

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:

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

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:

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

Therefore,

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

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 $(\sigma_1 + \sigma_2)$ -SubGaussian¹.

Let X_1 and X_2 be σ_1 -SubGaussian and σ_2 -SubGaussian random variables, respectively.

This means that for any $\lambda \in \mathbb{R}$,

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

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

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

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 $p, q \geq 0$), we get:

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

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

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

Therefore,

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

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

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

Thus,

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

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

¹Use the Hölder inequality $(\mathbb{E}[XY]) \leq (\mathbb{E}[X^p])^{1/p} (\mathbb{E}[Y^q])^{1/q}$ if $\frac{1}{p} + \frac{1}{q} = 1$ and $p, q \geq 0$ 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 $\lambda \in \mathbb{R}$,

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

The moment generating function of X , $\mathbb{E}[e^{\lambda X}]$, 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 $\mathbb{E}[e^{\lambda X}]$ 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 $\lambda \in \mathbb{R}$,

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

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

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

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:

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

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 $\lambda \in \mathbb{R}$,

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

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

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

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,

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

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:

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

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.

²Hint: 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.

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

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

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 $(C\sigma)$ -SubGaussian³.

³Hint: You may use the fact that for a non-negative random variable Y , $\mathbb{E}[Y] = \int_0^\infty \Pr(Y \geq x)dx$