\frac{1}{n}\sum_{i=1}^{n}(X_i - p_i) \geq t\right] = \mathbb{P}\left[\sum_{i=1}^{n}(X_i - p_i) \geq nt\right]$$

$$= \mathbb{P}[X_n - p_n \geq nt], \quad \text{where } X_n = \sum_{i=1}^{n} X_i \quad p_n = \sum_{i=1}^{n} p_i$$

Apply Chernoff's bound:

$$\mathbb{P}[X_n - p_n \geq nt] = \mathbb{P}[e^{\lambda(X_n - p_n)} \geq e^{\lambda n t}] \leq \frac{\mathbb{E}[e^{\lambda(X_n - p_n)}]}{e^{\lambda n t}}$$

$$\mathbb{P}[X_n - p_n \geq nt] \leq e^{-\lambda n t} \prod_{i=1}^{n} \mathbb{E}[e^{\lambda(X_i - p_i)}]$$

The MGF for each X_i:

$$\mathbb{E}[e^{\lambda(X_i - p_i)}] = e^{-\lambda p_i} \, \mathbb{E}[e^{\lambda X_i}] = e^{-\lambda p_i}(1 - p_i + p_i e^{\lambda})$$

Combine terms:

$$\mathbb{P}[X_n - p_n \geq nt] \leq e^{-\lambda n t} \prod_{i=1}^{n} e^{-\lambda p_i}(1 - p_i + p_i e^{\lambda})$$

$$= e^{-\lambda(nt + p_n)} \prod_{i=1}^{n} (1 - p_i + p_i e^{\lambda})$$

The tighest bound (by Chernoff's Ineq) requires:

$$\log \mathbb{P}[X_n - p_n \geq nt] \leq \inf_{\lambda \geq 0} \left\{ -\lambda(nt + p_n) + \sum_{i=1}^{n} \log(1 - p_i + p_i e^{\lambda}) \right\}$$

From the TA notes (01 NOV 2024) we can assume p_n is small. This special case simplifies the logarithm with the approximation:

$$\log \mathbb{P}[X_n - p_n \geq nt] \leq \inf_{\lambda \geq 0} \left\{ -\lambda(nt + p_n) + \sum_{i=1}^{n} p_i(e^{\lambda} - 1) \right\}$$

Optimize λ by differentiating w.r.t. λ and setting it to zero:

$$-(nt + p_n) + e^{\lambda} p_n = 0$$

$$e^{\lambda} = nt/p_n + 1$$

$$\lambda = \log(nt/p_n + 1)$$

$$= \log\left(\frac{nt}{\sum_{i=1}^{n} p_i} + 1\right)$$

$\therefore$ substituting λ back into the RHS of the tail bound:

$$\log \mathbb{P}[X_n - p_n \geq nt] \leq -\log(nt/p_n + 1)(nt + p_n) + \sum_{i=1}^{n} \log(1 - p_i + p_i(nt/p_n + 1))$$

$$\text{vs. } \log(2) - 2nt^2 \quad \text{(from slide \# 13)}$$

(I don't know how to interpret this, but I think the new bound is sharper than the expression on slide #13. This is because $2nt^2$ grows faster than the linear terms in n and will dominate for large n.)

Problem 5: Prove the given tail bound using Chernoff's method

To prove the following tail bound:

$$\mathbb{P}[X-\mu \geq t] \leq \begin{cases} e^{-t^2/2v^2} & \text{if } 0 \leq t \leq v^2/a \quad \leftarrow \text{Gaussian} \\ e^{-t/2a} & \text{if } t > v^2/a \quad \leftarrow \text{Exponential} \end{cases}$$

We start w/ Markov's Inequality to bound the probability:

$$\mathbb{P}[X-\mu \geq t] = \mathbb{P}[e^{\lambda(X-\mu)} \geq e^{\lambda t}] \leq \frac{\mathbb{E}[e^{\lambda(X-\mu)}]}{e^{\lambda t}}$$

By the sub-exponential property (pg 27 of textbook), the MGF bound is:

$$\mathbb{E}[e^{\lambda(X-\mu)}] \leq e^{v^2\lambda^2/2} \quad \text{for all } \lambda \in [0, 1/a]$$

Thus by Chernoff's bound:

$$\mathbb{P}[X-\mu \geq t] \leq \inf_{\lambda \in [0, 1/a]} \exp\underbrace{\left(-\lambda t + \frac{\lambda^2 v^2}{2}\right)}_{g(\lambda, t)}$$

Optimize λ by differentiating $g(\lambda, t)$ w.r.t. λ and setting it to zero:

$$\frac{\partial}{\partial \lambda} g(\lambda, t) = -t + \lambda v^2 = 0$$

$$\lambda = t/v^2$$

Substitute in λ to $g(\lambda, t)$ for the two cases in the initial tail bound:

I: $0 \leq t \leq v^2/a$

$$g(t/v^2, t) = -t\frac{t}{v^2} + \frac{v^2}{2}\left(\frac{t}{v^2}\right)^2 = -\frac{t^2}{2v^2}$$

$$\therefore \mathbb{P}[X-\mu \geq t] \leq e^{-t^2/2v^2} \quad \text{for } 0 \leq t \leq v^2/a$$

II: $t > v^2/a$. In this range, $\lambda = t/v^2$ exceeds $1/a$, so we use the boundary case $\lambda = 1/a$ instead:

$$g(1/a, t) = -\frac{t}{a} + \frac{v^2}{2}\left(\frac{1}{a}\right)^2 \leq -\frac{t}{2a} \quad \left(\text{since } \frac{v^2}{a} \leq t\right)$$

$$\therefore \mathbb{P}[X-\mu \geq t] \leq e^{-t/2a} \quad \text{for } t > v^2/a$$

Problem 6: Orlicz norms

(a) If $\|X\|_{\psi_q} < +\infty$, show there exists positive constants c_1, c_2 s.t.

$$\mathbb{P}[\,|X| > t\,] \leq c_1 \exp(-c_2 t^q) \quad \text{for all } t > 0$$

The Orlicz norm $\|X\|_{\psi_2}$ is defined by:

$$\|X\|_{\psi_2} = \inf\{t > 0 \mid \mathbb{E}[\psi_2(t^{-1}|X|)] \leq 1\}$$

where $\psi_2(u) = \exp(u^q) - 1$. Since $\|X\|_{\psi_q} < \infty$, there is $t = \|X\|_{\psi_2}$ s.t.

$$\mathbb{E}\left[\exp\left(\left[\frac{|X|}{\|X\|_{\psi_2}}\right]^q\right) - 1\right] \leq 1$$

$$\mathbb{E}\left[\exp\left(\left[\frac{|X|}{\|X\|_{\psi_2}}\right]^q\right)\right] \leq 2$$

Use Markov's inequality for any $t > 0$:

$$\mathbb{P}[\,|X| > t\,] = \mathbb{P}\left[\exp\left(\left[\frac{|X|}{\|X\|_{\psi_q}}\right]^q\right) > \exp\left(\left[\frac{t}{\|X\|_{\psi_2}}\right]^q\right)\right]$$

$$\leq \mathbb{E}\left[\exp\left(\left[\frac{|X|}{\|X\|_{\psi_q}}\right]^q\right)\right] \Big/ \exp\left(\left[\frac{t}{\|X\|_{\psi_2}}\right]^q\right)$$

Since the numerator is bounded by 2, we get:

$$\mathbb{P}[\,|X| > t\,] \leq 2 \Big/ \exp\left(\left[\frac{t}{\|X\|_{\psi_2}}\right]^q\right)$$

which holds when $c_1 = 2$ and $c_2 = \|X\|_{\psi_2}^{-q}$.

(b) Suppose that X satisfies the above tail bound. Show that $\|X\|_{\psi_q}$ is finite.

I don't know where to begin on this one and I'm very tired :(

Problem 7: Let X have χ_n^2 distribution. Show that with probability least $1-\delta$ $X \leq n + 2n\log(1/\delta) + 2\log(1/\delta)$

Let X have χ_n^2 distribution means $X = \sum_{i=1}^{n} Z_i^2$, where each Z_i is an independent standard normal r.v.

The MGF of a chi-squared r.v. is given by:

$$\mathbb{E}[e^{\lambda X}] = (1-2\lambda)^{-\frac{n}{2}} \quad \text{for } \lambda < 1/2$$

Use Markov's inequality to bound the probability:

$$\mathbb{P}[X > t] = \mathbb{P}[e^{\lambda X} > e^{\lambda t}]$$
$$\leq \mathbb{E}[e^{\lambda X}]\, e^{-\lambda t} = (1-2\lambda)^{-n/2} e^{-\lambda t}$$

$$\log \mathbb{P}[X > t] \leq \inf_{\lambda \in (0, 1/2)} \{ \underbrace{-\tfrac{n}{2}\log(1-2\lambda) - \lambda t}_{g(\lambda)} \}$$

optimize λ by differentiating $g(\lambda)$ w.r.t. λ and setting it to zero:

$$\frac{\partial}{\partial\lambda} g(\lambda) = \frac{n}{1-2\lambda} - t = 0$$
$$\lambda = \frac{t-n}{2t}$$

Substitute λ back into the probability bound:

$$1-2\lambda = n/t \quad \text{and} \quad \lambda t = \tfrac{1}{2}(t-n)$$
$$\mathbb{P}[X > t] \leq \left(\tfrac{t}{n}\right)^{\frac{n}{2}} e^{-\frac{1}{2}(t-n)}$$

Set $\mathbb{P}[X>t] = \delta$ and solve for t:

$$\left(\tfrac{t}{n}\right)^{\frac{n}{2}} e^{-\frac{1}{2}(t-n)} = \delta$$
$$\tfrac{n}{2}\log\left(\tfrac{t}{n}\right) - \tfrac{1}{2}(t-n) = \log(\delta)$$
$$t-n-n\log\left(\tfrac{t}{n}\right)) = -2\log(\delta)$$

The expression for t requires an approximation such as:

$$t = n + 2n\log(1/\delta) + 2n(1/\delta) \quad \text{(as given in the problem)}$$

We can verify this approximation holds by plugging it back in:

$$t-n-2n\log(t/n) = 2n\log(1/\delta) + 2\log(1/\delta) - n\log\left(1 + 2\log(1/\delta) + \tfrac{2}{n}\log(1/\delta)\right)$$

For large n and small δ, the log terms dominate and the approximation is reasonable.

$\therefore$ with probability at least $1-\delta$

$$X \leq n + 2n\log(1/\delta) + 2n\log(1/\delta)$$

Problem 8: Show an upper bound $\mathbb{E}[\max_{1\leq i\leq N}(X_i - p)] \leq 2\sqrt{p\log(N)} + 2\log(N)$ for $X_i \sim \chi^2_p$

From prop. 2.10 of the textbook, we have an MGF for any r.v. satisfying the Berstein condition:

$$\mathbb{E}[e^{\lambda(X_i-\mu)}] \leq e^{\frac{\lambda^2\sigma^2/2}{1-b|\lambda|}} \quad \text{for all } |\lambda| < 1/b$$

For $X \sim \chi^2_p$, $\mathbb{E}[X_i] = p$ and $\mathbb{V}[X_i] = 2p$, giving $\sigma^2 = 2p$ and $b = 2$.

For $|\lambda| < 1/2$, this simplifies to:

$$\mathbb{E}[e^{\lambda(X_i-p)}] \leq e^{\frac{\lambda^2 2p}{2(1-2|\lambda|)}} = e^{\frac{\lambda^2 p}{1-2|\lambda|}}$$

Referencing lecture slide 22, we maximize over N iid χ_p r.v.:

$$\mathbb{E}[\max_{1\leq i\leq N}(X_i-p)] \leq \frac{1}{\lambda}\log\left(\sum_{i=1}^{N}\mathbb{E}[e^{\lambda(X_i-p)}]\right)$$

$$= \frac{1}{\lambda}\left(\log(N) + \frac{\lambda^2 p}{1-2\lambda}\right)$$

Take $\lambda = \sqrt{\frac{\log(N)}{p}}$ which satisfies $\lambda < 1/2$ for large N (as $\log(N)$ grows slower than p for large p):

$$\mathbb{E}[\max_{1\leq i\leq N}(X_i-p)] \leq \frac{1}{\sqrt{\frac{\log(N)}{p}}}\left[\log(N) + \frac{\left(\sqrt{\frac{\log(N)}{p}}\right)^2 p}{1-2\sqrt{\frac{\log(N)}{p}}}\right]$$

Simplifying, this becomes:

$$\mathbb{E}[\max_{1\leq i\leq N}(X_i-p)] \leq 2\sqrt{p\log(N)} + 2\log(N)$$