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

\[ P \left( \left| \|X\|_2 - \sqrt{n} \right| \geq t \right) \leq 2 e^{-c t^2/n} \]
Here $C > 0$ is an absolute constant, and

$$K = \max_i \|x\|_2$$

(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 $\varepsilon, \delta > 0$, then $|\varepsilon - 1| \geq \delta$

implies that

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

**Proof.** Exercise. Hint: write $\varepsilon = 1 + \alpha$, so $\alpha \geq -1$. Then consider the cases where $|\alpha| \leq 1$ and $\alpha \geq 1$. 3
Proof of Theorem.

We have

\[ \left| \|x\|_2 - \sqrt{n} \right| \geq t \implies \left| \frac{1}{\sqrt{n}} \|x\|_2 - 1 \right| \geq \frac{t}{\sqrt{n}} \]

(Lemma)

\[ \implies \left| \frac{1}{n} \|x\|_2^2 - 1 \right| \geq u, \]

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

\[ P\left( \left| \|x\|_2 - \sqrt{n} \right| \geq t \right) \leq P\left( \left| \frac{1}{n} \|x\|_2^2 - 1 \right| \geq u \right). \]
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

$$\|X_i^2 - 1\|_{\psi_1} \leq C \|X_i^2\|_{\psi_1} \quad \text{(centering)}$$

$$= C \|X_i\|_{\psi_2}^2 \quad \text{(Lemma 2.7.6)}$$

$$\leq CK^2 \quad \text{(by hypothesis)}$$
If we set
\[ L := \max_i \| X_i^2 - 1 \|_{\psi_i}, \]
then we've just shown that \( L \leq CK^2 \).

Applying Bernstein's inequality, we find that
\[
P \left( \left| \frac{1}{n} \sum_{i=1}^{n} (X_i^2 - 1) \right| \right) 
\leq 2 \exp \left( -c \cdot n \cdot \min \left\{ \frac{u}{L}, \frac{u^2}{L^2} \right\} \right)
\]
\[ \text{[ } c > 0 \text{ is an absolute constant} \]
\[ \text{For simplicity, we assume } c \geq 1. \]
\[
\leq 2 \exp \left( -c \cdot n \cdot \min \left\{ \frac{u}{CK^2}, \frac{u^2}{C^2 K^4} \right\} \right)
\]
\[
\leq 2 \exp \left( -c \cdot n \cdot \min \left\{ \frac{u}{CK^4}, \frac{u^2}{C^2 K^4} \right\} \right)
\]
\[ \text{[ provided that } K \geq 1 \] \]
\[
\leq 2 \exp \left( - \frac{c \cdot n}{C^2 K^4} \cdot \min \{u, u^3\} \right)
\]
\[
= 2 \exp \left( - \frac{\tilde{c} \cdot n}{K^4} \cdot \frac{t^2}{n} \right).
\]

After recalling that \( u = \max \{\frac{t}{\sqrt{n}}, \frac{t^2}{n}\} \), verify that \( \min \{u, u^3\} = \frac{t^2}{n} \).

It follows that
\[
P \left( \left| \sum X_i - \frac{t}{\sqrt{n}} \right| \geq t \right) \leq 2 \exp \left( - \frac{\tilde{c} \cdot t^2}{K^4} \right),
\]
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:
\[
E \exp \left( X_i^2 / t^2 \right) \geq \exp \left( E \left( X_i^2 / t \right) \right) = \exp \left( \frac{1}{t^2} \right),
\]
\[
\text{[since } E X_i^2 = 1 \text{]}.
\]

7
which implies that \( \|X_i\|_{\psi_2} \geq 1 \) (check!). Thus \( K = \max_i \|X_i\|_{\psi_2} \geq 1 \).

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:

\[
E X = (EX_1, \ldots, EX_n).
\]

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

\[
cov(X) = E (X - \mu)(X - \mu)^T.
\]

where \( \mu = EX \). This is an \( nxn \) 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

$$\text{cov}(x) = \mathbb{E} (x-\mu)(x-\mu)^T$$

$$= \mathbb{E} (x.x^T - \mu.x^T - x.\mu^T + \mu.\mu^T)$$

$$= \mathbb{E}[x.x^T] - \mu.\mu^T - \mu.\mu^T + \mu.\mu^T$$

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

We also define the \textbf{second moment matrix} of \( X \) as

\[
\Sigma = \Sigma(X) := E[X \cdot X^T],
\]

and so \( \text{cov}(X) = \Sigma - \mu \cdot \mu^T \).

Hence, if \( X \) has mean zero, then \( \text{cov}(X) = \Sigma \).

\textbf{Remark.} The matrix \( \Sigma \) is also \( n \times n \)

symmetric, and positive-semidefinite.

Since \( \Sigma \) is a real, symmetric matrix, we can apply the \textbf{Spectral Theorem} to write

\[
\Sigma = U \cdot D \cdot U^T.
\]
Here, $\mathbf{U}$ is an orthogonal matrix ($\mathbf{U}^{-1} = \mathbf{U}^T$), whose columns, $\mathbf{u}_1, \ldots, \mathbf{u}_n$ are linearly independent eigenvectors of $\Sigma$.

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

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

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

$$
\Sigma = \sum_{i=1}^{n} s_i \cdot \mathbf{u}_i \cdot \mathbf{u}_i^T.
$$

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

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