# TOPICS IN PROBABILITY. PART I: CONCENTRATION

## Exercise sheet 2: Variance bounds

### 1. Warm up

**Exercise 1.** *Let $X$ be any (possibly vector-valued) random variable, $f$ be a measurable real-valued function on the state space of $X$. Show that*

$$\mathrm{Var}[f(X)] \le \frac{1}{4}(\sup f - \inf f)^2 \qquad \textit{and} \qquad \mathrm{Var}[f(X)] \le \mathbb{E}[(f(X) - \inf f)^2].$$

**Exercise 2** (Square root of Chi-squared distribution)**.**

*Let $Z$ be a non-negative random variable such that $Z^2$ is chi-squared distributed with $D$ degrees of freedom. Prove that*

$$\sqrt{D} - 1 \le \mathbb{E}[Z] \le \sqrt{D}.$$

*Hint: Recall how chi-squared distributed random variable is related to a Gaussian vector.*

### 2. An alternative proof of Efron-Stein inequality

**Exercise 3** (Proof using martingales)**.** *Let $X_1, \dots, X_n$ be independent, $f : \mathbb{R}^n \to \mathbb{R}$ such that $\mathbb{E}[f^2] < \infty$. Prove that*

$$\mathrm{Var}[f] \le \sum_{i=1}^{n} \mathbb{E}\big[\mathrm{Var}[f|(X_j)_{j\neq i}]\big]$$

*proceeding as follows:*

1. *Consider $S_m = \mathbb{E}[f|X_1, \dots, X_m]$ for $m \ge 1$ and $S_0 = \mathbb{E}[f]$ and show that it is a martingale satisfying $S_n = f$;*
2. *Prove that for a square-integrable martingale $S_n$ with $S_0 = \mathbb{E}[S_n]$,*
$$\mathrm{Var}[S_n] = \sum_{i=1}^{n} \mathbb{E}[(S_i - S_{i-1})^2];$$
3. *Show that*
$$S_i - S_{i-1} = \mathbb{E}\big[f - \mathbb{E}[f|(X_j)_{j\neq i}]\big|X_1, \dots, X_i\big];$$
4. *Conclude.*

### 3. Applications of Efron-Stein inequality

**Exercise 4** (Among Lipschitz functions the sum has the largest variance)**.**

*Consider the class $\mathcal{F}$ of functions $f : \mathbb{R}^n \to \mathbb{R}$ that are Lipschitz w.r.t. $\ell^1$ distance, i.e., if $x = (x_1, \dots, x_n) \in \mathbb{R}^n$ and $y = (y_1, \dots, y_n) \in \mathbb{R}^n$, then $|f(x) - f(y)| \le \sum_{i=1}^n |x_i - y_i|$. Let $X = (X_1, \dots, X_n)$ be a vector of independent variables of finite variance. Use the Efron-Stein inequality to show that the maximal value of $\mathrm{Var}[f(X)]$ over $f \in \mathcal{F}$ is attained by the function $f(x) = \sum_{i=1}^n x_i$.*

**Exercise 5** (Rademacher processes).

*Let $\varepsilon_1, \ldots, \varepsilon_n$ be independent Rademacher variables, i.e., Bernoulli random variables taking values $\pm 1$ with probability $1/2$, let $T \subset \mathbb{R}^n$. First check the following easy identity:*

$$\sup_{t\in T} \operatorname{Var}\left[\sum_{k=1}^{n} \varepsilon_k t_k\right] = \sup_{t\in T} \sum_{k=1}^{n} t_k^2.$$

*Now prove that*

$$\operatorname{Var}\left[\sup_{t\in T} \sum_{k=1}^{n} \varepsilon_k t_k\right] \le 4 \sup_{t\in T} \sum_{k=1}^{n} t_k^2.$$

*Thus, taking the supremum inside the variance costs at most a constant factor.*
*Remark: one can get constant 2 instead of 4 in the above inequality.*

**Exercise 6** (Triangles in Erdös-Rényi graph).

*Let $Z$ be the number of triangles in a random graph $G \sim \mathcal{G}(n,p)$, where $\mathcal{G}(n,p)$ denote the Erdös-Rényi model, which is constructed on a set of $n$ vertices by connecting every pair of distinct vertices independently with probability $p$ (or alternatively delete edges independently from the complete graph on $n$ vertices with probability $1-p$). A triangle is a complete three-vertex subgraph. Calculate the variance of $Z$ and compare it with what you get by using the Efron-Stein inequality to estimate it.*