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

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].

[1] Use 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.

Exercise 3

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

[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.

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 $\mathrm{var}(X) = 1$ 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 $(C\sigma)$-SubGaussian[3].

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