Homework 2 HDP

Bizzaro Davide, Schiavo Leonardo

October 2021

We use the sloppiness of notation of calling each constant C, despite the fact that it may be a different constant from passage to passage.

Exercise 1. We want to deduce from the theorem of the concentration of the norm that

$$\sqrt{n} - CK^2 \leq \mathbb{E}||X||_2 \leq \sqrt{n} + CK^2 \tag{0.1}$$

The theorem states that if $X = (X_1, \dots, X_n) \in \mathbb{R}^n$ is a random vector with independent sub-Gaussian coordinates X_i that satisfy $\mathbb{E}X_i^2 = 1$, then

$$\big|\big| ||X||_2 - \sqrt{n} \big|\big|_{\psi_2} \leq CK^2 \tag{0.2}$$

where $K := \max_i ||X_i||_{\psi_2}$ and C is an absolute constant.

This implies that $||X||_2 - \sqrt{n}$ is sub-Gaussian, and by formula 2.15 of the textbook (with $p = 1$) we have that

$$\mathbb{E}\big| ||X||_2 - \sqrt{n} \big| = \big|\big| ||X||_2 - \sqrt{n} \big|\big|_{L^1} \leq CK^2.$$

Combining this with Jensen's inequality

$$\big| \mathbb{E}||X||_2 - \sqrt{n} \big| \leq \mathbb{E}\big| ||X||_2 - \sqrt{n} \big|$$

we obtain the thesis

$$\big| \mathbb{E}||X||_2 - \sqrt{n} \big| \leq CK^2. \tag{0.3}$$

Exercise 2. We want to deduce from the theorem of the concentration of the norm that

$$Var(||X||_2) \leq CK^4 \tag{0.4}$$

Using again the fact that $||X||_2 - \sqrt{n}$ is sub-Gaussian with $\big|\big| ||X||_2 - \sqrt{n} \big|\big|_{\psi_2} \leq CK^2$, and using formula 2.15 of the textbook (this time with $p = 2$) we have that

$$\big|\big| ||X||_2 - \sqrt{n} \big|\big|_{L^2} \leq C \big|\big| ||X||_2 - \sqrt{n} \big|\big|_{\psi_2} \leq CK^2$$

so that squaring both sides

$$\mathbb{E}\big(||X||_2 - \sqrt{n} \big)^2 \leq CK^4. \tag{0.5}$$

Now we can see through a calculation that

$$Var(||X||_2) \leq \mathbb{E}\big(||X||_2 - \sqrt{n} \big)^2. \tag{0.6}$$

Indeed

$$(\mathbb{E}||X||_2)^2 - 2\sqrt{n}\mathbb{E}||X||_2 + n = \big(\mathbb{E}||X||_2 - \sqrt{n} \big)^2 \geq 0$$

and so

$$Var(||X||_2) = \mathbb{E}||X||_2^2 - (\mathbb{E}||X||_2)^2 \leq \mathbb{E}||X||_2^2 - \big(2\sqrt{n}\mathbb{E}||X||_2 - n \big) = \mathbb{E}\big(||X||_2 - \sqrt{n} \big)^2. \tag{0.7}$$

Finally, it follows from 0.5 and 0.6 that

$$Var(||X||_2) \leq \mathbb{E}\big(||X||_2 - \sqrt{n} \big)^2 \leq CK^4,$$

as we wanted.

Exercise 3. Let $X = (X_1, \dots, X_n) \in \mathbb{R}^n$ be a random vector with independent coordinates X_i that satisfy $\mathbb{E}X_i^2 = 1$ and $\mathbb{E}X_i^4 \le K^4$. Show that $Var(||X||_2) \le CK^4$.

We first show that

$$\mathbb{E}\big[(||X||_2^2 - n)^2\big] \le nK^4 \tag{0.8}$$

Indeed,

$$\begin{aligned}
\mathbb{E}\big[(||X||_2^2 - n)^2\big] &= \mathbb{E}\big[\big(\sum_{i=1}^n X_i^2\big)^2 - 2n\big(\sum_{i=1}^n X_i^2\big) + n^2\big] \\
&= \mathbb{E}\big[\big(\sum_{i=1}^n X_i^2\big)^2\big] - 2n\sum_{i=1}^n \mathbb{E}[X_i^2] + n^2 \\
&= \mathbb{E}\big[\big(\sum_{i=1}^n X_i^2\big)^2\big] - n^2 \\
&= \mathbb{E}\big[\sum_{i=1}^n X_i^4\big] + 2\mathbb{E}\big[\sum_{i<j=1}^n X_i^2 X_j^2\big] - n^2 \\
&= \mathbb{E}\big[\sum_{i=1}^n X_i^4\big] + 2\sum_{i<j=1}^n \mathbb{E}X_i^2\mathbb{E}X_j^2 - n^2 \\
&\le nK^4 + n(n-1) - n^2 \\
&= nK^4 - n \\
&\le nK^4
\end{aligned}$$

Now we observe that

$$\big(||X||_2^2 - n\big)^2 = \big(||X||_2 - \sqrt{n}\big)^2\big(||X||_2 + \sqrt{n}\big)^2 \ge n\big(||X||_2 - \sqrt{n}\big)^2$$

and thus

$$nK^4 \ge \mathbb{E}\big[\big(||X||_2^2 - n\big)^2\big] \ge n\mathbb{E}\big[\big(||X||_2 - \sqrt{n}\big)^2\big].$$

So we have that

$$\mathbb{E}\big[\big(||X||_2 - \sqrt{n}\big)^2\big] \le K^4$$

and now we can conclude recalling 0.7 from exercise 2:

$$Var(||X||_2) = \mathbb{E}||X||_2^2 - (\mathbb{E}||X||_2)^2 \le \mathbb{E}\big[\big(||X||_2 - \sqrt{n}\big)^2\big] \le K^4.$$

Exercise 4. Let $X = (X_1, \dots, X_n) \in \mathbb{R}^n$ be a random vector with independent coordinates X_i with continuous distributions. Assume that the densities of X_i are uniformly bounded by 1. Show that, for any $\epsilon > 0$, we have

$$\mathbb{P}\big[||X||_2 \le \epsilon\sqrt{n}\big] \le (C\epsilon)^n.$$

We first notice that since $X_1, \dots, X_n$ are independent, then $X_1^2, \dots, X_n^2$ are independent too. Therefore we can use the result of Exercise 2.2.10 of the textbook, obtaining that

$$\mathbb{P}\big[\sum_{i=1}^n X_i^2 \le \epsilon n\big] \le (e\epsilon)^n \tag{0.9}$$

for all $\epsilon > 0$. Hence we have that

$$\mathbb{P}\big[||X||_2 \le \sqrt{\epsilon n}\big] \le (e\epsilon)^n \tag{0.10}$$

and changing ϵ with ϵ^2 we have the thesis: assuming $\epsilon < 1$ (otherwise just consider $C = 1$ and the thesis is clearly satisfied),

$$\mathbb{P}\big[||X||_2 \le \epsilon\sqrt{n}\big] \le e^n\epsilon^{2n} \le e^n\epsilon^n = (C\epsilon)^n. \tag{0.11}$$

Exercise 5. Let X and Y be independent mean-zero isotropic random vectors in $\mathbb{R}^n$. We want to check that

$$\mathbb{E}||X - Y||_2^2 = 2n. \tag{0.12}$$

We have

$$
\begin{aligned}
\mathbb{E}||X-Y||_2^2 &= \sum_{i=1}^{n} \mathbb{E}(X_i - Y_i)^2 \\
&= \sum_{i=1}^{n} \mathbb{E}[X_i^2 + Y_i^2 - 2X_iY_i] \\
&= \sum_{i=1}^{n} (\mathbb{E}X_i^2 + \mathbb{E}Y_i^2 - 2\mathbb{E}X_i\mathbb{E}Y_i) \\
&= \sum_{i=1}^{n} (\mathbb{E}X_i^2 + \mathbb{E}Y_i^2)
\end{aligned}
$$

where we used the independence of the variables and the fact that they have mean zero. Now, since X and Y are isotropic, it follows that $\mathbb{E}X_i^2 = \mathbb{E}Y_i^2 = 1$ for all $i = 1, \ldots, n$. In the end

$$
\mathbb{E}||X-Y||_2^2 = \sum_{i=1}^{n} (\mathbb{E}X_i^2 + \mathbb{E}Y_i^2) = \sum_{i=1}^{n} 2 = 2n, \tag{0.13}
$$

as we wanted.

Exercise 6. We want to prove that $X \sim N(0, \mathbb{I}_n)$ is isotropic.

By the definition we need to prove that $\mathbb{E}XX^T = \mathbb{I}_n$. We just do the calculation on the coordinates:

$$
\mathbb{E}[XX^T]_{ij} = \mathbb{E}[X_iX_j] = \mathrm{E}[X_iX_j] - \mathbb{E}[X_i]\mathrm{E}[X_j] = Cov[X_i, X_j] = Cov[X]_{ij} = \delta_{ij} \tag{0.14}
$$

where we used the fact that the coordinates of X are independent with standard normal distribution, so they have mean zero, they are uncorrelated, and they have variance 1.

Exercise 7. We want to show that, given a random vector X in $\mathbb{R}^n$, it has normal distribution (say $X \sim N(\mu, \Sigma)$) if and only if $\langle X, x\rangle$ has a normal distribution for all $x \in \mathbb{R}^n$.

Let us first suppose that $X \sim N(\mu, \Sigma)$. By definition $X := \mu + \Sigma^{1/2}Z$, where Z has the standard normal distribution. Hence,

$$
\langle X, x\rangle = \langle \mu + \Sigma^{1/2}Z, x\rangle = \langle \mu, x\rangle + \langle Z, (\Sigma^{1/2})^T x\rangle = \langle \mu, x\rangle + \langle Z, y\rangle \tag{0.15}
$$

where $y := (\Sigma^{1/2})^T x = \Sigma^{1/2}x$. We know that the coordinates Z_i ($i = 1, \ldots, n$) of Z are independent random variables with distribution $N(0,1)$ and that any linear combination of independent normally distributed random variables is normally distributed. In particular

$$
\langle X, x\rangle = \langle \mu, x\rangle + \langle Z, y\rangle = \langle \mu, x\rangle + \sum_{i=1}^{n} y_i Z_i \sim N\left(\langle \mu, x\rangle, \sum_{i=1}^{n} y_i^2\right) \tag{0.16}
$$

.

Now, suppose that $\langle X, x\rangle$ has a normal distribution for all $x \in \mathbb{R}^n$. Without loss of generality we can subtract from X its mean and therefore assume X to have mean zero. Let $\Sigma := cov(X) = \mathbb{E}(XX^T)$ be the covariance matrix associated to X and let $\tilde{X} \sim N(0, \Sigma)$. If we prove that $\langle X, x\rangle$ and $\langle \tilde{X}, x\rangle$ have the same distribution for all $x \in \mathbb{R}^n$, then by Cramér-Wold's theorem it follows that X and $\tilde{X}$ have the same distribution and we are done. Now let X_i and x_i (for $i = 1, \ldots, n$) denote the coordinates of X and x respectively, and let $\Sigma_{i,j} := cov(X_i, X_j)$ (for $i, j = 1, \ldots, n$) denote the element of Σ in position (i, j). By hypothesis $\langle X, x\rangle = \sum_{i=1}^{n} x_i X_i$ is normally distributed. Moreover, $\mathbb{E}\langle X, x\rangle = \sum_{i=1}^{n} x_i \mathbb{E}X_i = 0$ and it has to be

$$
Var(\langle X, x\rangle) = Var(\sum_{i=1}^{n} x_i X_i) = \sum_{i,j=1}^{n} x_i x_j Cov(X_i, X_j) = \sum_{i,j=1}^{n} x_i x_j \Sigma_{i,j} = \langle x, \Sigma x\rangle. \tag{0.17}
$$

Therefore $\langle X, x\rangle \sim N(0, \langle x, \Sigma x\rangle)$. On the other hand by 0.16 we have that $\langle \tilde{X}, x\rangle$ is a normal distribution with mean 0 and variance

$$
\sum_{i=1}^{n} y_i^2 = \langle y, y\rangle = \langle \Sigma^{1/2}x, \Sigma^{1/2}x\rangle = \langle x, (\Sigma^{1/2})^T \Sigma^{1/2} x\rangle = \langle x, \Sigma x\rangle. \tag{0.18}
$$

By the basic fact that a normal distribution is uniquely characterized by its mean and variance, $\langle X, x\rangle$ and $\langle \tilde{X}, x\rangle$ have the same distribution for all $x \in \mathbb{R}^n$, and so by Cramér-Wold's theorem we are finished.

Exercise 8. Given $X \sim N(0, \mathbb{I}_n)$, we want to prove that

1. $||X||_2$ and $Y := \frac{X}{||X||_2}$ are independent;
2. $\sqrt{n}Y \sim Unif(\sqrt{n}S^{n-1})$.

Solution.

1. Let us consider the map

$$\Phi \colon \mathbb{R}^n \to \mathbb{R}^+ \times S^{n-1}, \quad x \mapsto \left(||x||_2, \frac{x}{||x||_2} \right) \tag{0.19}$$

and the change of variables to spherical coordinates

$$id \times \Psi \colon \mathbb{R}^+ \times S^{n-1} \to \mathbb{R}^+ \times [0,\pi]^{n-2} \times [0, 2\pi[, \quad (\rho, \theta) \mapsto (\rho, \phi_1, \dots, \phi_{n-2}, \phi_{n-1}) \tag{0.20}$$

Now, for all measurable subsets A_1 of $\mathbb{R}^+$ and A_2 of S^{n-1} we have

$$\mathbb{P}\left(||X||_2 \in A_1, Y \in A_2\right) = \mathbb{P}(X \in \Phi^{-1}(A_1 \times A_2)) = \int_{\Phi^{-1}(A_1 \times A_2)} (2\pi)^{-n/2} e^{-||x||_2^2/2} dx. \tag{0.21}$$

It is a classical result (see e.g. here for a proof) that

$$|\det(D((id \times \Psi) \circ \Phi)^{-1}(\rho, \phi_1, \dots, \phi_{n-1}))| = \rho^{n-1} \sin^{n-2}\phi_1 \sin^{n-3}\phi_2 \cdots \sin\phi_{n-2}, \tag{0.22}$$

so by the change of variables formula and Fubini-Tonelli's theorem

$$\int_{\Phi^{-1}(A_1 \times A_2)} (2\pi)^{-n/2} e^{-||x||_2^2/2} dx = \int_{A_1 \times \Psi(A_2)} (2\pi)^{-n/2} e^{-\rho^2/2} \rho^{n-1} \sin^{n-2}\phi_1 \dots \sin\phi_{n-2} d\rho d\phi_1 \cdots d\phi_{n-1} =$$
$$= \left(\int_{A_1} (2\pi)^{-n/2} S e^{-\rho^2/2} \rho^{n-1} d\rho \right) \left(\int_{\Psi(A_2)} \frac{1}{S} \sin^{n-2}\phi_1 \cdots \sin\phi_{n-2} d\phi_1 \cdots d\phi_{n-1} \right)$$

where

$$S := \int_{[0,\pi]^{n-2} \times [0,2\pi[} \sin^{n-2}\phi_1 \cdots \sin\phi_{n-2} d\phi_1 \cdots d\phi_{n-1} = \frac{2\pi^{n/2}}{\Gamma(n/2)} \tag{0.23}$$

is the volume of the $(n-1)$-dimensional sphere. From this it is easy to see that $(2\pi)^{-n/2} S e^{-\rho^2/2} \rho^{n-1}$ is the marginal probability density function for $||X||_2$, $(1/S)\sin^{n-2}\phi_1 \cdots \sin\phi_{n-2}$ is the marginal probability density function for $\Psi(Y)$, and $||X||_2$ and $\Psi(Y)$ are independent (because the joint probability density function is exactly the product of the marginal ones). But then also $||X||_2$ and $Y = \Psi^{-1}(\Psi(Y))$ (recalling that Ψ is bijective) are independent.

2. We can see that $\sqrt{n}Y \sim Unif(\sqrt{n}S^{n-1})$ from the fact that the probability density function

$$(1/S)\sin^{n-2}\phi_1 \cdots \sin\phi_{n-2}$$

found before for $\Psi(Y)$ (i.e. Y expressed in spherical coordinates) is exactly that of a uniform distribution on the sphere (since $\sin^{n-2}\phi_1 \cdots \sin\phi_{n-2} d\phi_1 \cdots d\phi_{n-1}$ is the volume element of the sphere). Another way to see it could be that $\sqrt{n}Y$ clearly takes values on $\sqrt{n}S^{n-1}$ and its distribution is rotation invariant. In fact, $Y := \frac{X}{||X||_2}$ and we know that X is rotation invariant (i.e. $X \sim UX$, where U is any orthogonal $(n \times n)$-matrix) because it has a standard normal distribution.

Exercise 9. We want to show that a random vector $X = (X_1, \dots, X_n)^T$ is sub-Gaussian if each X_i sub-Gaussian. By definition X is sub-Gaussian if $\langle X, x \rangle := \sum_{i=1}^n x_i X_i$ is sub-Gaussian for every $x \in \mathbb{R}^n$. Now, for any such x, the random variable $\sum_{i=1}^n x_i X_i$ is sub-Gaussian because it is a linear combination of sub-Gaussian random variables and we have seen that the product of a sub-Gaussian random variable by a scalar is sub-Gaussian and the sum of two sub-Gaussian random variables is sub-Gaussian. (These two facts can be deduced, for example, from the fact that a random variable is sub-Gaussian if and only if its sub-Gaussian norm is finite (prop. 2.5.2 of the textbook) and that the sub-Gaussian norm is indeed a norm (proved in the previous homework)). This concludes the proof.