Concentration & Projections aléatoires

Exercice 1 – ***Variables sous-Gaussiennes***. Une variable aléatoire $X \in \mathbb{R}$ centrée est dite sous-Gaussienne avec variance σ^2 si $\forall s \in \mathbb{R}, \mathbb{E}[\exp(sX)] \leq \exp\left(\frac{\sigma^2 s^2}{2}\right)$.

1. Soit X une variable sous-Gaussienne avec variance σ^2, montrer que $\forall t > 0, \mathbb{P}[X > t] \leq \exp\left(-\frac{t^2}{2\sigma^2}\right)$.
2. Soit X une variable aléatoire telle que $\mathbb{E}(X) = 0$ et $X \in [a, b]$ presque sûrement. Montrer que X est sous-Gaussienne de variance $\frac{(b-a)^2}{4}$.
3. Soient $X_1, ..., X_n$ des variables aléatoires indépendantes telles que presque sûrement : $\forall i, X_i \in [a_i, b_i]$. Soit $\bar{X} = \frac{1}{n}\sum_{i=1}^n X_i$. Montrer que $\forall t > 0$:
$$\mathbb{P}(\bar{X} - \mathbb{E}(\bar{X}) > t) \leq \exp\left(-\frac{2n^2t^2}{\sum_{i=1}^n (b_i - a_i)^2}\right) \quad \text{(Inégalité de Hoeffding)}$$
4. Soient $X_1, ..., X_n$ des variables aléatoires sous-Gaussiennes avec variance σ^2. Montrer que :
$$\mathbb{E}[\max_{1\leq i\leq n} X_i] \leq \sigma\sqrt{2\log(n)} \quad \& \quad \mathbb{E}[\max_{1\leq i\leq n} |X_i|] \leq \sigma\sqrt{2\log(2n)}$$
ainsi que $\forall t > 0$:
$$\mathbb{P}(\max_{1\leq i\leq n} X_i > t) \leq ne^{-\frac{t^2}{2\sigma^2}} \quad \& \quad \mathbb{P}(\max_{1\leq i\leq n} |X_i| > t) \leq 2ne^{-\frac{t^2}{2\sigma^2}}$$

Exercice 2 – ***Johnson-Lindenstrauss***.

1. Soit $k \in \mathbb{N}$ et $\mathcal{X}_k^2 = \sum_{i=1}^k Z_i^2$ où $\forall i, Z_i \sim \mathcal{N}(0, 1)$. Soit $\epsilon \in [0, \frac{1}{2}]$, montrer que $\mathbb{P}(\mathcal{X}_k^2 \geq (1+\epsilon)k) \leq \exp\left(-\frac{k}{4}(\epsilon^2 - \epsilon^3)\right)$ et $\mathbb{P}(\mathcal{X}_k^2 \leq (1-\epsilon)k) \leq \exp\left(-\frac{k}{4}(\epsilon^2 - \epsilon^3)\right)$.
2. Soit $A \in \mathbb{R}^{k\times d}$ telle que $\forall (i, j), A_{ij} \sim \mathcal{N}(0, 1)$. Montrer que : $\forall x \in \mathbb{R}^d$, $\mathbb{E}\left[||\frac{1}{\sqrt{k}}Ax||^2\right] = \mathbb{E}\left[||x||^2\right]$.
3. Montrer que : $\mathbb{P}\left((1-\epsilon)||x||^2 \leq ||\frac{1}{\sqrt{k}}Ax||^2 \leq (1+\epsilon)||x||^2\right) \geq 1 - 2e^{-(\epsilon^2-\epsilon^3)\frac{k}{4}}$.
4. Soient $x_1, ..., x_n \in \mathbb{R}^d$, et $k = \frac{20\log n}{\epsilon^2}$, montrer qu'il existe une application Lipschitz $f : \mathbb{R}^d \longrightarrow \mathbb{R}^k$ telle que $\forall (i, j)$:
$$(1-\epsilon)||x_i - x_j||^2 \leq ||f(x_i) - f(x_j)||^2 \leq (1+\epsilon)||x_i - x_j||^2$$

Exercice 3 – ***Inégalité de Khintchine***. Soit $\{X_i\}_{i\in\mathbb{N}}$ une suite de variables aléatoires telles que : $\forall i, \mathbb{P}(X_i = +1) = \mathbb{P}(X_i = -1) = \frac{1}{2}$ (distribution de Rademacher).

1. Montrer que $\forall u \in \mathbb{R}, \mathbb{E}(e^{uX_1}) \leq e^{\frac{1}{2}u^2}$.
2. Soit $(a_i)_{1\leq i\leq n} \in \mathbb{R}^n$ tel que $\sum_{k=1}^n a_k^2 = 1$, soit $S = \sum_{k=1}^n a_k X_k$. Montrer que $\forall t > 0, \mathbb{P}(|S| > t) \leq 2e^{-\frac{1}{2}t^2}$.
3. Montrer que : $\forall p > 0, \left[\mathbb{E}\left(|\sum_{k=1}^n a_k X_k|^p\right)\right]^{\frac{1}{p}} \leq C_p\left(\sum_k a_k^2\right)^{\frac{1}{2}}$ pour $C_p > 0$.
4. Soit $S_n = \sum_{k=1}^n X_k$. Montrer que $\forall \rho > 1$:
$$\lim_{n\longrightarrow+\infty} \frac{S_n}{\sqrt{n(\log n)^\rho}} = 0 \quad \text{presque sûrement.}$$