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*

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

Proof. Note that $\mathrm{Var}[f(X)] = \mathrm{Var}[f(X) - a] \leq \mathbb{E}[(f(X) - a)^2]$ for all $a \in \mathbb{R}$. The second inequality follows by choosing $a = \inf f$. As for the first inequality, consider $a = \frac{\inf f + \sup f}{2}$, then almost surely $|f(X) - a| \leq \frac{\sup f - \inf f}{2}$. □

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

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

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

Proof. Recall that Z^2 has the same law as $|X|^2 = \sum_{i=1}^{D} X_i^2$, where $X = (X_1, \ldots, X_D)$ is a Gaussian vector with standard normal iid coordinates. By Jensen's inequality we get

$$\mathbb{E}[Z] = \mathbb{E}[\sqrt{Z^2}] \leq \sqrt{\mathbb{E}[Z^2]} = \sqrt{\sum_{i=1}^{D} \mathbb{E}[X_i^2]} = \sqrt{D}.$$

For the lower bound note that by Hölder's inequality we get,

$$\mathbb{E}[Z^2] = \mathbb{E}[Z^{2/3+4/3}] \leq \mathbb{E}[Z]^{2/3}\mathbb{E}[Z^4]^{1/3}.$$

In particular, since $\mathbb{E}[Z^2] = D$ and $\mathbb{E}[Z^4] = \sum_{i=1}^{D} \mathbb{E}[X_i^4] + 2\sum_{j<i} \mathbb{E}[X_i^2]\mathbb{E}[X_j^2] = 3D + D(D-1) = D(D+2)$, we get

$$\mathbb{E}[Z] \geq \sqrt{\frac{\mathbb{E}[Z^2]^3}{\mathbb{E}[Z^4]}} = \sqrt{\frac{D^3}{D(D+2)}} \geq \sqrt{D} - 1.$$

The latter inequality is just a computation. □

2. An alternative proof of Efron-Stein inequality

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

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

proceeding as follows:

(1) *Consider* $S_m = \mathbb{E}[f|X_1, \dots, X_m]$ *for* $m \geq 1$ *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.*

Proof. Let $S_0 = \mathbb{E}[f]$ and $S_m = \mathbb{E}[f|X_1, \dots, X_m]$ for $1 \leq m \leq n$. By tower-property of conditional expectation and measurability of $f = f(X_1, \dots, X_n)$ w.r.t. $\mathcal{F}_n \coloneqq \sigma(X_1, \dots, X_n)$, $(S_m)_{m=0}^n$ is clearly an $(\mathcal{F}_m)_m$-martingale satisfying $S_n = f$. Moreover, since f is square-integrable, so is S_m for any m. Note that if $1 \leq i < j \leq n$, by "taking out what's known" property of conditional expectation, we get

$$\mathbb{E}[(S_i - S_{i-1})(S_j - S_{j-1})] = \mathbb{E}[(S_i - S_{i-1})\underbrace{\mathbb{E}[S_j - S_{j-1}|\mathcal{F}_{j-1}]}_{=0}] = 0.$$

Therefore,

$$\mathrm{Var}[S_n] = \mathbb{E}[(S_n - \mathbb{E}[S_n])^2] = \sum_{1\leq i,j\leq n} \mathbb{E}[(S_i - S_{i-1})(S_j - S_{j-1})] = \sum_{i=1}^{n} \mathbb{E}[(S_i - S_{i-1})^2].$$

Furthermore, since S_{i-1} is $\mathcal{F}_i$-measurable

$$S_i - S_{i-1} = \mathbb{E}[f - S_{i-1}|X_1, \dots, X_i];$$

but by tower property we also have,

$$S_{i-1} = \mathbb{E}[f|X_1, \dots, X_{i-1}] = \mathbb{E}\big[\mathbb{E}[f|X_1, \dots, X_{i-1}, X_{i+1}, \dots, X_n]\big|X_1, \dots, X_{i-1}\big],$$

which is independent of X_i since $(X_j)_j$ are i.i.d, and thus,

$$S_{i-1} = \mathbb{E}\big[\mathbb{E}[f|X_1, \dots, X_{i-1}, X_{i+1}, \dots, X_n]\big|X_1, \dots, X_i\big]$$

almost surely. Hence, as desired $S_i - S_{i-1} = \mathbb{E}\big[f - \mathbb{E}[f|(X_j)_{j\neq i}]\big|X_1, \dots, X_i\big]$. We are ready to conclude:

$$\begin{aligned}
\mathrm{Var}[f] &= \sum_{i=1}^{n} \mathbb{E}\Big[\mathbb{E}\big[f - \mathbb{E}[f|(X_j)_{j\neq i}]\big|X_1, \dots, X_i\big]^2\Big] \\
&\overset{\text{Jensen}}{\leq} \sum_{i=1}^{n} \mathbb{E}\Big[\mathbb{E}\big[(f - \mathbb{E}[f|(X_j)_{j\neq i}])^2\big|X_1, \dots, X_i\big]\Big] \\
&= \sum_{i=1}^{n} \mathbb{E}\Big[\mathbb{E}\big[(f - \mathbb{E}[f|(X_j)_{j\neq i}])^2\big]\Big] \\
&= \sum_{i=1}^{n} \mathbb{E}\Big[\underbrace{\mathbb{E}\big[(f - \mathbb{E}[f|(X_j)_{j\neq i}])^2\big|(X_j)_{j\neq i}\big]}_{=\mathrm{Var}[f|(X_j)_{j\neq i}]}\Big].
\end{aligned}$$

□

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. ℓ^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 $f(x) = \sum_{i=1}^n x_i$.

Proof. By Efron-Stein inequality for any $f \in \mathcal{F}$,

$$\begin{aligned}\mathrm{Var}[f(X)] &\le \frac{1}{2}\sum_{i=1}^n \mathbb{E}\left[(f(X) - (f(X))_i')^2\right] \le \frac{1}{2}\sum_{i=1}^n \mathbb{E}\left[|X_i - X_i'|^2\right] \\ &= \sum_{i=1}^n (\mathbb{E}[X_i^2] - \mathbb{E}[X_i]^2) = \sum_{i=1}^n \mathrm{Var}[X_i] = \mathrm{Var}\left[\sum_{i=1}^n X_i\right] = \mathrm{Var}[f_0(X)],\end{aligned}$$

where X' is an independent copy of X and $(f(X))_i' = f(X_1, \dots, X_{i-1}, X_i', X_{i+1}, \dots, X_n)$. The upper estimate finishes the proof. □

Exercise 5 (Rademacher processes)**.**

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

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

Now prove that

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

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.

Proof. The first identity follows from the fact that the variance of sum of independent random variables is the sum of variances of these variables, and $\mathrm{Var}[\varepsilon_k t_k] = t_k^2$.

To prove the second bound, let us set $Z := \sup_{t\in T} \sum_{k=1}^n \varepsilon_k t_k$. Let $\varepsilon_1', \dots, \varepsilon_n'$ be independent copies of $\varepsilon_1, \dots, \varepsilon_n$ and $Z_i' := \sup_{t\in T}\left[\sum_{j:j\neq i}^n \varepsilon_j t_j + \varepsilon_i' t_i\right]$. Let t^* be a random vector such that $Z = \sum_{i=1}^n \varepsilon_i t_i^*$. Assume that $\sup_{t\in T}\sum_{k=1}^n t_k^2$ is finite, since otherwise there is nothing to show. Note also that this implies that T has to be bounded and hence Z, Z' are well-defined square-integrable variables. Then for all $i = 1, \dots, n$,

$$Z - Z_i' \le (\varepsilon_i - \varepsilon_i')t_i^* \qquad \text{and so} \qquad (Z - Z_i')_+^2 \le (\varepsilon_i - \varepsilon_i')^2 (t_i^*)^2,$$

where the first inequality follows since for any two real-valued functions f, g on common domain, $(f - g)(u) \ge f(u) - \sup g$. The second inequality follows since if $f \le a$ and x is such that $f_+(x) = 0$, then clearly $f_+(x)^2 \le 0 \le a^2(x)$, if x is such that $f_+(x) \neq 0$, then

$f(x) = f_+(x) \geq 0$ and so $0 \leq f(x) \leq a(x)$ and $f_+^2(x) = f^2(x) \leq a^2(x)$. Using that ε_i' is independent of $\varepsilon_1, \ldots, \varepsilon_n$ we obtain by properties of conditional expectation,

$$\mathbb{E}\left[(Z - Z_i')_+^2\right] \leq \mathbb{E}\left[\mathbb{E}[(\varepsilon_i - \varepsilon_i')^2 (t_i^*)^2 | \varepsilon_1, \ldots, \varepsilon_n]\right] = 2\mathbb{E}[(t_i^*)^2].$$

By the Efron-Stein inequality we therefore get,

$$\mathrm{Var}[Z] \leq 2\mathbb{E}\left[\sum_{i=1}^n t_i^*\right] \leq 2 \sup_{t \in \overline{T}} \sum_{k=1}^n t_k^2 = 2 \sup_{t \in T} \sum_{k=1}^n t_k^2.$$

□

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.

Proof. Let us first compute the variance explicitly. For this we denote the set of vertices $V = \{1, \ldots, n\}$, and introduce the random variables for $i < j < k$: $X_{i,j,k} = \mathbf{1}_{\{i,j,k\} \text{ is a triangle}}$. In particular, $X_{i,j,k}$ is Bernoulli distributed with probability p^3 (probability that all three edges are present). Let Z be the number of triangles in G, then $Z = \sum_{i<j<k}^n X_{i,j,k}$ and $Z^2 = \sum_{i<j<k}^n \sum_{p<q<r}^n X_{i,j,k} X_{p,q,r}$. Hence,

$$\mathbb{E}[Z] = \sum_{i<j<k}^n p^3 = \binom{n}{3} p^3;$$

$$\mathbb{E}[Z^2] = \sum_{i<j<k}^n \sum_{p<q<r}^n \mathbb{E}[X_{i,j,k} X_{p,q,r}].$$

To compute the variance we need expressions for $\mathbb{E}[X_{i,j,k} X_{p,q,r}]$ with $i < j < k, p < q < r$. Let us consider possible cases:

- $\{i,j,k\} = \{p,q,r\}$ so that there must be three distinct edges present. Thus, $\mathbb{E}[X_{i,j,k} X_{p,q,r}] = p^3$. There are exactly $\binom{n}{3}$ elements of such form in the expression for $\mathbb{E}[N^2]$.
- $\{i,j,k\}$ and $\{p,q,r\}$ have two common elements meaning that there must be 5 distinct edges present. Thus, $\mathbb{E}[X_{i,j,k} X_{p,q,r}] = p^5$. There are $2\binom{n}{4}\binom{4}{2} = 12\binom{n}{4}$ elements of this type.
- $\{i,j,k\}$ and $\{p,q,r\}$ have one common element so that there must be 6 distinct edges present. Hence, $\mathbb{E}[X_{i,j,k} X_{p,q,r}] = p^6$. There are $\binom{n}{5}\binom{5}{1}\binom{4}{2} = 30\binom{n}{5}$ such elements.
- $\{i,j,k\}$ and $\{p,q,r\}$ have no common elements, meaning that there must be 6 distinct edges present. Hence, $\mathbb{E}[X_{i,j,k} X_{p,q,r}] = p^6$. There are $\binom{n}{6}\binom{6}{3} = 20\binom{n}{6}$ such elements.

Therefore,

$$\mathrm{Var}[Z] = \mathbb{E}[Z^2] - \mathbb{E}[Z]^2 = \binom{n}{3} p^3 + 12\binom{n}{4} p^5 + 30\binom{n}{5} p^6 + 20\binom{n}{6} p^6 - \binom{n}{3}^2 p^6.$$

To use the Efron-Stein estimate we need to slightly modify the variables we are working with. We write $m = \binom{n}{2}$ for the number of edges in the complete graph, and let $Y_1, \ldots, Y_m$ be iid

Bernoulli variables with probability p (edge is present or not). Let T be a set of all triples of edges forming a triangle. Then $Z = \sum_{t=\{p,q,r\}\in T} \prod_{i\in t} X_i$. Let us write $I_t \coloneqq \prod_{i\in t} X_i$ (equal in law to I_t^2). So we get,

$$\begin{aligned}\mathbb{E}[Z] &= \sum_{t\in T} \mathbb{E}[I_t];\\ \mathrm{Var}[Z] &= \sum_{t\in T}(\mathbb{E}[I_t^2] - \mathbb{E}[I_t]^2) + \sum_{s\neq t\in T: s\cap t\neq\emptyset} (\mathbb{E}[I_t I_s] - \mathbb{E}[I_t]\mathbb{E}[I_s])\\ &= \sum_{t\in T}(\mathbb{E}[I_t] - \mathbb{E}[I_t]^2) + \sum_{s\neq t\in T: s\cap t\neq\emptyset} (\mathbb{E}[I_t I_s] - \mathbb{E}[I_t]\mathbb{E}[I_s]).\end{aligned}$$

Define $\delta = (\sum_{s,t\in T: s\cap t\neq\emptyset} \mathbb{E}[I_t I_s])/\mathbb{E}[Z]$. A direct upper bound for the variance is

$$\mathrm{Var}[Z] \leq \sum_{t\in T} \mathbb{E}[I_t] + \sum_{s,t\in T: s\cap t\neq\emptyset} \mathbb{E}[I_t I_s] = (1+\delta)\mathbb{E}[Z].$$

On the other hand, by Efron-Stein we get

$$\begin{aligned}\mathrm{Var}[Z] &\leq \sum_{i=1}^m \mathbb{E}[(Z - Z_i')_-^2] \leq \sum_{i=1}^m \mathbb{E}\left[\left(\sum_{t: i\in t} I_t\right)^2\right] = \sum_i \mathbb{E}\left[\sum_{t: i\in t} I_t + \sum_{s\neq t: i\in s\cap t} I_t I_s\right]\\ &= \sum_t \sum_{i\in t} \mathbb{E}[I_t] + \sum_{t\neq s} \sum_{i\in t\cap s} \mathbb{E}[I_t I_s] \leq 3(1+\delta)\mathbb{E}[Z]\end{aligned}$$

where Z_k' depends on $X_1, \ldots, X_{k-1}, X_k', X_{k+1}, \ldots, X_m$ as Z on $X_1, \ldots, X_m$ with X_k''s are independent copies of X_k's. The last inequality comes from the fact that the cardinality of $t \in T$ is 3.

In particular, the Efron-Stein bound is 3 times larger than the direct bound on the variance.

Remark: note that the latter approach works for more general sets T with elements of bounded cardinality. □