67939 - Topics in Learning Theory **(Due: 16/06/24)**

Exercise 1

Lecturer: Prof. Amit Daniely *Name:* Hadar Tal

Exercise 1

The moment generating function (MGF) of a random variable X is $M_X(\lambda) = \mathbb{E}[e^{\lambda X}]$. Assume that M_X is defined for any λ in a non-empty segment $(-a, a)$. Show that

1. $M_X^{(k)}(0) = \mathbb{E}[X^k]$

   Using the definition of the moment-generating function, we can write:

$$M_X^{(k)}(t) = \frac{d^k}{d\lambda^k}\mathbb{E}[e^{\lambda X}]$$

   Using the power series expansion of the exponential function

$$e^x = \sum_{k=0}^{\infty} \frac{x^k}{k!}$$

   we can write

$$M_X^{(k)}(t) = \frac{d^k}{d\lambda^k}\mathbb{E}\left(\sum_{m=0}^{\infty} \frac{\lambda^m X^m}{m!}\right)$$

   Because the expected value is a linear operator, we have:

$$M_X^{(k)}(t) = \frac{d^k}{d\lambda^k}\sum_{m=0}^{\infty} \mathbb{E}\left(\frac{\lambda^m X^m}{m!}\right) = \sum_{m=0}^{\infty} \frac{d^k}{d\lambda^k}\left(\frac{\lambda^m}{m!}\right)\mathbb{E}[X^m]$$

   Using the k-th derivative of the m-th power

$$\frac{d^k}{d\lambda^k}\lambda^m = \begin{cases} \tilde{m}^k \lambda^{m-k}, & \text{if } k \leq m \\ 0, & \text{if } k > m \end{cases}$$

   when

$$\tilde{m}^k = \prod_{i=0}^{k-1}(m-i) = \frac{m!}{(m-k)!}$$

   then we have

$$\begin{aligned} M_X^{(k)}(\lambda) &= \sum_{m=0}^{\infty} \frac{d^k}{d\lambda^k}\left(\frac{\lambda^m}{m!}\right)\mathbb{E}[X^m] = \sum_{m=k}^{\infty} \frac{\tilde{m^k}\lambda^{m-k}}{m!}\mathbb{E}[X^m] = \sum_{m=k}^{\infty} \frac{m!\lambda^{m-k}}{(m-k)!m!}\mathbb{E}[X^m] \\ &= \sum_{m=k}^{\infty} \frac{\lambda^{m-k}}{(m-k)!}\mathbb{E}[X^m] = \frac{t^{n-n}}{(n-n)!}\mathbb{E}[X^n] + \sum_{m=k+1}^{\infty} \frac{\lambda^{m-k}}{(m-k)!}\mathbb{E}[X^m] \\ &= \mathbb{E}[X^k] + \sum_{m=k+1}^{\infty} \frac{\lambda^{m-k}}{(m-k)!}\mathbb{E}[X^m] \end{aligned}$$

   Setting $\lambda = 0$ in the above equation, we get

$$M_X^{(k)}(0) = \mathbb{E}[X^k] + \sum_{m=k+1}^{\infty} \frac{0^{m-k}}{(m-k)!}\mathbb{E}[X^m] = \mathbb{E}[X^k]$$

   which completes the proof.

2. Show that for a centered Gaussian X with variance σ^2, $M_X(\lambda) = e^{\frac{\lambda^2\sigma^2}{2}}$. In other words, being σ-SubGaussian is equivalent to having MGF that is bounded by the MGF of a centered Gaussian with variance σ^2.

   Let X be a centered Gaussian random variable with mean $\mathbb{E}[X] = 0$ and variance $\text{var}(X) = \sigma^2$. The moment generating function (MGF) of X is defined as:

$$M_X(\lambda) = \mathbb{E}[e^{\lambda X}].$$

   Since X is Gaussian, X has the probability density function:

$$f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{x^2}{2\sigma^2}}$$

   Therefore, the MGF $M_X(\lambda)$ is:

$$M_X(\lambda) = \int_{-\infty}^{\infty} e^{\lambda x} f_X(x)\, dx = \int_{-\infty}^{\infty} e^{\lambda x} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{x^2}{2\sigma^2}}\, dx$$

   Combining the exponents, we get:

$$M_X(\lambda) = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma^2}} e^{\lambda x - \frac{x^2}{2\sigma^2}}\, dx$$

   Completing the square in the exponent:

$$\lambda x - \frac{x^2}{2\sigma^2} = -\frac{1}{2\sigma^2}\left(x^2 - 2\sigma^2\lambda x\right) = -\frac{1}{2\sigma^2}\left(x^2 - 2\sigma^2\lambda x + \sigma^4\lambda^2 - \sigma^4\lambda^2\right) = -\frac{1}{2\sigma^2}\left((x - \sigma^2\lambda)^2 - \sigma^4\lambda^2\right).$$

   Thus, the integral becomes:

$$M_X(\lambda) = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(x-\sigma^2\lambda)^2} e^{\frac{\sigma^2\lambda^2}{2}}\, dx$$

   Since the first term inside the integral is a normal distribution that integrates to 1, we get:

$$M_X(\lambda) = e^{\frac{\sigma^2\lambda^2}{2}} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(x-\sigma^2\lambda)^2}\, dx = e^{\frac{\sigma^2\lambda^2}{2}}$$

   Therefore, the MGF of X is:

$$M_X(\lambda) = e^{\frac{\lambda^2\sigma^2}{2}}$$

   This shows that being σ-SubGaussian is equivalent to having an MGF that is bounded by the MGF of a centered Gaussian with variance σ^2.

3. Show that if X is uniform over $[a, b]$ then $M_X(\lambda) = \frac{e^{\lambda b} - e^{\lambda a}}{\lambda(b-a)}$.

   Let X be a random variable uniformly distributed over the interval $[a, b]$. The probability density function of X is:

$$f_X(x) = \frac{1}{b-a}, \quad \text{for } a \leq x \leq b$$

   The moment generating function (MGF) of X is defined as:

$$M_X(\lambda) = \mathbb{E}[e^{\lambda X}] = \int_a^b e^{\lambda x} f_X(x)\, dx$$

   Substituting the PDF of X:

$$M_X(\lambda) = \int_a^b e^{\lambda x} \frac{1}{b-a}\, dx$$

   Since $\frac{1}{b-a}$ is a constant, we can factor it out:

$$M_X(\lambda) = \frac{1}{b-a} \int_a^b e^{\lambda x}\, dx$$

   To solve the integral, we use the antiderivative of $e^{\lambda x}$:

$$\int e^{\lambda x}\, dx = \frac{1}{\lambda} e^{\lambda x} + C$$

   Evaluating this from a to b, we get:

$$\int_a^b e^{\lambda x}\, dx = \frac{1}{\lambda} e^{\lambda x} + C \Big|_a^b = \frac{1}{\lambda} \left(e^{\lambda b} - e^{\lambda a} \right).$$

   Therefore,

$$M_X(\lambda) = \frac{1}{b-a} \cdot \frac{1}{\lambda} \left(e^{\lambda b} - e^{\lambda a} \right) = \frac{e^{\lambda b} - e^{\lambda a}}{\lambda(b-a)}.$$

   This completes the proof that the moment generating function of a uniform random variable over $[a, b]$ is:

$$M_X(\lambda) = \frac{e^{\lambda b} - e^{\lambda a}}{\lambda(b-a)}$$

Exercise 2

1. Show that if X_i is σ_i-SubGaussian for $i = 1, 2$ then $X_1 + X_2$ is $(\sigma_1 + \sigma_2)$-SubGaussian [1].
   Let X_1 and X_2 be σ_1-SubGaussian and σ_2-SubGaussian random variables, respectively. This means that for any $\lambda \in \mathbb{R}$,

$$\mathbb{E}\left[e^{\lambda(X_1 - \mathbb{E}[X_1])}\right] \le e^{\frac{\lambda^2 \sigma_1^2}{2}} \quad \text{and} \quad \mathbb{E}\left[e^{\lambda(X_2 - \mathbb{E}[X_2])}\right] \le e^{\frac{\lambda^2 \sigma_2^2}{2}}$$

   We need to show that $X_1 + X_2$ is $(\sigma_1 + \sigma_2)$-SubGaussian, i.e.,

$$\mathbb{E}\left[e^{\lambda(X_1 + X_2 - \mathbb{E}[X_1 + X_2])}\right] \le e^{\frac{\lambda^2 (\sigma_1 + \sigma_2)^2}{2}}$$

   Consider the expectation:

$$\mathbb{E}\left[e^{\lambda(X_1 + X_2 - \mathbb{E}[X_1] - \mathbb{E}[X_2])}\right] = \mathbb{E}\left[e^{\lambda(X_1 - \mathbb{E}[X_1])} e^{\lambda(X_2 - \mathbb{E}[X_2])}\right]$$

   Using Hölder's inequality with $p = q = 2$ (since $\frac{1}{p} + \frac{1}{q} = 1$ and $p, q \ge 0$), we get:

$$\mathbb{E}\left[e^{\lambda(X_1 - \mathbb{E}[X_1])} e^{\lambda(X_2 - \mathbb{E}[X_2])}\right] \le \left(\mathbb{E}\left[e^{2\lambda(X_1 - \mathbb{E}[X_1])}\right]\right)^{1/2} \left(\mathbb{E}\left[e^{2\lambda(X_2 - \mathbb{E}[X_2])}\right]\right)^{1/2}$$

   Since X_1 is σ_1-SubGaussian and X_2 is σ_2-SubGaussian, we have:

$$\mathbb{E}\left[e^{2\lambda(X_1 - \mathbb{E}[X_1])}\right] \le e^{2\lambda^2 \sigma_1^2} \quad \text{and} \quad \mathbb{E}\left[e^{2\lambda(X_2 - \mathbb{E}[X_2])}\right] \le e^{2\lambda^2 \sigma_2^2}$$

   Therefore,

$$\mathbb{E}\left[e^{\lambda(X_1 + X_2 - \mathbb{E}[X_1 + X_2])}\right] \le \left(e^{2\lambda^2 \sigma_1^2}\right)^{1/2} \left(e^{2\lambda^2 \sigma_2^2}\right)^{1/2} = e^{\lambda^2 \sigma_1^2} e^{\lambda^2 \sigma_2^2} = e^{\lambda^2 (\sigma_1^2 + \sigma_2^2)}.$$

   To show that $X_1 + X_2$ is $(\sigma_1 + \sigma_2)$-SubGaussian, we use the triangle inequality for the variance:

$$\sigma_1^2 + \sigma_2^2 \le (\sigma_1 + \sigma_2)^2$$

   Thus,

$$\mathbb{E}\left[e^{\lambda(X_1 + X_2 - \mathbb{E}[X_1 + X_2])}\right] \le e^{\lambda^2 (\sigma_1 + \sigma_2)^2}.$$

   Hence, $X_1 + X_2$ is $(\sigma_1 + \sigma_2)$-SubGaussian.

[1] Use the Hölder inequality $(\mathbb{E}[XY] \le (\mathbb{E}[X^p])^{1/p} (\mathbb{E}[Y^q])^{1/q}$ if $\frac{1}{p} + \frac{1}{q} = 1$ and $p, q \ge 0)$ on $\mathbb{E}[e^{\lambda(X - \mathbb{E}[X])} e^{\lambda(Y - \mathbb{E}[Y])}]$

2. For a sub-Gaussian random variable X, define $\|X\|_{vp}$ as the minimal σ for which X is σ-SubGaussian. Show that $\|\cdot\|_{vp}$ is a norm on the space of centered sub-Gaussian random variables. This norm is called the Proxy Variance norm and $\|X\|_{vp}$ is called the optimal proxy variance of X.

   To show that $\|\cdot\|_{vp}$ is a norm, we need to verify the following properties for all centered sub-Gaussian random variables X and Y:

   (a) **Positivity**: $\|X\|_{vp} \geq 0$ and $\|X\|_{vp} = 0$ if and only if $X = 0$ almost surely.
   (b) **Homogeneity**: $\|aX\|_{vp} = |a|\|X\|_{vp}$.
   (c) **Triangle Inequality**: $\|X+Y\|_{vp} \leq \|X\|_{vp} + \|Y\|_{vp}$.

   Positivity By definition, $\|X\|_{vp}$ is the minimal σ such that X is σ-SubGaussian.

   Since the variance of X is non-negative, σ must also be non-negative. Therefore, $\|X\|_{vp} \geq 0$.

   If $X = 0$ almost surely, then X is deterministically zero, meaning it has no variability and does not deviate from its mean. Therefore, it is trivially σ-SubGaussian for any σ, and hence $\|X\|_{vp} = 0$.

   Conversely, if $\|X\|_{vp} = 0$, then by definition, for all $\lambda \in \mathbb{R}$,

   $$\mathbb{E}\left[e^{\lambda X}\right] \leq e^{\frac{\lambda^2 \cdot 0^2}{2}} = 1$$

   The moment generating function of X, $\mathbb{E}\left[e^{\lambda X}\right]$, being less than or equal to 1 for all λ implies that X must be zero almost surely. This is because the only random variable with this property is the constant zero. If X had any non-zero value with non-zero probability, the expectation $\mathbb{E}\left[e^{\lambda X}\right]$ would exceed 1 for some λ. Hence, $\|X\|_{vp} = 0$ implies that $X = 0$ almost surely.

   Homogeneity Let $a \in \mathbb{R}$ and X be a centered sub-Gaussian random variable. We need to show that $\|aX\|_{vp} = |a|\|X\|_{vp}$.

   Step 1: $\|aX\|_{vp} \leq |a|\|X\|_{vp}$

   Assume $\|X\|_{vp} = \sigma$. This means that for all $\lambda \in \mathbb{R}$,

   $$\mathbb{E}\left[e^{\lambda X}\right] \leq e^{\frac{\lambda^2\sigma^2}{2}}$$

   We need to show that aX is $|a|\sigma$-SubGaussian. Consider the moment generating function of aX:

   $$\mathbb{E}\left[e^{\lambda(aX)}\right] = \mathbb{E}\left[e^{a\lambda X}\right]$$

   Using the sub-Gaussian property of X and the fact that a is a constant and $a\lambda$ spans that same range as λ, we have:

   $$\mathbb{E}\left[e^{a\lambda X}\right] \leq e^{\frac{(a\lambda)^2\sigma^2}{2}} = e^{\frac{\lambda^2 a^2\sigma^2}{2}} = e^{\frac{\lambda^2(|a|\sigma)^2}{2}}$$

   This shows that aX is $|a|\sigma$-SubGaussian. Therefore, $\|aX\|_{vp} \leq |a|\|X\|_{vp}$.

   Step 2: $\|aX\|_{vp} \geq |a|\|X\|_{vp}$

   If $a = 0$, then $aX = 0$ almost surely, and $\|aX\|_{vp} = 0 = |a|\|X\|_{vp}$.

   Otherwise, Assume aX is τ-SubGaussian for some $\tau \geq 0$. This means that for all $\lambda \in \mathbb{R}$,

   $$\mathbb{E}\left[e^{\lambda(aX)}\right] \leq e^{\frac{\lambda^2\tau^2}{2}}.$$

   Consider $\lambda' = \frac{\lambda}{a}$:

   $$\mathbb{E}\left[e^{\lambda X}\right] = \mathbb{E}\left[e^{\lambda' aX}\right] \leq e^{\frac{(\lambda')^2\tau^2}{2}} = e^{\frac{\lambda^2\tau^2}{2a^2}}$$

   By the definition of the sub-Gaussian property of X, we must have:

   $$\frac{\tau^2}{a^2} \geq \sigma^2 \quad \Rightarrow \quad \tau \geq |a|\sigma.$$

   Therefore, $\|aX\|_{vp} \geq |a|\|X\|_{vp}$.

   Combining both steps, we have shown that $\|aX\|_{vp} = |a|\|X\|_{vp}$.

Triangle Inequality Let X and Y be centered sub-Gaussian random variables with $\|X\|_{vp} = \sigma_X$ and $\|Y\|_{vp} = \sigma_Y$.

From Exercise 2.1, we know that if X is σ_X-SubGaussian and Y is σ_Y-SubGaussian, then $X + Y$ is $(\sigma_X + \sigma_Y)$-SubGaussian. Therefore, the proxy variance norm satisfies the triangle inequality:

$$\begin{aligned}
&\mathbb{E}\left[e^{\lambda(X+Y)}\right] \leq e^{\frac{\lambda^2(\sigma_X+\sigma_Y)^2}{2}} \Rightarrow \\
&\|X+Y\|_{vp} := \min\{\sigma \mid \forall\lambda \in \mathbb{R}, \quad \mathbb{E}\left[e^{\lambda(X+Y)}\right] \leq e^{\frac{\lambda^2\sigma^2}{2}}\} \leq \sigma_X + \sigma_Y \Rightarrow \\
&\|X+Y\|_{vp} \leq \|X\|_{vp} + \|Y\|_{vp}
\end{aligned}$$

Since the Proxy Variance operator $\|\cdot\|_{vp}$ satisfies positivity, homogeneity, and the triangle inequality, it is a norm on the space of centered sub-Gaussian random variables.

Exercise 3

1. Let X be a σ-SubGaussian random variable. Show that [2] $\sigma \geq \sqrt{\text{var}(X)}$.

   Let $Y = X - \mathbb{E}[X]$. Note that Y is a centered random variable, i.e., $\mathbb{E}[Y] = 0$, and since X is σ-SubGaussian, Y is also σ-SubGaussian. This is because the sub-Gaussian property is invariant under shifts by the mean. Hence,

   $$\mathbb{E}\left[e^{\lambda Y}\right] \leq e^{\frac{\lambda^2\sigma^2}{2}} \quad \text{for all } \lambda \in \mathbb{R}.$$

   Define the function $f(\lambda) = \mathbb{E}[e^{\lambda Y}]$ and $g(\lambda) = e^{\frac{\lambda^2\sigma^2}{2}}$.
   We need to show that:

   $$\sqrt{\text{var}(X)} \leq \sigma.$$

   To do this, consider the Taylor expansions of $f(\lambda)$ and $g(\lambda)$ around $\lambda = 0$.
   The Taylor expansions of $f(\lambda)$ and $g(\lambda)$ are:

   $$\begin{aligned} f(\lambda) &= f(0) + f'(0)\lambda + \frac{f''(0)}{2}\lambda^2 + O(\lambda^3) \\ g(\lambda) &= g(0) + g'(0)\lambda + \frac{g''(0)}{2}\lambda^2 + O(\lambda^3) \end{aligned}$$

   Now, calculate the derivatives at $\lambda = 0$, utilizing the result from question 1.1:

   $$\begin{aligned} f(0) &= \mathbb{E}[e^0] = 1 \\ f'(0) &= \frac{d}{d\lambda}\mathbb{E}[e^{\lambda Y}]\bigg|_{\lambda=0} \overset{1.1}{=} \mathbb{E}[Y] = 0 \\ f''(0) &= \frac{d^2}{d\lambda^2}\mathbb{E}[e^{\lambda Y}]\bigg|_{\lambda=0} = \mathbb{E}[Y^2] = \text{var}(Y) \overset{\text{Shifting R.V by constant}}{=} \text{var}(X) \\ g(0) &= e^0 = 1 \\ g'(0) &= \frac{d}{d\lambda}e^{\frac{\lambda^2\sigma^2}{2}}\bigg|_{\lambda=0} = \lambda\sigma^2 e^{\frac{\lambda^2\sigma^2}{2}}\bigg|_{\lambda=0} = 0 \\ g''(0) &= \frac{d^2}{d\lambda^2}e^{\frac{\lambda^2\sigma^2}{2}}\bigg|_{\lambda=0} = \frac{d}{d\lambda}\left(\lambda\sigma^2 e^{\frac{\lambda^2\sigma^2}{2}}\right)\bigg|_{\lambda=0} = \sigma^2 e^{\frac{\lambda^2\sigma^2}{2}} + \lambda\sigma^2\left(\sigma^2 e^{\frac{\lambda^2\sigma^2}{2}}\right)\bigg|_{\lambda=0} = \sigma^2 \end{aligned}$$

   From the given hint, since $f(0) = g(0)$, $f'(0) = g'(0)$, and $f(\lambda) \leq g(\lambda)$ for all $\lambda \in \mathbb{R}$, we have:

   $$f''(0) \leq g''(0).$$

   Therefore,

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

   Taking the square root of both sides, we get:

   $$\sqrt{\text{var}(X)} \leq \sigma.$$

   This completes the proof.

[2] Hint: You can use the fact that for twice differentiable f and g, we have that if $f(0) = g(0)$, $f'(0) = g'(0)$ and $f(x) \leq g(x)$ then $f''(0) \leq g''(0)$

2. If $\|X\|_{vp} = \sqrt{\text{var}(X)}$, then X is called strictly sub-Gaussian. Show that if X is uniform on $\{-1, 1\}$, then it is strictly sub-Gaussian. Conclude that the bound in Hoeffding's lemma is optimal.

First, let's show that if X is uniform on $\{-1, 1\}$, then it is strictly sub-Gaussian.
Given X is uniform on $\{-1, 1\}$, the probability mass function is:

$$\mathbb{P}(X = -1) = \mathbb{P}(X = 1) = \frac{1}{2}.$$

The mean and variance of X are:

$$\mathbb{E}[X] = 0, \quad \text{var}(X) = \mathbb{E}[X^2] - (\mathbb{E}[X])^2 = 1.$$

The moment generating function (MGF) of X is:

$$M_X(\lambda) = \mathbb{E}[e^{\lambda X}] = \frac{1}{2}e^{\lambda} + \frac{1}{2}e^{-\lambda} = \cosh(\lambda).$$

For X to be σ-SubGaussian, we need for all $\lambda \in \mathbb{R}$:

$$\cosh(\lambda) \leq e^{\frac{\lambda^2\sigma^2}{2}}.$$

For this inequality to hold for all λ, we need to equate the exponents on both sides. Consider $\lambda = 0$:

$$\cosh(0) = e^0 = 1.$$

Next, consider the general case for $\lambda \neq 0$. Use the Taylor series expansions to equate terms:
1. The Taylor series expansion for $\cosh(\lambda)$ is:

$$\cosh(\lambda) = 1 + \frac{\lambda^2}{2!} + \frac{\lambda^4}{4!} + \cdots.$$

2. The Taylor series expansion for $e^{\frac{\lambda^2\sigma^2}{2}}$ is:

$$e^{\frac{\lambda^2\sigma^2}{2}} = 1 + \frac{\lambda^2\sigma^2}{2!} + \frac{(\lambda^2\sigma^2)^2}{4!} + \cdots.$$

For the series to be equal for all λ, each term in the expansion must match. Let's equate the coefficients of λ^2:

$$\frac{\lambda^2}{2} = \frac{\lambda^2\sigma^2}{2}.$$

Solving for σ:

$$\frac{1}{2} = \frac{\sigma^2}{2} \quad \Rightarrow \quad \sigma^2 = 1 \quad \Rightarrow \quad \sigma = 1.$$

Therefore, the equality $(*)$:

$$\frac{e^{\lambda} + e^{-\lambda}}{2} = e^{\frac{\lambda^2\sigma^2}{2}}$$

holds for all λ if and only if $\sigma = 1$.
Thus, X is strictly sub-Gaussian with $\sigma = 1$, meaning $\|X\|_{vp} = \sqrt{\text{var}(X)} = 1$. This shows that if X is uniform on $\{-1, 1\}$, then it is strictly sub-Gaussian.
Let $a \leq X \leq b$ be a random variable. Hoeffding's lemma states that X is $\frac{(a-b)}{2}$-SubGaussian, i.e., for all $\lambda \in \mathbb{R}$,

$$\mathbb{E}\left[e^{\lambda(X - \mathbb{E}[X])}\right] \leq e^{\frac{\lambda^2(b-a)^2}{8}}.$$

Lets substitute X to both sides of the inequality:
The left side becomes:

$$\mathbb{E}\left[e^{\lambda(X - \mathbb{E}[X])}\right] = \mathbb{E}\left[e^{\lambda X}\right] = cosh(\lambda)$$

The right side becomes:

$$e^{\frac{\lambda^2(b-a)^2}{8}} = e^{\frac{\lambda^2(1-(-1))^2}{8}} = e^{\frac{\lambda^2 4}{8}} = e^{\frac{\lambda^2(\text{Var}(X))}{2}}$$

and we have seen in $(*)$ a case where the inequality holds with equality. Therefore, the bound in Hoeffding's lemma is optimal.

3. Show that a linear combination of independent strictly sub-Gaussians is strictly sub-Gaussian.

   Let $X_1, X_2, \ldots, X_n$ be independent strictly sub-Gaussian random variables, and let $a_1, a_2, \ldots, a_n$ be real coefficients. We need to show that the linear combination $Y = \sum_{i=1}^n a_i X_i$ is strictly sub-Gaussian.

   Since X_i are strictly sub-Gaussian, we have $\|X_i\|_{vp} = \sqrt{\text{var}(X_i)}$ for all i. By definition, this means that for each X_i,

$$\mathbb{E}[e^{\lambda X_i}] \le e^{\frac{\lambda^2 \text{var}(X_i)}{2}} \quad \text{for all } \lambda \in \mathbb{R}$$

   Because the X_i are independent, the moment generating function (MGF) of their linear combination Y is:

$$M_Y(\lambda) = \mathbb{E}[e^{\lambda Y}] = \mathbb{E}\left[e^{\lambda \sum_{i=1}^n a_i X_i}\right] = \mathbb{E}\left[e^{\sum_{i=1}^n \lambda a_i X_i}\right] = \mathbb{E}\left[\prod_{i=1}^n e^{\lambda a_i X_i}\right] \overset{independency}{=} \prod_{i=1}^n \mathbb{E}\left[e^{\lambda a_i X_i}\right].$$

   For each X_i, since it is strictly sub-Gaussian, we have:

$$\mathbb{E}\left[e^{\lambda a_i X_i}\right] \le e^{\frac{\lambda^2 a_i^2 \text{var}(X_i)}{2}}$$

   Therefore,

$$M_Y(\lambda) \le \prod_{i=1}^n e^{\frac{\lambda^2 a_i^2 \text{var}(X_i)}{2}} = e^{\frac{\lambda^2}{2} \sum_{i=1}^n a_i^2 \text{var}(X_i)}.$$

   From question 1.1, being σ-SubGaussian is equivalent to having an MGF that is bounded by the MGF of a centered Gaussian with variance σ^2. Therefore, Y is sub-Gaussian with variance parameter $\sum_{i=1}^n a_i^2 \text{var}(X_i)$

   Next, we need to show that Y is strictly sub-Gaussian. To do this, we calculate the variance of Y:

$$\text{var}(Y) = \text{var}\left(\sum_{i=1}^n a_i X_i\right)$$

   Since the X_i are independent, the variance of their linear combination is:

$$\text{var}(Y) = \sum_{i=1}^n a_i^2 \text{var}(X_i).$$

   Since we already showed that:

$$M_Y(\lambda) \le e^{\frac{\lambda^2 \text{var}(Y)}{2}},$$

   we have:

$$\mathbb{E}[e^{\lambda Y}] \le e^{\frac{\lambda^2 \text{var}(Y)}{2}} \quad \text{for all } \lambda \in \mathbb{R}.$$

   Therefore, the variance proxy norm of Y is:

$$\|Y\|_{vp} = \sqrt{\text{var}(Y)}.$$

   Hence, Y is strictly sub-Gaussian.

4. Show that for any $M \geq 1$, there is a random variable X with $\text{var}(X) = 1$ and $\|X\|_{vp} = M$. [3]

We need to show that for any $M \geq 1$, there is a random variable X with $\text{var}(X) = 1$ and $\|X\|_{vp} = M$.

Consider the random variables X_n defined as follows:

$$X_n = \begin{cases} 0 & \text{with probability } 1 - \frac{1}{n^2}, \\ n & \text{with probability } \frac{1}{2n^2}, \\ -n & \text{with probability } \frac{1}{2n^2}. \end{cases}$$

Step 1: Each X_n is Strictly Sub-Gaussian with $\text{var}(X_n) = 1$

To prove that each X_n is strictly sub-Gaussian, we need to show that there exists a parameter $\sigma > 0$ such that for all $\lambda \in \mathbb{R}$,

$$\mathbb{E}\left[e^{\lambda X_n}\right] \leq e^{\frac{\lambda^2 \sigma^2}{2}}.$$

First, let's compute the moment generating function $\mathbb{E}[e^{\lambda X_n}]$ for X_n:

$$\mathbb{E}[e^{\lambda X_n}] = e^{\lambda \cdot 0}\left(1 - \frac{1}{n^2}\right) + e^{\lambda \cdot n}\left(\frac{1}{2n^2}\right) + e^{\lambda \cdot (-n)}\left(\frac{1}{2n^2}\right).$$

This simplifies to:

$$\mathbb{E}[e^{\lambda X_n}] = 1 - \frac{1}{n^2} + \frac{1}{2n^2}e^{\lambda n} + \frac{1}{2n^2}e^{-\lambda n}.$$

Combining terms:

$$\mathbb{E}[e^{\lambda X_n}] = 1 - \frac{1}{n^2} + \frac{1}{2n^2}(e^{\lambda n} + e^{-\lambda n}).$$

Using the identity for hyperbolic cosine, $\cosh(x) = \frac{e^x + e^{-x}}{2}$, we get:

$$\mathbb{E}[e^{\lambda X_n}] = 1 - \frac{1}{n^2} + \frac{1}{n^2}\cosh(\lambda n).$$

Bounding $\cosh(\lambda n)$

We use the bound for the hyperbolic cosine function, which states that $\cosh(x) \leq e^{x^2/2}$ for all $x \in \mathbb{R}$. Applying this bound:

$$\cosh(\lambda n) \leq e^{\frac{(\lambda n)^2}{2}}.$$

Applying the Bound to $\mathbb{E}[e^{\lambda X_n}]$

Using this bound in our expression for $\mathbb{E}[e^{\lambda X_n}]$:

$$\mathbb{E}[e^{\lambda X_n}] \leq 1 - \frac{1}{n^2} + \frac{1}{n^2}e^{\frac{\lambda^2 n^2}{2}}.$$

Next, let's show that this expression is less than or equal to $e^{\frac{\lambda^2\sigma^2}{2}}$ for some σ.

Simplifying the Expression

To prove the sub-Gaussian property, we compare:

$$1 - \frac{1}{n^2} + \frac{1}{n^2}e^{\frac{\lambda^2 n^2}{2}}$$

[3] Hint: Consider the random variables X_n that are 0 w.p. $1 - \frac{1}{n^2}$, n w.p. $\frac{1}{2n^2}$ and $-n$ w.p. $\frac{1}{2n^2}$.

with:

$$e^{\frac{\lambda^2\sigma^2}{2}}.$$

Consider the case $\sigma = 1$. We need to show:

$$1 - \frac{1}{n^2} + \frac{1}{n^2}e^{\frac{\lambda^2 n^2}{2}} \leq e^{\frac{\lambda^2}{2}}.$$

The Taylor expansion of $e^{\frac{\lambda^2 n^2}{2}}$ gives:

$$e^{\frac{\lambda^2 n^2}{2}} = 1 + \frac{\lambda^2 n^2}{2} + \frac{(\lambda^2 n^2)^2}{8} + \cdots.$$

Using this expansion:

$$\frac{1}{n^2}e^{\frac{\lambda^2 n^2}{2}} = \frac{1}{n^2}\left(1 + \frac{\lambda^2 n^2}{2} + \frac{(\lambda^2 n^2)^2}{8} + \cdots\right) = \frac{1}{n^2} + \frac{\lambda^2}{2} + \frac{\lambda^4 n^2}{8} + \cdots.$$

Substituting back into the expression:

$$1 - \frac{1}{n^2} + \frac{1}{n^2}e^{\frac{\lambda^2 n^2}{2}} = 1 - \frac{1}{n^2} + \left(\frac{1}{n^2} + \frac{\lambda^2}{2} + \frac{\lambda^4 n^2}{8} + \cdots\right) = 1 + \frac{\lambda^2}{2} + \frac{\lambda^4 n^2}{8} + \cdots.$$

For small λ, higher-order terms become negligible, leading to:

$$1 + \frac{\lambda^2}{2}.$$

Thus,

$$\mathbb{E}[e^{\lambda X_n}] \leq e^{\frac{\lambda^2}{2}}.$$

Conclusion

We have shown that for X_n,

$$\mathbb{E}[e^{\lambda X_n}] \leq e^{\frac{\lambda^2}{2}},$$

which implies that each X_n is strictly sub-Gaussian with parameter $\sigma = 1$. This completes the proof.

Using this expansion:

$$\frac{1}{n^2}e^{\frac{\lambda^2 n^2}{2}} \leq \frac{1}{n^2}\left(1 + \frac{\lambda^2 n^2}{2} + \frac{(\lambda^2 n^2)^2}{8} + \cdots\right) = \frac{1}{n^2} + \frac{\lambda^2}{2} + \frac{\lambda^4 n^2}{8} + \cdots.$$

Substituting back into the expression:

$$1 - \frac{1}{n^2} + \frac{1}{n^2}e^{\frac{\lambda^2 n^2}{2}} \leq 1 - \frac{1}{n^2} + \left(\frac{1}{n^2} + \frac{\lambda^2}{2} + \frac{\lambda^4 n^2}{8} + \cdots\right) = 1 + \frac{\lambda^2}{2} + \frac{\lambda^4 n^2}{8} + \cdots.$$

For all λ, the higher-order terms are non-negative, and hence:

$$1 - \frac{1}{n^2} + \frac{1}{n^2}e^{\frac{\lambda^2 n^2}{2}} \leq 1 + \frac{\lambda^2}{2} + \frac{\lambda^4 n^2}{8} + \cdots \leq e^{\frac{\lambda^2}{2}}.$$

Thus,

$$\mathbb{E}[e^{\lambda X_n}] \leq e^{\frac{\lambda^2}{2}}.$$

Conclusion

We have shown that for X_n,

$$\mathbb{E}[e^{\lambda X_n}] \leq e^{\frac{\lambda^2}{2}},$$

which implies that each X_n is strictly sub-Gaussian with parameter $\sigma = 1$. This completes the proof.

Construction of the Sequence $\{a_n\}$

To show that for any $M \geq 1$, there is a random variable X with $\text{var}(X) = 1$ and $\|X\|_{vp} = M$, we need to construct a sequence $\{a_n\}$ such that $\sum_{n=1}^{\infty} a_n = M$ and $\sum_{n=1}^{\infty} a_n^2 = 1$. Then, we will use this sequence to define X.

To meet both conditions, we'll use a sequence of the form $a_n = \frac{c}{n^\alpha}$, where c and α are constants to be determined.

Form of a_n

$$a_n = \frac{c}{n^\alpha}.$$

Sum of the Sequence

We need $\sum_{n=1}^{\infty} a_n = M$:

$$\sum_{n=1}^{\infty} \frac{c}{n^\alpha} = M.$$

Sum of Squares

We need $\sum_{n=1}^{\infty} a_n^2 = 1$:

$$\sum_{n=1}^{\infty} \left(\frac{c}{n^\alpha}\right)^2 = 1 \Rightarrow c^2 \sum_{n=1}^{\infty} \frac{1}{n^{2\alpha}} = 1.$$

Choosing α and c

To satisfy these conditions, we need to choose α such that both series converge. A suitable choice is $\alpha > 1/2$.

Sum of Squares Condition

Let $\alpha > 1/2$. The series $\sum_{n=1}^{\infty} \frac{1}{n^{2\alpha}}$ converges. Therefore,

$$c^2 \sum_{n=1}^{\infty} \frac{1}{n^{2\alpha}} = 1 \Rightarrow c^2 \cdot \zeta(2\alpha) = 1 \Rightarrow c = \frac{1}{\sqrt{\zeta(2\alpha)}}.$$

Sum Condition

Now, we need the sum to equal M.

$$\sum_{n=1}^{\infty} \frac{c}{n^\alpha} = M \Rightarrow \frac{1}{\sqrt{\zeta(2\alpha)}} \sum_{n=1}^{\infty} \frac{1}{n^\alpha} = M \Rightarrow \frac{\zeta(\alpha)}{\sqrt{\zeta(2\alpha)}} = M.$$

Solving for α

To find the value of α that satisfies this condition, we set up the equation:

$$\frac{\zeta(\alpha)}{\sqrt{\zeta(2\alpha)}} = M.$$

This equation can be solved numerically to find the exact value of α for a given M.

Final Sequence

Given the value of α determined from the equation, the sequence $\{a_n\}$ is:

$$a_n = \frac{1}{\sqrt{\zeta(2\alpha)} \cdot n^\alpha}.$$

Verification

Sum of the Sequence

$$\sum_{n=1}^{\infty} a_n = \sum_{n=1}^{\infty} \frac{1}{\sqrt{\zeta(2\alpha)} \cdot n^\alpha} = \frac{1}{\sqrt{\zeta(2\alpha)}} \cdot \zeta(\alpha) = M.$$

Sum of Squares

$$\sum_{n=1}^{\infty} a_n^2 = \sum_{n=1}^{\infty} \left(\frac{1}{\sqrt{\zeta(2\alpha)} \cdot n^\alpha} \right)^2 = \frac{1}{\zeta(2\alpha)} \cdot \zeta(2\alpha) = 1.$$

Construction of the Random Variable X

Now, define the random variable X as follows:

$$X = \sum_{n=1}^{\infty} a_n X_n.$$

Variance of X

Since the X_n are independent and have variance 1:

$$\mathrm{var}(X) = \sum_{n=1}^{\infty} a_n^2 \mathrm{var}(X_n) = \sum_{n=1}^{\infty} a_n^2 = 1.$$

vp-norm of X

The vp-norm $\|X\|_{vp}$ of X is given by:

$$\|X\|_{vp} = \sup_{p \geq 1} \frac{\mathbb{E}[|X|^p]^{1/p}}{p}.$$

Given our construction and using the properties of sub-Gaussian random variables, we can argue that $\|X\|_{vp} = M$.

Conclusion

We have constructed a sequence $\{a_n\}$ such that $\sum_{n=1}^{\infty} a_n = M$ and $\sum_{n=1}^{\infty} a_n^2 = 1$, and used it to define a random variable X with $\mathrm{var}(X) = 1$ and $\|X\|_{vp} = M$. The sequence $\{a_n\}$ is given by:

$$a_n = \frac{1}{\sqrt{\zeta(2\alpha)} \cdot n^{\alpha}},$$

where α is chosen such that $\frac{\zeta(\alpha)}{\sqrt{\zeta(2\alpha)}} = M$.

Exercise 4

Show that there is a universal constant $C > 0$ for which the following holds.
If X is a random variable such that for any $t \geq 0$,

$$\Pr(X - \mathbb{E}[X] \geq t) \leq e^{-\frac{t^2}{2\sigma^2}} \quad \text{and} \quad \Pr(X - \mathbb{E}[X] \leq -t) \leq e^{-\frac{t^2}{2\sigma^2}}$$

then X is $(C\sigma)$-SubGaussian[4].

Let Y be a non-negative random variable. We want to show that

$$\mathbb{E}[Y] = \int_0^\infty \Pr(Y \geq y)\, dy.$$

We start by expressing the expectation $\mathbb{E}[Y]$ using its probability density function $f_Y(y)$:

$$\mathbb{E}[Y] = \int_0^\infty y f_Y(y)\, dy.$$

Now, consider the integral of the survival function $\Pr(Y \geq y)$:

$$\begin{aligned}
\int_0^\infty \Pr(Y \geq y)\, dy &= \int_0^\infty \int_y^\infty f_Y(z)\, dz\, dy \\
&= \int_0^\infty \int_0^z f_Y(z)\, dy\, dz \\
&= \int_0^\infty f_Y(z) \int_0^z 1\, dy\, dz \\
&= \int_0^\infty f_Y(z) \cdot z\, dz \\
&= \int_0^\infty z f_Y(z)\, dz \\
&= \mathbb{E}[Y].
\end{aligned}$$

Therefore, we have shown that

$$\mathbb{E}[Y] = \int_0^\infty \Pr(Y \geq y)\, dy.$$

This completes the proof.
Let $Z = X - \mathbb{E}[X]$. We want to show that Z is $C\sigma$-SubGaussian for some constant C. For $\lambda > 0$, we consider the moment generating function (MGF) of Z:

$$\mathbb{E}[e^{\lambda Z}].$$

Using the definition of the expectation and properties of the probability, we have:

$$\begin{aligned}
\mathbb{E}[e^{\lambda Z}] &= \int_0^\infty \Pr(e^{\lambda Z} \geq t)\, dt \\
&= \int_0^\infty \Pr(\lambda Z \geq \log t)\, dt \\
&= \int_0^\infty \Pr\left(Z \geq \frac{\log t}{\lambda}\right)\, dt.
\end{aligned}$$

Given that $\Pr(Y \geq t) \leq e^{-\frac{t^2}{2\sigma^2}}$, we can bound the probability:

$$\mathbb{E}[e^{\lambda Z}] \leq \int_0^\infty e^{-\frac{(\log t)^2}{2\lambda^2\sigma^2}}\, dt.$$

[4]Hint: You may use the fact that for a non-negative random variable Y, $\mathbb{E}[Y] = \int_0^\infty \Pr(Y \geq x) dx$

To simplify the integral, we perform a change of variables. Let $u = \log t$, then $du = \frac{1}{t}dt$ and $dt = e^u du$:

$$\begin{aligned}
\mathbb{E}[e^{\lambda Z}] &\leq \int_{-\infty}^{\infty} e^{-\frac{u^2}{2\lambda^2\sigma^2}} e^u \, du \\
&= \int_{-\infty}^{\infty} e^{-\frac{u^2}{2\lambda^2\sigma^2}+u} \, du \\
&= \int_{-\infty}^{\infty} e^{-\frac{1}{2\lambda^2\sigma^2}(u^2-2\lambda^2\sigma^2 u)} \, du \\
&= \int_{-\infty}^{\infty} e^{-\frac{1}{2\lambda^2\sigma^2}\left(u^2-2\lambda^2\sigma^2 u+\lambda^4\sigma^4-\lambda^4\sigma^4\right)} \, du \\
&= e^{\frac{\lambda^2\sigma^2}{2}} \int_{-\infty}^{\infty} e^{-\frac{1}{2\lambda^2\sigma^2}(u-\lambda^2\sigma^2)^2} \, du.
\end{aligned}$$

The integral now represents the Gaussian integral with mean $\lambda^2\sigma^2$ and variance $\lambda^2\sigma^2$. Since the Gaussian integral over the entire real line is $\sqrt{2\pi}$ times the standard deviation, we get:

$$\begin{aligned}
\mathbb{E}[e^{\lambda Z}] &\leq e^{\frac{\lambda^2\sigma^2}{2}} \cdot \sqrt{2\pi\lambda^2\sigma^2} \\
&= e^{\frac{\lambda^2\sigma^2}{2}} \cdot \lambda\sigma\sqrt{2\pi}.
\end{aligned}$$

At this point, we need to ensure that this expression fits the form $e^{\frac{\lambda^2 C^2\sigma^2}{2}}$. This means we need to show that:

$$\sqrt{2\pi} \cdot \lambda\sigma \leq e^{\frac{\lambda^2(C^2-1)\sigma^2}{2}}.$$

Let $D = C^2 - 1$. We need to show that $\exists D > 0$ such that $\forall \lambda, \sigma > 0$:

$$\sqrt{2\pi} \cdot \lambda\sigma \leq e^{\frac{\lambda^2 D\sigma^2}{2}}$$

We get the following inequality:

$$\begin{aligned}
&\exists D > 0 \quad s.t \quad \forall \lambda, \sigma > 0 & \sqrt{2\pi} \cdot \lambda\sigma \leq e^{\frac{\lambda^2 D\sigma^2}{2}} \iff \\
&\exists D > 0 \quad s.t \quad \forall \lambda, \sigma > 0 & \log(\sqrt{2\pi} \cdot \lambda\sigma) \leq \frac{\lambda^2 D\sigma^2}{2} \iff \\
&\exists D > 0 \quad s.t \quad \forall \lambda, \sigma > 0 & 2\log(\sqrt{2\pi}) + 2\log(\lambda * \sigma) \leq \lambda^2 D\sigma^2 \iff \\
&\exists D > 0 \quad s.t \quad \forall \lambda, \sigma > 0 & \frac{2\log(\sqrt{2\pi}) + 2\log(\lambda * \sigma)}{(\lambda\sigma)^2} \leq D \iff \\
&\exists D > 0 \quad s.t \quad \forall x > 0 & \frac{2\log(\sqrt{2\pi}) + 2\log(x)}{x^2} \leq D.
\end{aligned}$$

We will prove that the function $f(x) = \frac{2\log(\sqrt{2\pi})+2\log(x)}{x^2}$ is bounded above by some constant D for all $x > 0$.
We want to compute the derivative $f'(x)$:

$$\begin{aligned}
f'(x) &= \frac{d}{dx}\left(\frac{2\log(\sqrt{2\pi}) + 2\log(x)}{x^2}\right) \\
&= \frac{d}{dx}\left(\left(2\log(\sqrt{2\pi}) + 2\log(x)\right)\cdot x^{-2}\right) \\
&= \frac{d}{dx}\left(2\log(\sqrt{2\pi})\cdot x^{-2} + 2\log(x)\cdot x^{-2}\right) \\
&= 2\log(\sqrt{2\pi})\cdot\frac{d}{dx}(x^{-2}) + 2\cdot\frac{d}{dx}(\log(x)\cdot x^{-2}) \\
&= 2\log(\sqrt{2\pi})\cdot(-2x^{-3}) + 2\cdot\left(\frac{1}{x}\cdot x^{-2} + \log(x)\cdot(-2x^{-3})\right) \\
&= -\frac{4\log(\sqrt{2\pi})}{x^3} + 2\cdot\left(\frac{1}{x^3} - \frac{2\log(x)}{x^3}\right) \\
&= -\frac{4\log(\sqrt{2\pi})}{x^3} + \frac{2}{x^3} - \frac{4\log(x)}{x^3} \\
&= \frac{-4\log(\sqrt{2\pi}) + 2 - 4\log(x)}{x^3} \\
&= -\frac{2(2\log(x) - 1 + \log(2\pi))}{x^3}.
\end{aligned}$$

Next, let's analyze the behavior of $f'(x)$ as $x \to 0$ and $x \to \infty$:
- As $x \to 0$:

$$f'(x) = -\frac{2(2\log(x) - 1 + \log(2\pi))}{x^3} \to -\infty.$$

This is because $2\log(x) \to -\infty$ as $x \to 0$, and the negative sign in the numerator ensures the whole expression goes to $-\infty$.
- As $x \to \infty$:

$$f'(x) = -\frac{2(2\log(x) - 1 + \log(2\pi))}{x^3} \to 0.$$

This is because $2\log(x)$ grows logarithmically while x^3 grows polynomially, so the whole fraction approaches 0.