RWTH Aachen University **Winter 2023/24**

Instructor: Prof. Dr. Semih Çaycı
Teaching Assistant: Johannes Müller, M.Sc.

Mathematical Foundations of Deep Learning (11.80020)
Assignment 0 (Voluntary exercises)

Reminder. For a real random variable X we call (if existent)

$$\varphi_X(\lambda) \coloneqq \log \mathbb{E}[e^{\lambda X}] \quad \text{and } \widetilde{\varphi}_X(\lambda) \coloneqq \log \mathbb{E}[e^{\lambda(X-\mathbb{E}X)}]$$

the *logarithmic moment generating function* or shortly *log-moment generating function* and the *centered logarithmic moment generating function* or shortly *centered log-moment generating function*, respectively. Note that $\varphi_X, \widetilde{\varphi}_X \colon \mathbb{R} \to \mathbb{R}_{\geq 0} \cup \{+\infty\}$ and we denote their *domains* by

$$\mathrm{dom}(\varphi_X) \coloneqq \{\lambda \in \mathbb{R} : \varphi_X(\lambda) < \infty\} = \mathrm{dom}(\widetilde{\varphi}_X) \coloneqq \{\lambda \in \mathbb{R} : \widetilde{\varphi}_X(\lambda) < \infty\}.$$

We call

$$\varphi_X^\star(\lambda) \coloneqq \sup_{\lambda>0}\{\lambda t - \varphi_X(\lambda)\} \quad \text{and } \widetilde{\varphi}_X^\star(\lambda) \coloneqq \sup_{\lambda>0}\{\lambda t - \widetilde{\varphi}_X(\lambda)\}$$

the *Cramer transform* of φ_X and $\widetilde{\varphi}_X$, respectively. We call

$$\varphi_X^*(\lambda) \coloneqq \sup_{\lambda\in\mathbb{R}}\{\lambda t - \varphi_X(\lambda)\} \quad \text{and } \widetilde{\varphi}_X^*(\lambda) \coloneqq \sup_{\lambda\in\mathbb{R}}\{\lambda t - \widetilde{\varphi}_X(\lambda)\}$$

the *Legendre transform* (or *Legendre-Fenchel transform* or *convex conjugate*) of φ_X and $\widetilde{\varphi}_X$, respectively. We call a real random variable *sub-Gaussian* with parameter σ^2 (or shortly σ^2-sub-Gaussian) if $\widetilde{\varphi}_X(\lambda) \leq \frac{\sigma^2\lambda^2}{2}$ for all $\lambda \in \mathbb{R}$.

Finally, recall *Hölder's inequality.* For this consider $p, q \in [1, \infty]$ such that $\frac{1}{p} + \frac{1}{q} = 1$, where $\frac{1}{\infty} \coloneqq 0$. For two real random variables X and Y it holds that

$$\mathbb{E}[|XY|] \leq \mathbb{E}[|X|^p]^{\frac{1}{p}} \cdot \mathbb{E}[|Y|^q]^{\frac{1}{q}}.$$

Q1. (Logarithmic moment generating function of a Gaussian) Compute the centered and non centered log-moment generating function and their Legendre transforms of a Gaussian random variable $X \sim \mathcal{N}(\mu, \sigma^2)$ with mean μ and variance σ^2.

Q2. (Stability of sub-Gaussianity) Let X_1 and X_2 be sub-Gaussian with parameters σ_1^2 and σ_2^2, respectively.

(a) Show that αX_1 is sub-Gaussian with parameter $\alpha^2\sigma_1^2$ for a constant $\alpha \in \mathbb{R}$.

(b) If X_1 and X_2 are independent, show that X_1+X_2 is sub-Gaussian with parameter $\sigma_1^2+\sigma_2^2$.

(c) Show that in general (without assuming independence), the random variable $X_1 + X_2$ is sub-Gaussian with parameter $2(\sigma_1^2 + \sigma_2^2)$. Next, show that $X_1 + X_2$ is sub-Gaussian with parameter $(\sigma_1 + \sigma_2)^2$, which improves the result. *Hint:* Use Cauchy-Schwarz and Hoelder's inequality respectively.

Q3. (Properties of logarithmic moment generating functions) For this exercise you can assume $\mathbb{E}X = 0$, but the statements hold in general.

(a) *Convexity.* Show that φ_X and $\widetilde{\varphi}_X$ are convex functions. *Hint:* Hölder's inequality

(b) *Semi-continuity.* Show that φ_X and $\widetilde{\varphi}_X$ lower semi-continuous, i.e., if $\lambda_n \to \lambda$ for $n \to \infty$ then

$$\liminf_{n\to\infty} \varphi_X(\lambda_n) \geq \varphi_X(\lambda) \quad \text{and} \quad \liminf_{n\to\infty} \widetilde{\varphi}_X(\lambda_n) \geq \widetilde{\varphi}_X(\lambda).$$

Hint: Fatou's lemma

(c) *Smoothness.* Show that φ_X and $\widetilde{\varphi}_X$ are smooth, i.e., infinitely many times continuously differentiable, function on their domains. *Hint:* It suffices to show that e^{φ_X} and $e^{\widetilde{\varphi}}$ are smooth for which you can use the dominated convergence theorem.

(d) *Derivatives.* Show that

$$\varphi_X'(\lambda) = \frac{\mathbb{E}[Xe^{\lambda X}]}{M(\lambda)} \quad \text{and } \widetilde{\varphi}_X'(\lambda) = \frac{\mathbb{E}[Xe^{\lambda(X-\mathbb{E}X)}]}{\widetilde{M}(\lambda)},$$

where $M(\lambda) := \mathbb{E}[e^{\lambda X}] = e^{\varphi_X(\lambda)}$ and $\widetilde{M}(\lambda) := \mathbb{E}[e^{\lambda(X-\mathbb{E}X)}] = e^{\widetilde{\varphi}_X(\lambda)}$ denote the expontial and centered exponential moments of X.

(e) *Existence of moments.* Assume that $\varphi_X(\lambda) < \infty$ or $\widetilde{\varphi}_X(\lambda) < \infty$ for all $\lambda \in (-\varepsilon, \varepsilon)$ for some $\varepsilon > 0$. Show that all moments exist, i.e., $\mathbb{E}[|X|^k] < \infty$ for all $k \in \mathbb{N}$.

(f) *Cramer transform equals Legendre transform.* Show that

$$\widetilde{\varphi}_X^\star(\lambda) = \sup_{\lambda > 0}\{\lambda t - \widetilde{\varphi}_X(\lambda)\} = \sup_{\lambda \in \mathbb{R}}\{\lambda t - \widetilde{\varphi}_X(\lambda)\} = \widetilde{\varphi}_X^*(\lambda) \quad \text{for all } t \geq 0.$$