TOPICS IN PROBABILITY. PART I: CONCENTRATION

Exercise sheet 3: Efron-Stein and Poincaré inequalities, Subgaussianity

1. Efron-Stein and Poincaré inequalities

Exercise 1 (Extending Gaussian Poincaré to C^1 functions)**.**

Let $f : \mathbb{R} \to \mathbb{R}$ be any continuously differentiable function and X be a standard normal random variable. Show that

$$\mathrm{Var}[f(X)] \leq \mathbb{E}[(f'(X))^2].$$

Proof. We can assume throughout the proof that $\mathbb{E}[(f'(X))^2] < \infty$ since otherwise there is nothing to show. Let η_ε be a standard mollifier and ψ_n be a smooth cut-off function with uniformly (independent of n) bounded gradient and taking values 1 in $[-n, n]$ and 0 in $[-n-1, n+1]^c$. Let $f_\varepsilon = f * \eta_\varepsilon$ and $f_{\varepsilon,n} = \psi_n f_\varepsilon$. Since $f \in C^1$, $(f_\varepsilon)' = (f')_\varepsilon$, f_ε and $(f')_\varepsilon$ converge uniformly over compacts towards f and f', respectively. The latter holds for $f_{\varepsilon,n}$ and $(f')_{\varepsilon,n}$. Note that these functions are C_c^∞. Hence, with the result from the lecture,

$$\mathrm{Var}[f_{\varepsilon,n}(X)] \leq \mathbb{E}\left[[(f_{\varepsilon,n})'(X)]^2\right] \leq \mathbb{E}\left[[((f')_\varepsilon)_n(X)]^2\right] + C\mathbb{E}\left[f_\varepsilon^2(X)\mathbf{1}_{n\leq|X|\leq n+1}\right].$$

By taking limit in ε, we get

$$\mathrm{Var}[(f)_n(X)] \leq \mathbb{E}\left[[(f')_n(X)]^2\right] + C\mathbb{E}\left[f^2(X)\mathbf{1}_{n\leq|X|\leq n+1}\right].$$

Since by assumption $\mathbb{E}[(f'(X))^2] < \infty$, $\mathbb{E}\left[[(f')_n(X)]^2\right]$ converges to $\mathbb{E}\left[(f'(X))^2\right]$ as $n \to \infty$. If $f(X) \in L^2(\mathbb{P})$, we thus can conclude. Otherwise we might distinguish between:

- $f^2(x)e^{-x^2/2} \to 0$ as $|x| \to \infty$;
- $\liminf_{x\to\infty} f^2(x)e^{-x^2/2} > 0$ for at least one of $+\infty$ or $-\infty$.

In the first case, $\mathbb{E}\left[f^2(X)\mathbf{1}_{|X|\in[n,n+1]}\right] \leq \max_{[-n-1,-n]\cup[n,n+1]}\left(f^2(x)e^{-x^2/2}\right) \to 0$ as $n \to \infty$. Therefore, if $\mathrm{Var}[f(X)]$ is well-defined (situation of our interest), i.e. that no $\infty - \infty$ occurs, by taking limit in n, we obtain

$$\mathrm{Var}[f(X)] \leq \mathbb{E}[(f'(X))^2].$$

In the case $\liminf_{x\to\infty} f^2(x)e^{-x^2/2} > 0$, we have that $f(x) = \mathcal{O}(e^{x^2/4})$ as $x \to \infty$. But this would imply that $(f'(x))^2 = \mathcal{O}(x^2e^{x^2/2})$, which is not integrable with respect to Gaussian measure. Hence, contradiction to $\mathbb{E}[(f'(X))^2] < \infty$. □

Exercise 2 (Gaussian Poincaré for Lipschitz functions is better than direct Efron-Stein)**.**

Let X be a standard Gaussian vector, f be L-Lipschitz function and set $Z = f(X)$. Compare the bounds for $\mathrm{Var}[Z]$ *obtained using Gaussian Poincaré inequality to the ones that you get by applying Efron-Stein inequality (in the most obvious way). What important feature does Poincaré inequality have compared to Efron-Stein in this situation?*

(1) Prove that for any Gaussian vector Y,

$$\mathrm{Var}[\max_{i=1}^{n} Y_i] \leq \max_{i=1}^{n} \mathrm{Var}[Y_i]$$

(2) Recall Exercise 2 Sheet 2. Use Gaussian Poincaré inequality to prove that

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

where Z be a non-negative random variable such that Z^2 has a chi-squared distribution with D degrees of freedom.

Can you get these bounds using Efron-Stein inequality?

Proof. Since f is L-Lipschitz, it is almost everywhere differentiable with $\|\nabla f\| \le L$. Via approximation argument as in the lecture, we obtain by Poincaré inequality that

$$\mathrm{Var}[Z] \le \mathbb{E}[\|\nabla f(X)\|^2] \le L^2.$$

On the other hand, a direct bound obtained by the Efron-Stein inequality is

$$\mathrm{Var}[Z] \le \frac{1}{2}\sum_{i=1}^{d} \mathbb{E}\left[(Z - Z_i')^2\right] \overset{\text{Lipschitz}}{\le} \frac{1}{2}L^2 \sum_{i=1}^{d} \mathbb{E}\left[(X_i - X_i')^2\right] = L^2 d.$$

In particular, we see that the most naive application of Efron-Stein inequality together with Lipschitz property gives a bound which depends on the dimension of the Gaussian vector. Gaussian Poincaré inequality, on contrary, provides a uniform bound independent of the dimension.

As for (1), there exists A such that AA^T is the covariance matrix of Y. In law, $\max_i Y_i = f(X)$, where $f(x) = \max_i (Ax)_i$. Note that f is L-Lipschitz with $L^2 = \max_i \sum_j a_{ij}^2 = \max_i \mathrm{Var}[Y_i]$. Indeed, $|f(x) - f(y)| \le \max_i |(A(x-y))_i| \le \max_i \sqrt{\sum_j a_{ij}^2}|x - y|$ by Cauchy-Schwarz. The first part of the exercise yields the desired result.

As for (2), recall that Z is distributed as absolute value of D-dimensional standard Gaussian vector, $|X|$. Note that $|\cdot|$ is 1-Lipschitz. Therefore, we get

$$\mathbb{E}[Z] = \sqrt{\mathbb{E}[Z^2] - \mathrm{Var}[Z]} = \sqrt{D - \mathrm{Var}[Z]} \ge \sqrt{D-1} \ge \sqrt{D} - 1.$$

Since Efron-Stein gives a bound on $\mathrm{Var}[Z]$ proportional to D, we won't be able to obtain the desired lower bound for $\mathbb{E}[Z]$ this way. □

Exercise 3 (Poisson Poincaré inequality).

Let $f : \mathbb{N}_0 \to \mathbb{R}$ be a real-valued function defined on non-negative integers, denote its discrete derivative as $Df(x) = f(x+1) - f(x)$. Let X be a Poisson random variable with intensity μ. Prove that

$$\mathrm{Var}[f(X)] \le \mu \mathbb{E}[(Df(X))^2].$$

Hint: infinite divisibility of Poisson distribution and Efron-Stein inequality might be useful.

Proof. Recall the Poisson limit theorem, which tells us that if $(X_i^n)_{i=1}^n, n \in \mathbb{N}$ are independent families of iid Bernoulli distributed random variables with parameter $p_n = \mu/n$, then $S_n :=$

$\sum_{i=1}^n X_i^n$ converges in distribution to X. Let us now compute $\mathrm{Var}[f(S_n)|X_j^n : j \neq i]$:

$$\begin{aligned}
\mathrm{Var}[f(S_n)|X_j^n : j \neq i] &= \mathbb{E}\left[\left(f(S_n) - \mathbb{E}[f(S_n)|X_j^n : j \neq i]\right)^2 \left|X_j^n : j \neq i\right.\right] \\
&= \mathbb{E}\left[\left(f(S_n) - \frac{\mu}{n} f(S_n - X_i + 1) - \left(1 - \frac{\mu}{n}\right) f(S_n - X_i)\right)^2 \left|X_j^n : j \neq i\right.\right] \\
&= \left(\frac{\mu}{n}\left(1 - \frac{\mu}{n}\right)^2 + \left(1 - \frac{\mu}{n}\right)\left(\frac{\mu}{n}\right)^2\right)(f(S_n - X_i + 1) - f(S_n - X_i))^2 \\
&= \frac{\mu}{n}\left(1 - \frac{\mu}{n}\right)(f(S_n - X_i + 1) - f(S_n - X_i))^2.
\end{aligned}$$

Therefore, by Efron-Stein inequality we obtain,

$$\begin{aligned}
\mathrm{Var}[f(S_n)] &\leq \frac{\mu}{n}\left(1 - \frac{\mu}{n}\right)\sum_i \mathbb{E}\left[(f(S_n - X_i + 1) - f(S_n - X_i))^2\right] \\
&= \mu\left(1 - \frac{\mu}{n}\right)\mathbb{E}\left[(f(S_{n-1} + 1) - f(S_{n-1}))^2\right] = \mu\left(1 - \frac{\mu}{n}\right)\mathbb{E}\left[(Df(S_{n-1}))^2\right].
\end{aligned}$$

Note that the rhs converges to $\mu\mathbb{E}[(Df(X))^2]$ as $n \to \infty$. $\square$

2. Subgaussianity

Exercise 4 (Subgaussian properties)**.**

Let X be a random variable. Show that the following properties are equivalent so that the parameters $C_i > 0$ appearing in the properties below differ from each other by an absolute constant factor, meaning that there exists $C > 0$ such that property i implies property j with parameter $C_j \leq CC_i$ for all $i, j = 1, \ldots, 5$

(1) The tails of X satisfy

$$\mathbb{P}[|X| \geq t] \leq 2e^{-t^2/C_1^2} \quad \textit{for all } t \geq 0.$$

(2) The moments of X satisfy

$$\|X\|_{\mathrm{L}^p} \leq C_2\sqrt{p} \quad \textit{for all } p \geq 1.$$

(3) The MGF of X^2 satisfies

$$\mathbb{E}\left[e^{\lambda^2 X^2}\right] \leq e^{C_3^2\lambda^2} \quad \textit{for all } |\lambda| \leq \frac{1}{C_3}.$$

(4) The MGF of X^2 is bounded at some point, namely

$$\mathbb{E}\left[e^{X^2/C_4^2}\right] \leq 2.$$

Moreover if $\mathbb{E}[X] = 0$, then the above properties are also equivalent to

(5) The MGF of X satisfies

$$\mathbb{E}\left[e^{\lambda X}\right] \leq e^{C_5^2\lambda^2} \quad \textit{for all } \lambda \in \mathbb{R}.$$

More precisely, show that

- (1) $\Rightarrow$ (2) *with* $C_2 \geq \sqrt{\pi}C_1$;
- (2) $\Rightarrow$ (3) *with* $C_3 \geq 2\sqrt{e}C_2$;
- (3) $\Rightarrow$ (4) *with* $C_4 \geq C_3/\sqrt{\log 2}$;
- (4) $\Rightarrow$ (1) *with* $C_1 \geq C_4$;

- (3) $\Rightarrow$ (5) *with* $C_5 \geq C_3$ *(under mean zero assumption);*
- (5) $\Rightarrow$ (1) *with* $C_1 \geq 2C_5$ *(under mean zero assumption).*

A random variable satisfying one of the above equivalent properties is called ***subgaussian****. To be more specific, we will call a random variable* σ^2***-subgaussian*** *if property* (5) *holds for* $X - \mathbb{E}[X]$ *with* $C_5^2 = \sigma^2/2$*. The constant* σ^2 *is called the* ***variance proxy****. The smallest such* σ^2 *is called the* ***optimal variance proxy****.*

Proof. Note that when showing that property i implies property j we can assume that i holds with $C_i = 1$, otherwise consider X/C_i. It then would be sufficient to show that C_j is a uniform constant (independent of any parameters such as p, λ, etc) as specified in the above bullet-points.

(1) $\Rightarrow$ (2): follows from the integral identity:

$$\mathbb{E}[|X|^p] \leq \int_0^\infty 2e^{-t^2} pt^{p-1} \mathrm{d}t = p\Gamma(p/2) \leq p^{p/2} \left(\max_{p \geq 1} \frac{\Gamma(p/2)^{1/p}}{p^{1/2-1/p}} \right)^p = (\sqrt{\pi})^p p^{p/2}.$$

Taking the p-th root yields (2) for any constant greater or equal than $\sqrt{\pi}$.

(2) $\Rightarrow$ (3): recall that by Stirling's formula we get $p! \geq (p/e)^p$. So,

$$\mathbb{E}\left[e^{\lambda^2 X^2}\right] = 1 + \sum_{p \geq 1} \frac{\lambda^{2p} \mathbb{E}\left[X^{2p}\right]}{p!} \leq 1 + \sum_{p \geq 1} \frac{(2p\lambda^2)^p}{(p/e)^p} = \frac{1}{1 - 2e\lambda^2}$$

provided that $2e\lambda^2 < 1$. Now we can use that on $x \in [0, 1/2]$, $1/(1-x) \leq e^{2x}$. Thus, for all $|\lambda| \leq \frac{1}{2\sqrt{e}}$ it holds $\mathbb{E}\left[e^{\lambda^2 X^2}\right] \leq e^{4e\lambda^2}$. Set $C_3 = 2\sqrt{e}(= 2\sqrt{e}C_2)$.

(3) $\Rightarrow$ (4): trivial.

(4) $\Rightarrow$ (1): By Markov's inequality we get $\mathbb{P}[|X| \geq t] \leq e^{-t^2}\mathbb{E}\left[e^{X^2}\right] \leq 2e^{-t^2}$. Set $C_1 = 1 = C_4$.

(3) $\Rightarrow$ (5): use that $e^x \leq x + e^{x^2}$ for all x. Since $\mathbb{E}[X] = 0$ in this case by assumption, we get that for $|\lambda| \leq 1$

$$\mathbb{E}\left[e^{\lambda X}\right] \leq \mathbb{E}\left[e^{\lambda^2 X^2}\right] \leq e^{\lambda^2}.$$

We have hence shown that the MGF of X is smaller or equal to e^{λ^2} on $\lambda \in [-1, 1]$. So let now $|\lambda| \geq 1$. Since $2\lambda x \leq x^2 + \lambda^2$ and by (3),

$$\mathbb{E}\left[e^{\lambda X}\right] \leq \mathbb{E}\left[e^{\lambda^2/2} e^{X^2/2}\right] \leq e^{\lambda^2/2} e^{1/2} \leq e^{\lambda^2}.$$

Set $C_5 = 1 = C_3$.

(5) $\Rightarrow$ (1): use exponential Markov's inequality and optimize over λ. Set $C_1 = 2C_5$. ☐

Exercise 5 (Hoeffding lemma)**.**

Let $a \leq X \leq b$ *a.s. for some* $a, b \in \mathbb{R}$*. Then* $\mathbb{E}\left[e^{\lambda(X - \mathbb{E}[X])}\right] \leq e^{\lambda^2 (b-a)^2/8}$*. That is,* X *is* $(b-a)^2/4$*-subgaussian.*

Hint: consider function $\psi(\lambda) = \log \mathbb{E}\left[e^{\lambda(X - \mathbb{E}[X])}\right]$*. Show that* $\psi''(\lambda)$ *can be interpreted as a variance of* $X - \mathbb{E}[X]$ *with respect to a new appropriately chosen probability measure. Use some suitable bound for variances that you know (sheet 2 might be helpful).*

Proof. Suppose wlog that $\mathbb{E}[X]=0$. In particular, we have $\psi(\lambda)=\log\mathbb{E}\left[e^{\lambda X}\right]$ and

$$\psi''(\lambda)=\frac{\mathbb{E}\left[X^2e^{\lambda X}\right]}{\mathbb{E}\left[e^{\lambda X}\right]}-\left(\frac{\mathbb{E}\left[Xe^{\lambda X}\right]}{\mathbb{E}\left[e^{\lambda X}\right]}\right)^2.$$

Let us define a new probability measure $\mathrm{d}\mathbb{Q}=\frac{e^{\lambda X}}{\mathbb{E}\left[e^{\lambda X}\right]}\mathrm{d}\mathbb{P}$. By rewriting the above equality with respect to this new measure we obtain,

$$\psi''(\lambda)=\mathbb{E}_{\mathbb{Q}}\left[X^2\right]-\mathbb{E}_{\mathbb{Q}}\left[X\right]^2=\mathrm{Var}_{\mathbb{Q}}[X].$$

By Exercise 1 Sheet 2 we know that the rhs is less or equal to $\frac{1}{4}(b-a)^2$. Now by fundamental theorem of calculus $(\psi(0)=0,\psi'(0)=0)$ we obtain

$$\psi(\lambda)=\int_0^\lambda\int_0^t\psi''(s)\mathrm{d}s\mathrm{d}t\leq\frac{\lambda^2(b-a)^2}{8}.$$

Exponentiation concludes the proof. □