Notes for High-Dimensional Probability: Random Vectors in High Dimensions

Isak Falk

April 20, 2021

1 Preliminaries

Proposition 1. *Let X be a real-values random variable and $Y = \sqrt{|X|}$. Then X is sub-exponential if and only if Y is sub-Gaussian, and in such case $\|X\|_{\psi_1} = \|Y\|_{\psi_2}^2$*

Proposition 2. *Let X be a real-valued sub-exponential variable, then the centered random variable $X - \mathbb{E}X$ is sub-exponential and*

$$\|X - \mathbb{E}X\|_{\psi_1} \leq (1 + \frac{2}{\ln 2})\|X\|_{\psi_2}. \tag{1}$$

Proof. The proof is analogous to the sub-Gaussian case but with different constant since the definition of the norm is different,

$$\|X - \mathbb{E}X\|_{\psi_1} \leq \|X\|_{\psi_1} + \|\mathbb{E}X\|_{\psi_1}, \tag{2}$$

and

$$\|\mathbb{E}X\|_{\psi_1} = \frac{|\mathbb{E}X|}{\ln 2} \leq \frac{\mathbb{E}|X|}{\ln 2} = \frac{\|X\|_1}{\ln 2} \leq \frac{2}{\ln 2}\|X\|_{\psi_1}, \tag{3}$$

where we have used the definition of sub-exponential norm for constant functions, Jensen and bound of L^p norm of X by sub-exponential norm. We thus have that $\|X - \mathbb{E}X\|_{\psi_1} \leq (1 + \frac{2}{\ln 2})\|X\|_{\psi_1}$ which is what we wanted to show.

□

Theorem 3 (Bernstein's theorem)**.** *Let $(X_i)_{i=1}^n$ be a sequence of independent real-valued zero-mean random variables such that $\|X_i\|_{\psi_1} < \infty$. Then, for every $t > 0$*

$$\mathbb{P}(|\frac{1}{n}\sum_{i=1}^n X_i| > t) \leq 2\exp(-cn\min(\frac{t^2}{K^2}, \frac{t}{K})), \tag{4}$$

where $K = \max_i \|X_i\|_{\psi_1}$ and $c > 0$ is some absolute constant.

2 Concentration of the Norm

Theorem 4 (Concentration of the L_2 norm)**.** *Let $X = (X_1, \dots, X_n) \in \mathbb{R}^n$ be a random vector with independent sub-gaussian coordinates X_i that satisfy $\mathbb{E}X_i^2 = 1$. Then*

$$\big\| \|X\|_2 - \sqrt{n} \big\|_{\psi_2} \leq CK^2, \tag{5}$$

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

Proof. We first note that $K \geq 1$. By Jensen's Inequality we have that $\mathbb{E}\exp(\frac{X_i^2}{t^2}) \geq \exp(\frac{\mathbb{E}X_i^2}{t^2}) = \exp(t^{-2})$ and using $t = 1$ we see that $\mathbb{E}\exp(X_i^2) \geq e > 2$ so $\|X_i\|_{\psi_2} \geq 1$ for all $i \in \{1, \dots, n\}$. Since $K = \max_i \|X_i\|_{\psi_2} \geq 1$ we are done.

Now consider the quantity $\frac{1}{n}\|X\|_2^2 - 1$ which we can write as

$$\frac{1}{n}\|X\|_2^2 - 1 = \frac{1}{n}\sum_{i=1}^{n}(X_i^2 - 1) = \frac{1}{n}\sum_{i=1}^{n} Y_i, \tag{6}$$

where $Y_i = X_i^2 - 1$. Since $\mathbb{E}X_i^2 = 1$ for any i, $(Y_i)_{i=1}^n$ is a vector of zero-centred random variables. Since $\|X_i\|_{\psi_2} < \infty$ we can show that $\|Y_i\|_{\psi_1} < \infty$ since

$$\begin{aligned} \big\|X_i^2 - 1\big\|_{\psi_1} &\leq C\big\|X_i^2\big\|_{\psi_1} && (7)\\ &= C\|X_i\|_{\psi_2}^2 && (8)\\ &\leq CK^2, && (9) \end{aligned}$$

using centring property of sub-exponentials proposition 2 (noting that $C > 1$), proposition 1 and definition of K. Since $(Y_i)_{i=1}^n$ are independent, zero-mean and by the above calculation sub-exponential, we can apply Bernstein's Inequality Thm. 3 which for any $u \geq 0$ means that

$$\mathbb{P}(|\frac{1}{n}\sum_{i=1}^{n} Y_i| > u) \leq 2\exp(-cn\min(\frac{u^2}{C^2K^4}, \frac{u}{CK^2})) \tag{10}$$

for some $c > 0$. Since $C, K > 1$ we have that $C^2 > C$ and $K^4 > K^2$, so we can write

$$\exp(-cn\min(\frac{u^2}{C^2K^4}, \frac{u}{CK^2})) \leq \exp(-\frac{cn}{C^4K^4}\min(u^2, u)). \tag{11}$$

So far we have proved a concentration bound on $\|X\|_2^2$. We will finish the proof by relating $\|X\|_2$ to $\|X\|_2^2$. Note the following; for any $z, \delta \geq 0$

$$|z - 1| \geq \delta \Rightarrow \big|z^2 - 1\big| \geq \max(\delta, \delta^2). \tag{12}$$

We show it by first noting that $\left|z^2 - 1\right| = |z-1||z+1| \geq |z+1|\delta$, and see that if $\delta \in [0,1)$ then $|z+1| \geq \delta$ and if $\delta \geq 1$ then since $|z+1| \geq 1$ we have that $\left|z^2+1\right| \geq \delta$. We can write this compactly as (12).

Now, consider any $\delta \geq 0$ and the expression $\left|\frac{1}{\sqrt{n}}\|X\|_2 - 1\right| \geq \delta$. Using (12) with $z = \frac{1}{\sqrt{n}}\|X\|_2$ we see that

$$\left|\frac{1}{\sqrt{n}}\|X\|_2 - 1\right| \geq \delta \Rightarrow \left|\frac{1}{n}\|X\|_2^2 - 1\right| \geq \max(\delta, \delta^2). \tag{13}$$

In terms of events, this means that

$$\mathbb{P}(|\frac{1}{\sqrt{n}}\|X\|_2 - 1| \geq \delta) \leq \mathbb{P}(|\frac{1}{n}\|X\|_2^2 - 1| \geq \max(\delta, \delta^2)) \leq 2\exp(-\frac{cn}{C^4K^4}\delta^2), \tag{14}$$

where in the final inequality we have used (11) together with

$$\delta^2 = \min(\max(\delta, \delta^2), \max(\delta, \delta^2)^2). \tag{15}$$

Letting $t = \delta\sqrt{n}$ we obtain the bound

$$\mathbb{P}(\left|\|X\|_2 - \sqrt{n}\right| \geq t) \leq 2\exp(-\frac{ct^2}{C^4K^4}), \tag{16}$$

for any $t \geq 0$ and this is equal to the conclusion of the theorem. $\square$

Remark 1. *The above bound tells us that with high probability, X takes values very close to the sphere of radius $\sqrt{n}$. In particular, for a fixed probability X stays within a constant distance from that sphere independently of the dimension n. This is due to the fact that $\|X\|_2^2$ has mean n and standard deviation $O(\sqrt{n})$ since*

$$\mathbb{V}(\|X\|_2^2) = \sum_{i=1} \mathbb{V}(X_i^2) = n\mathbb{V}(X_1^2), \tag{17}$$

due to independence of the coordinates of X, and so $\sqrt{\mathbb{V}(\|X\|_2^2)} = \sqrt{n} \cdot \mathrm{std}(X_1^2)$. $\sqrt{n \pm O(\sqrt{n})} = \sqrt{n} \pm O(1)$ since by Taylor expansion around $\sqrt{n}$ on the interval $[n - c\sqrt{n}, n + c\sqrt{n}]$ where c need to be chosen so that $n - c\sqrt{n} \geq 0$ we see that

$$\sqrt{n \pm c\sqrt{n}} = \sqrt{n} + R_1(c\sqrt{n}) \tag{18}$$

where $R_1(x) =$ [Need to fill in].

3 Covariance, second moments, whitening and isotropy

Definition 1. *The covariance matrix of a random vector $X \in \mathbb{R}^n$ is*

$$\mathrm{Cov}(X) = \mathbb{E}(X - \mathbb{E}X)(X - \mathbb{E}X)^\top = \mathbb{E}XX^\top - \mathbb{E}X(\mathbb{E}X)^\top. \tag{19}$$

Definition 2. *The second moment matrix of a random vector $X \in \mathbb{R}^n$ is*

$$\Sigma = \Sigma(X) = \mathbb{E}XX^\top. \tag{20}$$

Note that if $\mathbb{E}X = 0$, then $\mathrm{Cov}(X) = \Sigma(X)$. So we can mostly consider the second moment matrix without loss of generality in place of the covariance as long as we center our variable X. Both of the matrices are symmetric and positive semi-definite.

Definition 3. *A random vector $X \in \mathbb{R}^n$ is called* isotropic *if*

$$\Sigma(X) = \mathbb{E}XX^\top = I. \tag{21}$$

For a random variable $X \in \mathbb{R}^n$ with covariance matrix Σ and mean μ we can always put it into a isotropic form by whitening $Z = \Sigma^{-1/2}(X - \mu)$. Similarly, the transformation using a psd matrix Σ, $X = \mu + \Sigma^{1/2}Z$ of a mean-zero isotropic random vector $Z \in \mathbb{R}^n$ leads to a random vector with mean μ and covariance Σ. This observation means that in many cases we may focus on random isotropic mean-zero random vectors without loss of generality.

Lemma 5. *A random vector $X \in \mathbb{R}^n$ is isotropic if and only if*

$$\mathbb{E}\langle X, u\rangle = \|u\|^2 \tag{22}$$

for all $u \in \mathbb{R}^n$. Equivalently we could have used

$$\mathbb{E}\langle X, u\rangle = 1 \tag{23}$$

for all $u \in S^{n-1}$, the n-dimensional unit sphere.

Proof. Two symmetric matrices A, B are equal if and only if $x^\top Ax = x^\top B$ for any $x \in \mathbb{R}^n$. Thus X is isotropic if and only if

$$u^\top \mathbb{E}XX^\top u = \mathbb{E}\langle X, u\rangle = u^\top I u = \|u\|_2^2 \tag{24}$$

for all $u \in \mathbb{R}^n$ which is what we wanted to show. $\square$

Lemma 6. *Let X be an isotropic random vector in $\mathbb{R}^n$. The*

$$\mathbb{E}\|X\|_2^2 = n. \tag{25}$$

Moreover, if X, Y are two independent isotropic random vectors in $\mathbb{R}^n$, then

$$\mathbb{E}\langle X, Y\rangle^2 = n \tag{26}$$

Proof. First we have

$$\begin{aligned}
\mathbb{E}\|X\|_2^2 &= \mathbb{E}X^\top X && (27)\\
&= \mathbb{E}\mathrm{Tr}(XX^\top) && (28)\\
&= \mathrm{Tr}(\mathbb{E}XX^\top) && (29)\\
&= \mathrm{Tr}(I) && (30)\\
&= n. && (31)
\end{aligned}$$

For the second part we use the so called *law of total expectation* to write

$$\mathbb{E}\langle X, Y\rangle = \mathbb{E}_X\mathbb{E}_Y(\langle x, Y\rangle | X = x) = \mathbb{E}_X\|X\|_2^2 = n \tag{32}$$

by the Lemma 5 and reusing the first part of the proof. □

Corollary 7. *Let X, Y be independent mean-zero isotropic random vectors in R^n, then*

$$\mathbb{E}\|X - Y\|_2^2 = 2n \tag{33}$$

Proof.

$$\begin{aligned}
\mathbb{E}\|X - Y\|_2^2 &= \mathbb{E}\|X\|_2^2 + \mathbb{E}\|Y\|_2^2 - 2\mathbb{E}\langle X, Y\rangle && (34)\\
&= 2n - 0 && (35)\\
&= 2n && (36)
\end{aligned}$$

□

4 Examples of high-dimensional distributions

Let $X \sim \mathrm{Unif}(\sqrt{n}S^{n-1})$ mean that the law of X is the uniform measure on the zero-centered sphere of radius $\sqrt{n}$ in n dimensions.

Theorem 8 (Isotropy of uniform random variable on $\sqrt{n}S^{n-1}$)**.**