Probability in High Dimensions

Solutions to Exercises, Week 2

July 25, 2020

2 Variance bounds and Poincaré inequalities

2.1 Tensorization and bounded differences

Problem 2.1 (Banach-valued sums)**.** *Let $X_1, \ldots, X_n$ be independent random variables with values in a Banach space $(B, \|\cdot\|_B)$. Suppose these random variables are bounded in the sense that $\|X_i\|_B \leq C$ a.s. for every i. Show that*

$$\operatorname{Var}\left(\left\|\frac{1}{n}\sum_{k=1}^{n} X_k\right\|_B\right) \leq \frac{C^2}{n}.$$

Solution. Alternative 1:
Let z_1, z_2, a be elements of B with norm bounded by C and $n \in \mathbb{N}$. Then

$$\begin{aligned}\left\|\frac{z_1}{n} + a\right\|_B &\leq \left\|\frac{z_1}{n} - \frac{z_2}{n}\right\|_B + \left\|\frac{z_2}{n} + a\right\|_B \\ &\leq \frac{2C}{n} + \left\|\frac{z_2}{n} + a\right\|_B\end{aligned}$$

This implies

$$\sup_{\|z_1\|_B \leq C} \left\|\frac{z_1}{n} + a\right\|_B \leq \frac{2C}{n} + \left\|\frac{z_2}{n} + a\right\|_B$$

which in turn yields

$$\sup_{\|z_1\|_B \leq C} \left\|\frac{z_1}{n} + a\right\|_B \leq \frac{2C}{n} + \inf_{\|z_2\|_B \leq C} \left\|\frac{z_2}{n} + a\right\|_B.$$

Set $a = \frac{1}{n}\sum_{k \neq i} x_k$ and $f(x_1, \ldots, x_n) = \left\|\frac{1}{n}\sum_{i=1}^{n} x_i\right\|_B$ then

$$\begin{aligned}D_i f(x) &= \sup_{\|z\|_B \leq C} \left\|\frac{z}{n} + \frac{1}{n}\sum_{k \neq i} x_k\right\|_B - \inf_{\|z\|_B \leq C} \left\|\frac{z}{n} + \frac{1}{n}\sum_{k \neq i} x_k\right\|_B \\ &\leq \frac{2C}{n}.\end{aligned}$$

By Corollary 2.4 (Bounded Differences) we obtain

$$\operatorname{Var}(f(X_1, \ldots, X_n)) \leq \frac{1}{4}\,\mathbf{E}\left[\sum_{i=1}^{n} (D_i f)^2\right] \leq \frac{1}{4}\left[\sum_{i=1}^{n} \frac{4C^2}{n^2}\right] = \frac{C^2}{n}.$$

Alternative 2:
In the spirit of an earlier version of this solution it is also possible to find the same bound by the reverse triangle inequality

$$\begin{aligned}
D_i f(x) &= \sup_{\|z_1\|_B \le C} \left\| \frac{z}{n} + a \right\|_B + \sup_{\|z_2\|_B \le C} - \left\| \frac{z}{n} + a \right\|_B \\
&= \sup \left\{ \left\| \frac{z_1}{n} + a \right\|_B - \left\| \frac{z_2}{n} + a \right\|_B : \|z_1\|_B \le C, \|z_2\|_B \le C \right\} \\
&\le \frac{1}{n} \sup \left\{ \|z_1 - z_2\|_B : \|z_1\|_B \le C, \|z_2\|_B \le C \right\} \le \frac{2C}{n}.
\end{aligned}$$

Alternative 3:

$$\begin{aligned}
\|\frac{1}{n}\sum_{i=1}^n X_i\|_B &= \sup_{t\in B^\circ} \left\langle t, \frac{1}{n}\sum_{i=1}^n X_i \right\rangle \\
D_i^- &= \sup_{t\in B^\circ} \left\langle t, \frac{1}{n}\sum_{i=1}^n X_i \right\rangle - \inf_{\|z_i\|\le C} \sup_{t\in B^\circ} \left\langle t, \frac{1}{n}\sum_{j\neq i} X_i + \frac{z_i}{n} \right\rangle \\
&= \sup_{\|z_i\|\le C} \left[\sup_{t\in B^\circ} \left\langle t, \frac{1}{n}\sum_{i=1}^n X_i \right\rangle - \sup_{t\in B^\circ} \left\langle t, \frac{1}{n}\sum_{j\neq i} X_i + \frac{z_i}{n} \right\rangle \right]
\end{aligned}$$

and since $\sup f - \sup g \le \sup[f-g]$

$$\begin{aligned}
&= \sup_{\|z_i\|\le C} \left[\sup_{t\in B^\circ} \left\{ \left\langle t, \frac{1}{n}\sum_{i=1}^n X_i \right\rangle - \left\langle t, \frac{1}{n}\sum_{j\neq i} X_i + \frac{z_i}{n} \right\rangle \right\} \right] \\
&= \sup_{\|z_i\|\le C} \left[\sup_{t\in B^\circ} \left\langle t, \frac{z_i}{n} \right\rangle \right] \\
&\le \frac{C}{n} \sup_{\|w\|\le 1} \left[\sup_{t\in B^\circ} \langle t, w\rangle \right]
\end{aligned}$$

and since $\sup_{t\in B^\circ}\langle t, w\rangle = \|w\|_B$

$$= \frac{C}{n} \sup_{\|w\|\le 1} [\|w\|_B] \le \frac{C}{n}.$$

Problem 2.2 (Rademacher process)**.** *Let $\varepsilon_1, \ldots, \varepsilon_n$ be independent symmetric random Bernoulli random variables* $\mathbf{P}[\varepsilon_i = \pm 1] = \frac{1}{2}$ *(also called Rademacher variable), let $T \subseteq \mathrm{R}^n$. The following identity is completely trivial:*

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

Prove the following nontrivial fact:

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

Thus taking the supremum inside the variance costs at most a constant factor.

Solution. Let $\langle x, y\rangle := \sum_{k=1}^n x_k y_k$ denote the inner product between vectors x and y in $\mathbb{R}^n$. Define the function $f : \varepsilon \in \{-1,1\}^n \to f(\varepsilon) := \sup_{t\in T}\langle \varepsilon, t\rangle$. **If we assume that for any $\varepsilon \in \{-1,1\}^n$ the supremum in $\sup_{t\in T}\langle \varepsilon, t\rangle$ is attained (this would be the case if the set T is compact, by the extreme value theorem)**, then we can let $t(\varepsilon) \in \arg\max_{t\in T}\langle \varepsilon, t\rangle$ denote any of its optimizers. Then, for any $\varepsilon \in \{-1,1\}^n$ and any $i \in \{1,\dots,n\}$, if we let $z(\varepsilon) \in \arg\min_{z\in\{-1,1\}} f(\varepsilon_1,\dots,\varepsilon_{i-1}, z, \varepsilon_{i+1},\dots,\varepsilon_n)$ we have

$$\begin{aligned}
D_i^- f(\varepsilon) &:= f(\varepsilon) - \min_{z\in\{-1,1\}} f(\varepsilon_1,\dots,\varepsilon_{i-1}, z, \varepsilon_{i+1},\dots,\varepsilon_n)\\
&= f(\varepsilon) - f(\varepsilon_1,\dots,\varepsilon_{i-1}, z(\varepsilon), \varepsilon_{i+1},\dots,\varepsilon_n)\\
&= \sup_{t\in T}\langle \varepsilon, t\rangle - \sup_{t\in T}\langle (\varepsilon_1,\dots,\varepsilon_{i-1}, z(\varepsilon), \varepsilon_{i+1},\dots,\varepsilon_n), t\rangle\\
&\le \langle \varepsilon, t(\varepsilon)\rangle - \langle (\varepsilon_1,\dots,\varepsilon_{i-1}, z(\varepsilon), \varepsilon_{i+1},\dots,\varepsilon_n), t(\varepsilon)\rangle\\
&= \langle (0,\dots,0,\varepsilon_i - z(\varepsilon), 0,\dots,0), t(\varepsilon)\rangle\\
&\le 2|t(\varepsilon)_i|.
\end{aligned}$$

The proof follows by applying Corollary 2.4 upon noticing that since $t(\varepsilon) \in T$ we have

$$\sum_{i=1}^n (D_i^- f(\varepsilon))^2 \le 4\sum_{i=1}^n t(\varepsilon)_i^2 \le 4\sup_{t\in T}\sum_{i=1}^n t_i^2.$$

More generally if $\mathcal{T}$ is bounded then by continuity we can replace it by its closure which is compact. Finally, if $\mathcal{T}$ is unbounded there must be a basis vector $\mathbf{e}_i$ such that $\sup_{t\in\mathcal{T}}\langle \mathbf{e}_i, t\rangle = \infty$, and since any basis vectors $\mathbf{e}_i$ can be written as a linear combination of two elements of $\{\pm 1\}^n$, we can always find an element $\varepsilon \in \{\pm1\}^n$ such that $\sup_{t\in\mathcal{T}}\langle \varepsilon, t\rangle = \infty$. In this case, with ϵ uniform in $\{\pm1\}^n$, we have $\sup_{t\in\mathcal{T}}\langle \epsilon, t\rangle = \infty$ with positive probability and the inequality holds trivially.

Problem 2.3 (Bin packing)**.**

Solution. (a)

$$D_i f(X_1,\dots,X_n) = \sup_{z_i} f(X_1,\dots,X_{i-1}, z_i, X_{i+1},\dots,X_n) - \inf_{z_i} f(X_1,\dots,X_{i-1}, z_i, X_{i+1},\dots,X_n).$$

The sup will be at worst $B_n + 1$ and the inf at worst $B_n - 1$. But, suppose that $\inf_{z_i} = B_n - 1$. This means that if we reduce the size of X_i to 0, then we gain a bin. Therefore, there is an optimal allocation with B_n bin in which X_i is a single occupant of a bin. Thus in this case, since $X_i \le 1$ we can increase its $D_i \le 1$. (b) Trivial since the total size of the packages $\sum X_i$ must fit in the total space of the bins which is $B_n \times 1$. Thus $n\mathbb{E}[X_1] \le \mathbb{E}[B_n]$.

Problem 2.4 (Order Statistics and spacings)**.** *Let $X_1, ..., X_n$ be independent random variables, and denote by $X_{(1)} \ge ... \ge X_{(n)}$ their decreasing rearrangement. Show that*

$$\mathrm{Var}\left[X_{(k)}\right] \le k\mathbb{E}\left[(X_{(k)} - X_{(k+1)})^2\right] \tag{1}$$

for $1 \le k \le n/2$ and that

$$\mathrm{Var}\left[X_{(k)}\right] \le (n-k+1)\,\mathbb{E}\left[(X_{(k-1)} - X_{(k)})^2\right] \tag{2}$$

for $n/2 < k \le n$.

Solution. To prove first (1), we note that for $f(X) = X_{(k)}$

$$D_i^- f(X) = X_{(k)} - \inf_Z\ f(X_{1:i-1}, Z, X_{i+1:n})$$

where

$$\inf_Z f\left(X_{1:i-1}, Z, X_{i+1:n}\right) = \begin{cases} X_{(k)} & \text{if } X_i < X_{(k)}, \\ X_{(k+1)} & \text{if } X_i \geq X_{(k)}. \end{cases}$$

Hence it follows by Corollary 2.4. that we have

$$\begin{aligned} \operatorname{Var}\left[X_{(k)}\right] &\leq \mathbb{E}\left[\sum_{i=1}^{n}\left(D_i^- f(X)\right)^2\right] \\ &\leq \mathbb{E}\left[\sum_{i=1}^{n}\mathbb{I}\left(X_i \geq X_{(k)}\right)\left(X_{(k)} - X_{(k+1)}\right)^2\right] \\ &\leq k\mathbb{E}\left[\left(X_{(k)} - X_{(k+1)}\right)^2\right] \end{aligned}$$

where the last inequality follows from the definition of $X_{(k)}$. One can verify that having ties is not an issue. This establishes (1) for $1 \leq k \leq n-1$ but this inequality is not tight for large values of k. To establish (2), one uses the increasing rearrangement $X_{(1)} \leq ... \leq X_{(n)}$ and performs similar calculations.

Problem 2.5 (Convex Poincaré inequalities)**.** *Let $X_1, \dots, X_n$ be independent random variables taking values in $[a, b]$. The bounded difference inequalities of Corollary 2.4 estimate the variance* $\operatorname{Var}[f(X_1, \dots, X_n)]$ *in terms of* discrete *derivatives $D_i f$ or $D_i^- f$ of the function f. The goal of this problem is to show that if the function f is* convex*, then one can obtain a similar bound in terms of the ordinary notion of derivative $\nabla_i f(x) = \frac{\partial f(x)}{\partial x_i}$ in $\mathbb{R}^n$.*

a. *Show that if $g : \mathbb{R} \to \mathbb{R}$ is convex, then*

$$g(y) - g(x) \geq g'(x)(y - x) \quad \textit{for all } x, y \in \mathbb{R}.$$

b. *Show using part a. and Corollary 2.4 that if $f : \mathbb{R}^n \to \mathbb{R}$ is convex, then*

$$\operatorname{Var}[f(X_1, \dots, X_n)] \leq (b-a)^2\, \mathbf{E}[\|\nabla f(X_1, \dots, X_n)\|^2].$$

c. *Conclude that if f is convex and L-Lipschitz, i.e., $|f(x) - f(y)| \leq L\|x - y\|$ for all $x, y \in [a, b]^n$, then* $\operatorname{Var}[f(X_1, \dots, X_n)] \leq L^2(b-a)^2$.

Solution. a. This can be proven directly from the definition of convexity and the derivative: for any $t \in [0, 1]$ and $x, y \in \mathbb{R}$,

$$\begin{aligned} (1-t)g(x) + tg(y) &\geq g(x + t(y - x)) \\ g(y) - g(x) &\geq \frac{g(x + t(y-x)) - g(x)}{t}. \end{aligned}$$

Taking the limit as $t \to 0$ completes the proof.

b. Consider a function $g_i : [a, b] \to [a, b]$, defined as

$$g_i(z) \equiv f(x_1, \dots, x_{i-1}, z, x_{i+1}, \dots, x_n),$$

for some fixed $\{x_1, \dots, x_{i-1}, x_{i+1}, \dots, x_n\} \in [a, b]^{n-1}$ and $i \in \{1, \dots, n\}$. This function is convex since f is convex in all of its arguments. Thus, by part a.,

$$g_i(z) \geq g_i(x_i) + g_i'(x_i)(z - x_i)$$

for all $z, x_i \in [a,b]$. In other words, for all $\{x_1, \dots, x_n, z\} \in [a,b]^{n+1}$,

$$f(x_1, \dots, x_{i-1}, z, x_{i+1}, \dots x_n) \geq f(x) + \nabla_i f(x)(z - x_i),$$

where we write $x \equiv \{x_1, \dots, x_n\}$. Thus, we can bound $D_i^- f(x)$, for all $i \in \{1, \dots, n\}$, as

$$\begin{aligned} D_i^- f(x) &= f(x) - \inf_{a \leq z \leq b} f(x_1, \dots, x_{i-1}, z, x_{i+1}, \dots, x_n) \\ &\leq f(x) - \left[f(x) + \inf_{a \leq z \leq b} \nabla_i f(x)(z - x_i) \right] \\ &= - \inf_{a \leq z \leq b} \nabla_i f(x)(z - x_i). \end{aligned}$$

Finally, invoking Corollary 2.4, we have

$$\begin{aligned} \mathrm{Var}[f(X_1, \dots, X_n)] &\leq \mathbf{E}\left[\sum_{i=1}^n \left(D_i^- f(X_1, \dots, X_n) \right)^2 \right] \\ &\leq \mathbf{E}\left[\sum_{i=1}^n \left(\inf_{a \leq z \leq b} \nabla_i f(X_1, \dots, X_n)(z - X_i) \right)^2 \right] \\ &\leq \mathbf{E}\left[(b-a)^2 \sum_{i=1}^n \nabla_i f(X_1, \dots, X_n)^2 \right] \\ &= (b-a)^2 \, \mathbf{E}\left[\|\nabla_i f(X_1, \dots, X_n)\|^2 \right]. \end{aligned}$$

c. Since f is L-Lipschitz, we know that $\|f(x)\| \leq L$ for all $x \in [a,b]^n$. Thus, from part b.,

$$\mathrm{Var}[f(X_1, \dots, X_n)] \leq L^2 (b-a)^2.$$

2.2 Markov semigroups

Problem 2.6 (Some elementary identities)**.** *Let P_t be a Markov semigroup with generator $\mathcal{L}$ and stationary measure μ. Prove the following elementary facts:*

a) Show that $\mu(\mathcal{L}f) = 0$ for every $f \in \mathrm{Dom}(\mathcal{L})$.

b) If $\phi \colon \mathbb{R} \to \mathbb{R}$ is convex, then $P_t \phi(f) \geq \phi(P_t f)$ when $f, \phi(f) \in L^2(\mu)$.

c) If $\phi \colon \mathbb{R} \to \mathbb{R}$ is convex, then $\mathcal{L}\phi(f) \geq \phi'(f)\mathcal{L}f$ when $f, \phi(f) \in \mathrm{Dom}(\mathcal{L})$.

d) Let $f \in \mathrm{Dom}(\mathcal{L})$. Show that the following process is a martingale:

$$M_t^f := f(X_t) - \int_0^t \mathcal{L}f(X_s) \mathrm{d}s.$$

Solution. a) Recall that the generator is defined as a limit in $L^2(\mu)$, i.e. we have

$$\lim_{t \to 0} \left\| \frac{1}{t}(P_t f - f) - \mathcal{L}f \right\|_{L^2(\mu)} \to 0.$$

In particular, this implies

$$\lim_{t \to 0} \int \frac{1}{t} \left(P_t f - f \right) \mathrm{d}\mu = \int \mathcal{L}f \mathrm{d}\mu.$$

Since stationarity implies for all t

$$\frac{1}{t}\int (P_t f - f)\,\mathrm{d}\mu = \frac{1}{t}\mathbb{E}\left[\mathbb{E}\left[f(X_t) \mid X_0\right] - f(X_0)\right] = 0$$

we can conclude

$$\int \mathcal{L}f\mathrm{d}\mu = \lim_{t\to 0}\int \frac{1}{t}\left(P_t f - f\right)\mathrm{d}\mu = 0.$$

b) This is just the conditional version of Jensen's inequality, i.e. for any x

$$P_t\phi(f(x)) = \mathbb{E}\left[\phi(f(X_t)) \mid X_0 = x\right] \geq \phi\left(\mathbb{E}\left[f(X_t) \mid X_0 = x\right]\right) = \phi(P_t f(x)).$$

c) By convexity we have for any y, z that $\phi(y) - \phi(z) \geq \phi'(z)(y - z)$. Using Jensen's inequality from above and with $y = P_t f(x)$ and $z = f(x)$, we get

$$\begin{aligned} \frac{P_t(\phi(f_x)) - \phi(f(x)}{t} &\geq \frac{\phi(P_t f(x)) - \phi(f(x)}{t} \\ &\geq \frac{\phi'(f(x))\left(P_t f(x) - f(x)\right)}{t}. \end{aligned}$$

Taking the limit $t \to 0$ yields the required result.

d) We have

$$M_t^f := f(X_t) - \int_0^t \mathcal{L}f(X_s)\mathrm{d}s.$$

Let $\{\mathcal{F}_t\}_{t\geq 0}$ denote the natural filtration with respect to $X = (X_t)$ and $\tau \leq t$, then

$$\begin{aligned} \mathbb{E}\left[M_t^f \mid \mathcal{F}_\tau\right] &= \mathbb{E}\left[f(X_t) \mid \mathcal{F}_\tau\right] - \mathbb{E}\left[\int_0^t \mathcal{L}f(X_s)\mathrm{d}s \mid \mathcal{F}_\tau\right] \\ &= \mathbb{E}\left[f(X_t) \mid X_\tau\right] - \mathbb{E}\left[\int_\tau^t \mathcal{L}f(X_s)\mathrm{d}s \mid X_\tau\right] - \mathbb{E}\left[\int_0^\tau \mathcal{L}f(X_s)\mathrm{d}s \mid \mathcal{F}_\tau\right] \\ &= P_{t-\tau}f(X_\tau) - \mathbb{E}\left[\int_0^\tau \mathcal{L}f(X_s)\mathrm{d}s \mid X_\tau\right] - \int_0^\tau \mathcal{L}f(X_s)\mathrm{d}s \\ &= P_{t-\tau}f(X_\tau) - P_{t-\tau}f(X_\tau) + f(X_\tau) - \int_0^\tau \mathcal{L}f(X_s)\mathrm{d}s \\ &= f(X_\tau) - \int_0^\tau \mathcal{L}f(X_s)\mathrm{d}s = M_\tau^f, \end{aligned}$$

where we used that $\int_0^\tau \mathcal{L}f(X_s)\mathrm{d}s$ is $\mathcal{F}_\tau$ measurable in the second line and the Markov property in the third.