Ch. 3: Random Vectors

A **random vector** is simply a vector,

$$X = (X_1, \ldots, X_n) \in \mathbb{R}^n$$

whose coordinates are random variables (on the same probability space).

It's difficult to visualize n dimensions. Let's start by looking at the length of X,

$$\|X\|_2 = \sqrt{X_1^2 + \cdots + X_n^2} \in \mathbb{R}.$$

Note. This is a random **variable**!

Q. How long should we expect X to be?

Suppose $\mathbb{E}\, X_i^2 = 1$ for $1 \leq i \leq n$.

Then we have

$$\mathbb{E}\|X\|_2^2 = \mathbb{E}(X_1^2 + \cdots + X_n^2) = \sum_{i=1}^{n} \mathbb{E} X_i^2 = n.$$

That is, we have $\|X\|_2^2 = n$ on average, which suggests that $\|X\|_2 \approx \sqrt{n}$.

Thus, we might expect that $\|X\|_2 - \sqrt{n}$ is small (with high probability):

Theorem 3.1.1 (Concentration of the norm)
Let $X = (X_1, \ldots, X_n)$ have independent, sub-gaussian coordinates, each with $\mathbb{E} X_i^2 = 1$. Then

$$P\left(\left| \|X\|_2 - \sqrt{n} \right| \geq t\right) \leq 2e^{-ct^2/K^4}.$$

Here $C > 0$ is an absolute constant, and

$$K = \max_i \|X\|_{\psi_2}$$

(maximum of the sub-gaussian norms).

Our strategy. i) Try to replace $\|X\|_2$ with $\|X\|_2^2$

ii) Use Bernstein's inequality (Corollary 2.8.3)

For i), we use the following.

Lemma. If $z, \delta \geq 0$, then $|z-1| \geq \delta$ implies that

$$|z^2 - 1| \geq \max\{\delta, \delta^2\}.$$

Proof. Exercise. Hint: write $z = 1 + \alpha$, so $\alpha \geq -1$. Then consider the cases where $|\alpha| \leq 1$ and $\alpha \geq 1$.

Proof of Theorem.

We have

$$\big|\, \|X\|_2 - \sqrt{n} \,\big| \geq t \implies \left| \frac{1}{\sqrt{n}} \|X\|_2 - 1 \right| \geq \frac{t}{\sqrt{n}}$$

$$\xRightarrow{\text{(Lemma)}} \left| \frac{1}{n} \|X\|_2^2 - 1 \right| \geq u,$$

with $u = \max\left\{ \frac{t}{\sqrt{n}}, \frac{t^2}{n} \right\}$.

Remark. Given two events A and B, if B always happens whenever A does, then $P(A) \leq P(B)$.
(monotonicity)

That is, we see that

$$P\left(\big|\, \|X\|_2 - \sqrt{n} \,\big| \geq t \right)$$

$$\leq P\left(\left| \frac{1}{n} \|X\|_2^2 - 1 \right| \geq u \right).$$

Thus, it suffices to bound this last probability from above.

Next, observe that

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

a sum of independent, mean zero sub-exponential random variables (since X_i is sub-gaussian; see Lemma 2.7.6).

The sub-exponential norm of $X_i - 1$ satisfies

$$\begin{aligned}\|X_i^2 - 1\|_{\psi_1} &\leq C\|X_i^2\|_{\psi_1} && \text{(centering)}\\ &= C\|X_i\|_{\psi_2}^2 && \text{(Lemma 2.7.6)}\\ &\leq CK^2 && \text{(by hypothesis)}\end{aligned}$$

If we set

$$L := \max_i \| X_i^2 - 1 \|_{\psi_1},$$

then we've just shown that $L \leq CK^2$.

Applying Bernstein's inequality, we find that

$$P\left(\left| \frac{1}{n} \sum_{i=1}^{n} (X_i^2 - 1) \right| \right)$$

$$\leq 2\exp\left(-c \cdot n \cdot \min\left\{ \frac{u}{L}, \frac{u^2}{L^2} \right\} \right)$$

[$c > 0$ is an absolute constant

For simplicity, we assume $c \gtrsim 1$.]

$$\leq 2\exp\left(-c \cdot n \cdot \min\left\{ \frac{u}{CK^2}, \frac{u^2}{C^2K^4} \right\} \right)$$

$$\leq 2\exp\left(-c \cdot n \cdot \min\left\{ \frac{u}{C^2K^4}, \frac{u^2}{C^2K^4} \right\} \right)$$

[provided that $K \gtrsim 1$]

$$\leq 2\exp\left(-\frac{c \cdot n}{C^2K^4}\cdot \min\{u, u\}\right)$$

$$= 2\exp\left(-\frac{\tilde{c}\cdot n}{K^4}\cdot\frac{t^2}{n}\right).$$

[After recalling that $u = \max\{\frac{t}{\sqrt{n}}, \frac{t^2}{n}\}$, verify that $\min\{u, u^2\} = \frac{t^2}{n}$]

It follows that

$$P\left(\left|\,\|X\|_2 - \sqrt{n}\,\right| \geq t\right) \leq 2e^{-\tilde{c}\cdot t^2/K^4},$$

which is exactly what we wanted to show. ■

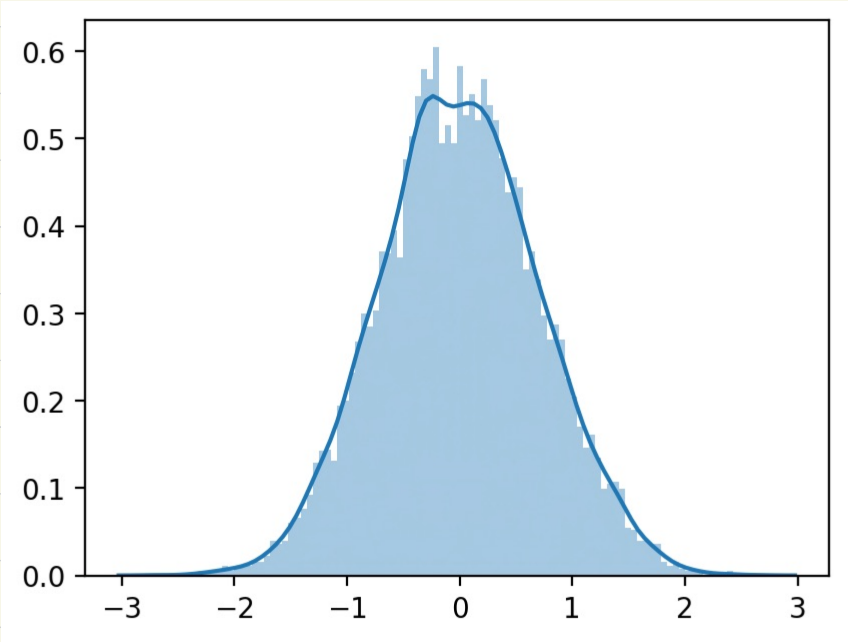
Recall that we assumed $K \geq 1$ in the proof above. Luckily, this turns out to be true! Why?
Use Jensen's inequality:

$$\mathbb{E}\exp(X_i^2/t^2) \geq \exp(\mathbb{E}(X_i^2/t)) = \exp\left(\frac{1}{t^2}\right),$$

[since $\mathbb{E}X_i^2 = 1$]

which implies that $\|X_i\|_{\psi_2} \gtrsim 1$ (check!).
Thus $K = \max_i \|X_i\|_{\psi_2} \gtrsim 1$.

The Theorem says that random vectors tend to cluster around the sphere of radius $\sqrt{n}$ centered at the origin.



The distribution of 10,000 samples of $\|X\|_2 - \sqrt{n}$ with $n = 5{,}000$.

Mean and Covariance

The **mean** of $X = (X_1, \ldots, X_n)$ is taken coordinate-wise:

$$\mathbb{E}X = (\mathbb{E}X_1, \ldots, \mathbb{E}X_n).$$

The higher-dimensional analogue of variance is the **covariance matrix** of X, given by

$$\operatorname{cov}(X) = \mathbb{E}(X-\mu)(X-\mu)^T,$$

[transpose]

where $\mu = \mathbb{E}X$. This is an $n \times n$ symmetric, **positive-semidefinite** matrix.

Def. A symmetric matrix A is called positive-semidefinite if

$$x^T A x \geq 0$$

for all $x \in \mathbb{R}^n$.

Positive-semidefinite matrices are special, as they have a unique (positive semi-definite) square root.

That is, if A is positive-semidefinite, then there is a unique positive-semidefinite matrix $B = B^T$ such that

$$A = BB = B^2.$$

We write $B = A^{\frac{1}{2}}$.

Note that

$$\operatorname{cov}(X) = \mathbb{E}(X-\mu)(X-\mu)^T$$

$$= \mathbb{E}\left(X \cdot X^T - \mu \cdot X^T - X \cdot \mu^T + \mu \cdot \mu^T\right)$$

$$= \mathbb{E}[X \cdot X^T] - \mu \cdot \mu^T - \mu \cdot \mu^T + \mu \cdot \mu^T$$

$$= \mathbb{E}[X \cdot X^T] - \mu \cdot \mu^T.$$

Compare this with $Var(Y) = \mathbb{E}[Y^2] - (\mathbb{E}Y)^2$.

We also define the **second moment matrix** of X as

$$\Sigma = \Sigma(X) := \mathbb{E}[X \cdot X^T],$$

and so $cov(X) = \Sigma - \mu \cdot \mu^T$.

Hence, if X has mean zero, then $cov(X) = \Sigma$.

Remark. The matrix Σ is also $n \times n$, symmetric, and positive-semidefinite.

Since Σ is a real, symmetric matrix, we can apply the **Spectral Theorem** to write

$$\Sigma = U \cdot D \cdot U^T.$$

Here, $\mathbf{U}$ is an orthogonal matrix ($U^{-1} = U^T$), whose columns, $u_1, \ldots, u_n$ are linearly independent eigenvectors of Σ.

If s_i is the eigenvalue associated to u_i ($\Sigma \cdot u_i = s_i \cdot u_i$), then $\mathbf{D}$ is the diagonal matrix of eigenvalues:

$$D = \begin{pmatrix} s_1 & & & 0 \\ & s_2 & & \\ & & \ddots & \\ 0 & & & s_n \end{pmatrix}.$$

The spectral decomposition is sometimes written in terms of the eigenvectors:

$$\Sigma = \sum_{i=1}^{n} s_i \cdot u_i \cdot u_i^T.$$

It's also common to order the eigenvalues in descending order, according to size:

$$s_1 \geq s_2 \geq \cdots \geq s_n \geq 0.$$