TOPICS IN PROBABILITY. PART I: CONCENTRATION

Exercise sheet 4: Concentration inequalities and their applications

Exercise 1 (Azuma-Hoeffding inequality and generalization). *Let $(\mathcal{F}_n)_n$ be a filtration, $(\Delta_n)_n$ be random variables satisfying*

- *(martingale difference property) Δ_k is $\mathcal{F}_k$-measurable and $\mathbb{E}[\Delta_k|\mathcal{F}_{k-1}] = 0$ almost surely;*
- *(predictable bounds) A_k, B_k are $\mathcal{F}_{k-1}$-measurable and $A_k \leq \Delta_k \leq B_k$ a.s.*

Prove that $\sum_{k=1}^n \Delta_k$ is subgaussian with variance proxy $\frac{1}{4}\sum_{k=1}^n \|B_k - A_k\|_\infty^2$ and conclude that

$$\mathbb{P}\Big[\sum_{k\leq n} \Delta_k \geq t\Big] \leq \exp\left(-\frac{2t^2}{\sum_{k=1}^n \|B_k - A_k\|_\infty^2}\right).$$

In the case $|B_k - A_k|$ is not uniformly bounded, this bound is useless. So, prove the following more general form:

$$\mathbb{P}\Big[\sum_{k\leq n} \Delta_k \geq t, \sum_{k\leq n}(B_k - A_k)^2 \leq c^2\Big] \leq e^{-2t^2/c^2}.$$

If you need a hint, see the footnote[1].

Exercise 2. *(Concentration of the norm of vector with bounded entries) Let X_i's be i.i.d. uniformly bounded random variables with $\mathbb{E}[X_i^2] = 1$, and set $X = (X_1, \ldots, X_n)$. Apply McDiarmid's theorem or Azuma-Hoeffding in the most natural way to find a bound for $\mathbb{P}[\|X\| - \mathbb{E}[\|X\|] \geq t]$ of the form e^{-t^2/c_n}. What do you get for c_n? Does it depend on n?*

Actually, it is possible (with what you've learnt so far) to get such a bound with an absolute (independent of n) constant c by proceeding in a slightly different way. More precisely, prove that there exists $C > 0$ (independent of n!) such that for all $t \geq 0$,

$$\mathbb{P}[\|X\| - \sqrt{n} \geq t] \leq e^{-t^2/C^2};$$
$$\mathbb{P}[|\|X\| - \sqrt{n}| \geq t] \leq 2e^{-t^2/C^2}.$$

Hint: First look at variables $Y_i = X_i^2 - \mathbb{E}[X_i^2])$ and find a suitable bound for $\mathbb{P}[\frac{1}{n}\sum_{i\leq n} Y_i \geq t]$ (similarly for the absolute value of the sum); use the fact that for $z, \delta \geq 0$, $|z-1| \geq \delta$ implies that $|z^2 - 1| \geq \max(\delta, \delta^2)$ and conclude.

Show further that there exists an absolute constant $K > 0$ (independent of n) such that $0 \leq \sqrt{n} - \mathbb{E}[\|X\|] \leq K$ and conclude that for any $t \geq 2K$,

$$\mathbb{P}[|\|X\| - \mathbb{E}[\|X\|]| \geq t] \leq 2e^{-t^2/(4C^2)}.$$

[1] Consider $\lambda \sum_{k=1}^n \Delta_k - \frac{\lambda^2}{8}\sum_{k=1}^n (B_k - A_k)^2$.

Exercise 3 (Borell-TIS: concentration of supremum of Gaussian)**.**

Let $X \in \mathbb{R}^n$ be a Gaussian vector. Let $\sigma^2 := \sup_{i=1}^n \mathrm{Var}[X_i]$. Recall that then $Z := \sup_i X_i$ satisfies $\mathrm{Var}[Z] \le \sigma^2$ and prove that for some suitable $c > 0$ and all $t > 0$,

$$\mathbb{P}[|Z - \mathbb{E}[Z]| \ge t] \le 2e^{-\frac{t^2}{2c\sigma^2}}.$$

Remark: one could get the inequality with $c = 1$, but one would need a sharper Gaussian concentration inequality than the one proven in the lecture.

Exercise 4 (Empirical frequencies)**.** *Let $(X_i)_i$ be i.i.d. random variables with distribution μ on a measurable space E. Set $N_n(C) := \#\{k \le n : X_k \in C\}/n$. By the law of large numbers, $N_n(C) \approx \mu(C)$ for $n \gg 1$. We would like to control the deviation between the true probability $\mu(C)$ and its empirical average $N_n(C)$ uniformly over some countable class $\mathcal{C}$ of measurable subsets of E. Thus, define $Z_n \sup_{C \in \mathcal{C}} |N_n(C) - \mu(C)|$. Prove that*

$$\mathbb{P}[Z_n - \mathbb{E}[Z_n] \ge t] \le e^{-2nt^2}.$$

Exercise 5 (Maximal eigenvalue of symmetric matrix with Rademacher entries)**.** *Let M be an $n \times n$ symmetric matrix with i.i.d. Rademacher entries $(M_{ij})_{i \le j}$. We are interested in the maximal eigenvalue of the matrix, $\lambda_{\max}(M)$. Show that $\mathrm{Var}[\lambda_{\max}(M)] \le 16$ and*

$$\mathbb{P}[\lambda_{\max}(M) - \mathbb{E}[\lambda_{\max}(M)] \ge t] \le e^{-t^2/(4n(n+1))} \wedge \frac{16}{t^2}.$$

Hint: recall that $\lambda_{\max}(M) = \sup_{v \in \mathbb{R}^n : |v|=1} \langle v, Mv \rangle$, use this representation to find an estimate on $D_{ij}^- f(M)$ with $f(M) = \lambda_{\max}(M)$ for the variance bound, and on $D_{ij} f(M)$ for the concentration bound.

Remark: It is actually possible (using the so-called Talagrand's concentration inequality) to show that $\lambda_{\max}(M)$ is 16-subgaussian as you might expect from the variance bound.