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

$$\mathbb{P}\Big[\sum_{k\leq n} \Delta_k \geq t\Big] \leq \exp\left(-\frac{2t^2}{\sum_{k=1}^n \|B_k - A_k\|_\infty^2}\right). \tag{0.1}$$

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

$$\mathbb{P}\Big[\sum_{k\leq n} \Delta_k \geq t, \sum_{k\leq n}(B_k - A_k)^2 \leq c^2\Big] \leq e^{-2t^2/c^2}. \tag{0.2}$$

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

Proof. Note that Hoeffding lemma (Exercise 5 Sheet 3) applied to Δ_k conditionally on $\mathcal{F}_{k-1}$ implies that

$$\mathbb{E}\left[e^{\lambda\Delta_k}|\mathcal{F}_{k-1}\right] \leq e^{\lambda^2(B_k-A_k)^2/8} \leq e^{\lambda^2\|B_k-A_k\|_\infty^2/8}.$$

For any $1 \leq k \leq n$, we get

$$\mathbb{E}\left[e^{\lambda\sum_{j=1}^k \Delta_j}\right] \overset{(A)}{=} \mathbb{E}\left[e^{\lambda\sum_{j=1}^{k-1}\Delta_j}\mathbb{E}[e^{\lambda\Delta_k}|\mathcal{F}_{k-1}]\right] \leq e^{\lambda^2\|B_k-A_k\|_\infty^2/8}\mathbb{E}\left[e^{\lambda\sum_{j=1}^{k-1}\Delta_j}\right].$$

It follows by induction that $\mathbb{E}\left[e^{\lambda\sum_{j=1}^n \Delta_j}\right] \leq e^{\frac{\lambda^2}{8}\sum_{j=1}^n\|B_k-A_k\|^2}$. Hence, $\sum_{j=1}^n \Delta_j$ is by definition $\frac{1}{4}\sum_{j=1}^n\|A_j - B_j\|_\infty^2$. An application of exponential Markov inequality optimized in λ directly gives (0.1). To obtain (0.2), we apply exponential Markov inequality to $M_n = \lambda\sum_{k=1}^n \Delta_k - \frac{\lambda^2}{8}\sum_{k=1}^n(B_k - A_k)^2$ for some $\lambda > 0$. Using the same trick as in (A) combined with Hoeffding lemma, we get

$$\begin{aligned}\mathbb{P}\Big[\sum_{k\leq n}\Delta_k \geq t, \sum_{k\leq n}(B_k - A_k)^2 \leq c^2\Big] &\leq \mathbb{P}[M_n \geq \lambda t - \lambda^2/8c^2] \leq e^{-\lambda t+\lambda^2c^2/8}\mathbb{E}[e^{\lambda M_n}] \\ &\leq e^{-\lambda t+\lambda^2c^2/8}\mathbb{E}[e^{\lambda M_{n-1}}] \leq \cdots \leq e^{-\lambda t+\lambda^2c^2/8}.\end{aligned}$$

Optimizing in $\lambda > 0$ yields the result. $\square$

---

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

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 $t \geq 2K$,

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

Proof. Let us first apply McDiarmid's theorem in the most obvious way, namely to $f(x) = \|x\|$ restricted to the domain $[-c, c]^n$, where $c > 0$ is a constant which bounds $|X_i|$. Clearly, $\|D_i f\| \leq c$, and thus we get

$$\mathbb{P}[\|X\| - \mathbb{E}[\|X\|] \geq t] \leq e^{-2t^2/\sum_{k=1}^n c^2} = e^{-2t^2/(nc^2)},$$

which is a poor bound that strongly depends on the dimension of the vector X. Thus, let's proceed in a different way. Consider $Y_i = X_i^2 - \mathbb{E}[X_i^2]$ – these are again uniformly bounded random variables (by c^2 in absolute value). By Azuma-Hoeffding with $\Delta_k = Y_k$ and $\mathcal{F}_k = \sigma(X_i : i \leq k)$,

$$\mathbb{P}\left[\frac{1}{n}\sum_{k=1}^n Y_k \geq t\right] \leq e^{-2(tn)^2/(n4c^4)} = e^{-nt^2/(2c^4)}.$$

And by applying the same argument to $-Y_k$'s, we get (A): $\mathbb{P}\left[|\frac{1}{n}\sum_{k=1}^n Y_k| \geq t\right] \leq 2e^{-nt^2/(2c^4)}$. In particular. we have shown that

$$\mathbb{P}\left[\frac{1}{n}\|X\|^2 - 1 \geq t\right] \leq e^{-nt^2/(2c^4)}; \qquad \mathbb{P}\left[\left|\frac{1}{n}\|X\|^2 - 1\right| \geq t\right] \leq 2e^{-nt^2/(2c^4)}.$$

Using the fact (check it) that for $z, \delta \geq 0$, $|z - 1| \geq \delta$ implies $|z^2 - 1| \geq \delta \vee \delta^2$ (and clearly $z - 1 \geq \delta$ implies that $z^2 - 1 \geq \delta \vee \delta^2$), we obtain

$$\mathbb{P}[\|X\| - \sqrt{n} \geq t] \leq \mathbb{P}\left[\frac{1}{n}\|X\|^2 - 1 \geq \frac{t}{\sqrt{n}} \vee \frac{t^2}{n}\right] \leq e^{-t^2/(2c^4)},$$

and analogously for $|\|X\| - \sqrt{n}|$. This completes the first part of the exercise. As for the second, first note that by Cauchy-Schwarz, $\mathbb{E}[\|X\|] \leq \sqrt{\mathbb{E}[\|X\|^2]} = \sqrt{n}$. Furthermore, recall that you have shown in the lecture that $\mathrm{Var}[\|X\|] \leq (2c)^2$ (since the norm is 1-Lipschitz and $-c \leq X_i \leq c$ are independent). The latter implies that $\mathbb{E}[\|X\|] - \sqrt{n} \geq \sqrt{n - (2c)^2} - \sqrt{n} \geq -(2c)^2/\sqrt{n}$ (for all n sufficiently large, for n uniformly bounded, the statement is anyway clear with a potentially slightly different (uniform) constant). The desired result follows from the observations that $|\|X\|-\mathbb{E}[\|X\|] \geq | \leq |\|X\|-\sqrt{n}|+\sqrt{n}-\mathbb{E}[\|X\|]$ and $\sqrt{n}-\mathbb{E}[\|X\|] \leq t/2$. Note that with the above more precise bound we actually get that for any $t > 0$,

$$\limsup_{n\to\infty} \mathbb{P}[|\|X\| - \mathbb{E}[\|X\|]| \geq t] \vee \mathbb{P}[|\|X\| - \sqrt{n}| \geq t] \leq e^{-t^2/(4C^2)}$$

□

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* $\mathrm{Var}[Z] \leq \sigma^2$ *and prove that for some suitable* $c > 0$ *and all* $t > 0$,

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

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.*

Proof. Let A be the square root of the positive-definite covariance matrix of X and let N be a standard Gaussian vector. Then $X \overset{\text{law}}{=} AN + \mu$ with $\mu = \mathbb{E}[X]$. Define $f : \mathbb{R}^n \to \mathbb{R}$ as $f(x) = \max_{i=1}^n (Ax + \mu)_i$. In particular, $f(N)$ has the same distribution as Z. We saw in the proof of Exercise 2 Sheet 3 that f is is σ-Lipschitz. Hence, we can apply Gaussian concentration inequality (proven in the lecture) and obtain

$$\mathbb{P}[|Z - \mathbb{E}[Z]| \geq t] = \mathbb{P}[|f(N) - \mathbb{E}[f(N)]| \geq t] \leq 2e^{-\frac{t^2}{2c\sigma^2}}$$

for a suitable constant $c > 0$. □

Exercise 4 (Empirical frequencies)**.** *Let* $(X_i)_i$ *be i.i.d. random variables with distribution* μ *on a measurable space* E*. Set* $N_n(C) := \#\{k \leq n : X_k \in C\}/n$*. 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*

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

Proof. Let us look at Z_n as a function f of variables $X_1, \dots, X_n$. For $1 \leq k \leq n$ and any $z, w \in E$,

$$\begin{aligned} &|f(X_1, \dots, X_{k-1}, z, X_{k+1}, \dots, X_n) - f(X_1, \dots, X_{k-1}, w, X_{k+1}, \dots, X_n)| \\ &\qquad \leq \sup_{C\in\mathcal{C}} |N_n(C, (X_1, \dots, z, \dots, X_n)) - N_n(C, (X_1, \dots, w, \dots, X_n))| \leq \frac{1}{n}. \end{aligned}$$

Therefore, $\|D_k Z_n\|_\infty \leq 1/n$, and by McDiarmid's theorem,

$$\mathbb{P}[Z_n - \mathbb{E}[Z_n] \geq t] \leq e^{-2t^2/(\sum_{k=1}^n 1/n^2)} = e^{-2nt^2}.$$

□

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* $(M_{ij})_{i\leq j}$*. We are interested in the maximal eigenvalue of the matrix,* $\lambda_{\max}(M)$*. Show that* $\mathrm{Var}[\lambda_{\max}(M)] \leq 16$ *and*

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

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.*

Proof. Recall that $\lambda_{\max}(M) = \sup_{v\in\mathbb{R}^n:|v|=1}\langle v, Mv\rangle$. Denote by $v_{\max}(M)$ an eigenvector of M with eigenvalue $\lambda_{\max}(M)$. Let $i \le j$ be fixed. Choose a symmetric matrix M_- such that

$$\lambda_{\max}(M_-) = \inf_{M_{ij}=\pm 1} \lambda_{\max}(M).$$

In particular, $(M_-)_{kl} = M_{kl}$ unless $\{k, l\} = \{i, j\}$, and thus,

$$\begin{aligned} D_{ij}^-\lambda_{\max}(M) &= \lambda_{\max}(M) - \lambda_{\max}(M_-) \le \langle v_{\max}(M), (M - M_-)v_{\max}(M)\rangle \\ &\le 2v_{\max}(M)_i v_{\max}(M)_j(M_{ij} - (M_-)_{ij}) \le 4|v_{\max}(M)_i||v_{\max}(M)_j| \end{aligned}$$

since both $M_{ij}, (M_-)_{ij}$ take values ± 1. By a corollary of Efron-Stein, we thus get

$$\mathrm{Var}[\lambda_{\max}(M)] \le \mathbb{E}\left[\sum_{i\le j} 16|v_{\max}(M)_i|^2|v_{\max}(M)_j|^2\right] \le 16.$$

Here we have used that $\sum_i |v_{\max}(M)_i|^2 = 1$.

By a fully analogous argument, $D_{ij}\lambda_{\max}(M) \le 4|v_{\max}(M_+^{(ij)})_i||v_{\max}(M_+^{(ij)})_j|$, where for each $i \le j$, $M_+^{(ij)}$ is a symmetric matrix $\lambda_{\max}(M_+^{(ij)}) = \sup_{M_{ij}=\pm 1} \lambda_{\max}(M)$. In particular, $\|D_{ij}\lambda_{\max}(M)\|_\infty \le 4$, since $|v| = 1$, and so each of its coordinates v_i is at most of absolute value one. Therefore, by McDiarmid's theorem,

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

The remaining bound follows from Chebychev inequality and the above estimate on the variance. $\square$