Probability in High Dimensions

Solutions to Exercises, Week 7

September 5, 2020

3.2 The martingale method

Problem 3.6.

Problem 3.7.

Problem 3.8. *(Empirical Frequencies)*

Solution. We use the following notations :

$$f(X) = Z_n(X) = \sup_C \left| \frac{\#\{k \in \{1, 2, ..., n\} : X_k \in C\}}{n} - \mu(C) \right|$$

$$Y_n(C, X) = \left| \frac{\#\{k \in \{1, 2, ..., n\} : X_k \in C\}}{n} - \mu(C) \right|$$

For any C, X and X' (such that X' is different from X only on the ith component), we have :

$$\begin{aligned} Y_n(C, X) - Z_n(X') &\leq Y_n(C, X) - Y_n(C, X') \\ &\leq \frac{1}{n} \end{aligned}$$

and by taking the supremum over C, we have:

$$Z_n(X) - Z_n(X') \leq \frac{1}{n}$$

which gives us :

$$D_i(f) \leq \frac{1}{n}$$

We conclude using McDiarmid's inequality.

Problem 3.9 (Sums in Hilbert space)**.** *Let $X_1, \dots, X_n$ be independent random variables with zero mean in a Hilbert space, and suppose that $\|X_k\| \leq C$ a.s. for every k.*

a.

b.

c. Conclude that for all $t \geq \mathbf{2}Cn^{-1/2}$ (note the 2 in the bound which differs from van Handel's notes),

$$\mathbf{P}\left[\left\| \frac{1}{n} \sum_{k=1}^{n} X_k \right\| \geq t \right] \leq e^{-nt^2/8C^2}.$$

d. Finally, argue that for all $t \geq 0$,

$$\mathbf{P}\left[\left\|\frac{1}{n}\sum_{k=1}^{n} X_k\right\| \geq t\right] \leq 2e^{-nt^2/8C^2}.$$

Solution. a.

b.

c. For any $t \geq \frac{2C}{\sqrt{n}}$,

$$\mathbf{P}\left[\left\|\frac{1}{n}\sum_{k=1}^{n} X_k\right\| \geq t\right] = \mathbf{P}\left[\left\|\frac{1}{n}\sum_{k=1}^{n} X_k\right\| \geq \mathbf{E}\left\|\frac{1}{n}\sum_{k=1}^{n} X_k\right\| + \tilde{t}\right],$$

where $\tilde{t} = t - \mathbf{E}\left\|\frac{1}{n}\sum_{k=1}^{n} X_k\right\|$. Note that $\tilde{t} \geq 0$ since $t \geq \frac{2C}{\sqrt{n}}$. Thus, from part (a),

$$\mathbf{P}\left[\left\|\frac{1}{n}\sum_{k=1}^{n} X_k\right\| \geq t\right] \leq e^{-n\tilde{t}^2/2C^2}.$$

Now, let us consider the difference $4\tilde{t}^2 - t^2$:

$$4\tilde{t}^2 - t^2 = (2\tilde{t} - t)(2\tilde{t} + t) = \left(t - 2\,\mathbf{E}\left\|\frac{1}{n}\sum_{k=1}^{n} X_k\right\|\right)\left(3t - 2\,\mathbf{E}\left\|\frac{1}{n}\sum_{k=1}^{n} X_k\right\|\right).$$

Since $t \geq \frac{2C}{\sqrt{n}} \geq 2\,\mathbf{E}\left\|\frac{1}{n}\sum_{k=1}^{n} X_k\right\|$, we have that $4\tilde{t}^2 - t^2 \geq 0$, i.e. $4\tilde{t}^2 \geq t^2$. Note that this does **NOT** hold true if t is only bounded below by $\frac{C}{\sqrt{n}}$ (implying a likely typo in van Handel's notes). Therefore,

$$\mathbf{P}\left[\left\|\frac{1}{n}\sum_{k=1}^{n} X_k\right\| \geq t\right] \leq e^{-n\tilde{t}^2/2C^2} \leq e^{-nt^2/8C^2}.$$

d. For the right-hand side to be non-trivial, $t \geq 0$ must satisfy

$$2e^{-nt^2/8C^2} \leq 1,$$

which we can rewrite as

$$t \geq \sqrt{2\log 2}\,\frac{2C}{\sqrt{n}} > \frac{2C}{\sqrt{n}},$$

since $\sqrt{2\log 2} > 1$. However, the bound holds for all $t \geq \frac{2C}{\sqrt{n}}$ by part (c). Thus, the bound is satisfied for all $t \geq 0$.

Problem 3.10. *An Erdös-Rényi random graph $G(n,p)$ is a graph on n vertices such that for every pair of vertices v, v' there is an edge between them with probability p, independently of the other edges. A coloring of the graph is the assignment of a color to each vertex such that every pair of vertices connected by an edge have a distinct color. The* chromatic number *χ is the minimal number of colors needed to color the graph. Show that*

$$\mathbf{P}\left[|\chi - \mathbf{E}\,\chi| \geq t\sqrt{n}\right] \leq 2e^{-t^2}.$$

It can be shown that the chromatic number satisfies $\mathbf{E}\,\chi \sim n/2\log_b n$ as $n \to \infty$, where $b = 1/(1-p)$. We therefore see that the fluctuations of the chromatic number are of much smaller ordr than its magnitude.

Solution. Describe a graph of n vertices by the set of edges (i, j), $i < j$, as

$$G(X_1, \dots, X_n) = \bigcup_{k=1}^{n} X_n,$$

where $X_k \subseteq \{(i, j) \in [k-1] \times \{k\}\}$ is the set of edges connecting vertex k to vertices with index less than k. Note this description only uses n variables, and for the ER random graph these variables are independent.

Apply McDiarmid's theorem with $f(X_1, \dots, X_n)$ as the chromatic number of the graph $G(X_1, \dots, X_n)$:

$$D_k f(x_{1:n}) = \sup_z f(x_{1:k-1}, z, x_{k+1:z}) - \inf_z f(x_{1:k-1}, z, x_{k+1:z}).$$

First notice that

$$\begin{aligned} \sup_z f(x_{1:k-1}, z, x_{k+1:z}) &\leq f(x) + 1 \\ \inf_z f(x_{1:k-1}, z, x_{k+1:z}) &\geq f(x) - 1, \end{aligned}$$

since in the first case, the worst case scenario is that you connect the k-th vertex to another vertex of the same colour, so in this case you can just give the k-th vertex a completely new, unused colour and be done with it. In the second case, at best you by removing an edge you can now give the k-th vertex the same colour as another vertex in the graph, thereby saving at best one colour.

But suppose first that

$$\sup_z f(x_{1:k-1}, z, x_{k+1:z}) = f(x) + 1;$$

this means that the k-th vertex currently shares its colour with another vertex that it is not connected to, since by adding the edge we have to give the k-th vertex(or the one connected to it) a new colour. But then if this is the case, even since the colour of the k-th vertex is already used by another vertex, by changing the colour of the k-th vertex we can't improve the chromatic number and thus

$$\inf_z f(x_{1:k-1}, z, x_{k+1:z}) = f(x).$$

Similarly if

$$\inf_z f(x_{1:k-1}, z, x_{k+1:z}) = f(x) - 1;$$

we conclude that the k-th vertex has a unique colour and thus

$$\sup_z f(x_{1:k-1}, z, x_{k+1:z}) = f(x).$$

Overall

$$\left| D_k f(x_{1:n}) = \sup_z f(x_{1:k-1}, z, x_{k+1:z}) - \inf_z f(x_{1:k-1}, z, x_{k+1:z}) \right| \leq 1.$$

This quantity is 0 or 1 as changing the set of edges to vertex k can at most require that vertex k take on a new colour.

Plugging in, we get

$$\begin{aligned} \mathbf{P}\left[f(X_1, \dots, X_n) - \mathbf{E}\, f(X_1, \dots, X_n) \geq t\right] &\leq e^{-2t^2 / \sum_{k=1}^{n} \|D_k f\|_\infty^2} \\ &= e^{-2t^2/n} \end{aligned}$$

which implies

$$\mathbf{P}\left[|f(X_1,\dots,X_n)-\mathbf{E}\,f(X_1,\dots,X_n)|\geq t\right]\leq 2e^{-2t^2/n},$$

which is better than the problem statement which claims $2e^{-t^2/n}$.

Problem 3.11. *(A Generalization of Azuma-Hoeffding) Given filtration* $\{\mathcal{F}_k\}_{k\leq n}$ *with* Δ_k *satisfying* $E[\Delta_k|\mathcal{F}_{k-1}]=0$ *and* $A_k\leq\Delta_k\leq B_k$ *show that*

$$\mathbb{P}\left[\sum_{k=1}^n\Delta_k\geq t\ \textit{and}\ \sum_{k=1}^n(B_k-A_k)^2\leq c^2\right]\leq e^{-2t^2/c^2} \tag{1}$$

Solution. We start with an assumption which we will prove at the end:

$$E\left[\exp\left\{\lambda\sum_{k=1}^n\Delta_k-\frac{\lambda^2}{8}\sum_{k=1}^n(B_k-A_k)^2\right\}\right]\leq 1. \tag{2}$$

Start by applying Markov inequality with $f(x)=e^{\lambda x}$ giving

$$\mathbb{P}\left(\sum_{k=1}^n\Delta_k-\frac{\lambda}{8}\sum_{k=1}^n(A_k-B_k)^2\geq t\right)e^{\lambda t}\leq E\left[\exp\left\{\lambda\sum_{k=1}^n\Delta_k-\frac{\lambda^2}{8}\sum_{k=1}^n(B_k-A_k)^2\right\}\right]. \tag{3}$$

Focusing on the left probability, recall that for any events A,B we have $\mathbb{P}(A)\geq\mathbb{P}(A\cap B)$. This means

$$\mathbb{P}\left(\left\{\sum_{k=1}^n\Delta_k-\frac{\lambda}{8}\sum_{k=1}^n(A_k-B_k)^2\geq t\right\}\cap\left\{\sum_{k=1}^n(A_k-B_k)^2\leq c^2\right\}\right)\leq\mathbb{P}\left(\sum_{k=1}^n\Delta_k-\frac{\lambda}{8}\sum_{k=1}^n(A_k-B_k)^2\geq t\right). \tag{4}$$

Additionally observe that we have inequality

$$\mathbb{P}\left(\left\{\sum_{k=1}^n\Delta_k-\frac{\lambda c^2}{8}\geq t\right\}\cap\left\{\sum_{k=1}^n(A_k-B_k)^2\leq c^2\right\}\right)\leq \tag{5}$$

$$\mathbb{P}\left(\left\{\sum_{k=1}^n\Delta_k-\frac{\lambda}{8}\sum_{k=1}^n(A_k-B_k)^2\geq t\right\}\cap\left\{\sum_{k=1}^n(A_k-B_k)^2\leq c^2\right\}\right). \tag{6}$$

Going back to the expectation we get

$$\mathbb{P}\left(\left\{\sum_{k=1}^n\Delta_k-\frac{\lambda c^2}{8}\geq t\right\}\cap\left\{\sum_{k=1}^n(A_k-B_k)^2\leq c^2\right\}\right)\leq e^{-\lambda t} \tag{7}$$

now optimising with $\frac{\lambda c^2}{8}=t\implies\lambda=\frac{8t}{c^2}$

$$\mathbb{P}\left(\left\{\sum_{k=1}^n\Delta_k\geq 2t\right\}\cap\left\{\sum_{k=1}^n(A_k-B_k)^2\leq c^2\right\}\right)\leq e^{-8t^2/c^2}. \tag{8}$$

Finally let $2t = s$ to get the result.
Now for each $k = 1, \ldots, n$ we have from Hoeffdings Lemma 3.6 $E[e^{\lambda \Delta_k} | \mathcal{F}_{k-1}] \leq e^{\lambda^2 (B_k - A_k)^2/8}$ therefore

$$E\left[\exp\left\{\lambda \sum_{k=1}^{n} \Delta_k - \frac{\lambda^2}{8} \sum_{k=1}^{n} (B_k - A_k)^2\right\}\right] = \tag{9}$$

$$E\left[\exp\left\{\lambda \sum_{k=2}^{n} \Delta_k - \frac{\lambda^2}{8} \sum_{k=2}^{n} (B_k - A_k)^2\right\} E\left[e^{\lambda \Delta_k - \frac{\lambda^2}{8}(B_1 - A_k)^2} | \mathcal{F}_1\right]\right] \leq \tag{10}$$

$$E\left[\exp\left\{\lambda \sum_{k=2}^{n} \Delta_k - \frac{\lambda^2}{8} \sum_{k=2}^{n} (B_k - A_k)^2\right\}\right] \leq 1 \tag{11}$$

where we have iteratively taken the expectation in.