

High-Dimensional Probability: Answers, Theorems, and Definitions

Last revised on June 10, 2021

- Companion notes for *High-Dimensional Probability*, by Roman Vershynin. Link to book (PDF available online): www.math.uci.edu/~rvershyn/papers/HDP-book/HDP-book.html.
- **Disclaimer:** These notes compile my answers to the exercises, and lift the required theorems and definitions from the book. I wrote these notes to aid my personal study of the book. Read them at your own risk!*

Contents

0	Appetizer: Using probability to cover a geometric set	2
1	Preliminaries on random variables	5
1.1	Basic quantities	5
1.2	Some classical inequalities	5
1.3	Limits theorems	6

*Scribe: Alex Bie, alexbie98@gmail.com.

0 Appetizer: Using probability to cover a geometric set

A point $x \in \mathbb{R}^n$ is a **convex combination** of points $x_1, \dots, x_m \in \mathbb{R}^n$ if

$$x = \sum_{i=1}^m \lambda_i x_i \quad \text{with each } \lambda_i \geq 0 \quad \text{and} \quad \sum_{i=1}^m \lambda_i = 1.$$

The **convex hull** of $T \subseteq \mathbb{R}^n$, $\text{conv}(T)$, is the set of all convex combinations of T .

Theorem 0.0.1 (Carathéodory's Theorem). Let $x \in \text{conv}(T)$. There exists $k \leq n + 1$ points $x_1, \dots, x_k \in T$ such that x is a convex combination of x_1, \dots, x_k .

The result says we can obtain any point in the convex hull of T using at most a dimension-dependent number of points. Let the **diameter** of a set T be defined as $\text{diam}(T) = \sup\{\|x - y\|_2 : x, y \in T\}$.

Theorem 0.0.2 (Approximate Carathéodory's Theorem). Let $\text{diam}(T) = 1$. Let $x \in \text{conv}(T)$. For any k , there exists k points $x_1, \dots, x_k \in T$ such that

$$\left\| x - \frac{1}{k} \sum_{j=1}^k x_j \right\|_2 \leq \frac{1}{\sqrt{k}}$$

Proof. Suppose $|T| = m$. WLOG we can assume T is bounded by 1 in $\|\cdot\|_2$. We write $x = \sum_{i=1}^m \lambda_i x_i$, and interpret λ_i as probabilities. We define the random variable

$$X = x_i \text{ with probability } \lambda_i$$

for $i = 1, \dots, m$. We can verify that $\mathbb{E}X = \sum_{i=1}^m \lambda_i x_i = x$. Taking $X_1, \dots, X_k \stackrel{\text{iid}}{\sim} X$. It remains to analyse the quantity $\mathbb{E}\|x - \frac{1}{k} \sum_{j=1}^k X_j\|_2^2$.

$$\begin{aligned} \mathbb{E} \left\| x - \frac{1}{k} \sum_{j=1}^k X_j \right\|_2^2 &\leq \frac{1}{k^2} \mathbb{E} \left\| \sum_{j=1}^k X_j - x \right\|_2^2 \\ &= \frac{1}{k^2} \sum_{j=1}^k \mathbb{E} \|X_j - x\|_2^2 && \text{(by Exercise 0.0.3 (a))} \\ &= \frac{1}{k} \mathbb{E} \|X - x\|_2^2 \end{aligned}$$

Applying the result of Exercise 0.0.3 (b), we obtain

$$\mathbb{E}\|X - x\|_2^2 = \mathbb{E}\|X\|_2^2 - \|\mathbb{E}X\|_2^2 \leq \mathbb{E}\|X\|_2^2 \leq 1$$

Plugging this in above, we obtain the desired bound in expectation, hence there must exist a realization of the X_j , x_1, \dots, x_k , such that the bound holds. \square

Exercise 0.0.3. Check the following identities for random vectors.

(a) Let X_1, \dots, X_k be independent, mean zero random vectors in \mathbb{R}^n . Show that

$$\mathbb{E} \left\| \sum_{j=1}^k X_j \right\|_2^2 = \mathbb{E} \sum_{j=1}^k \|X_j\|_2^2$$

Answer.

$$\begin{aligned}
\mathbb{E} \left\| \sum_{j=1}^k X_j \right\|_2^2 &= \sum_{i=1}^n \mathbb{E} \left(\sum_{j=1}^m X_j^{(i)} \right)^2 = \sum_{i=1}^n \text{Var} \left(\sum_{j=1}^m X_j^{(i)} \right) && \text{(by mean zero)} \\
&= \sum_{i=1}^n \sum_{j=1}^m \text{Var} \left(X_j^{(i)} \right) && \text{(by independence)} \\
&= \sum_{i=1}^n \sum_{j=1}^m \mathbb{E} \left(X_j^{(i)} \right)^2 && \text{(by mean zero)} \\
&= \mathbb{E} \sum_{j=1}^m \|X_j\|_2^2
\end{aligned}$$

□

Among other things, this result implies that the expected squared distance of a random walk (starting from the origin) is equal to sum of the expected squared distances of each step.

(b) Let X be a random vector in \mathbb{R}^n . Show that

$$\mathbb{E} \|X - \mathbb{E}X\|_2^2 = \mathbb{E} \|X\|_2^2 - \|\mathbb{E}X\|_2^2$$

Answer.

$$\begin{aligned}
\mathbb{E} \|X - \mathbb{E}X\|_2^2 &= \mathbb{E} \sum_{i=1}^n \left(X^{(i)} - (\mathbb{E}X)^{(i)} \right)^2 = \sum_{i=1}^n \text{Var}(X^{(i)}) = \sum_{i=1}^n \mathbb{E} \left(X^{(i)} \right)^2 - \left(\mathbb{E}X^{(i)} \right)^2 \\
&= \mathbb{E} \|X\|_2^2 - \|\mathbb{E}X\|_2^2
\end{aligned}$$

□

Corollary 0.0.4 (Covering polytopes by balls). Let $P \subseteq \mathbb{R}^n$ be a polytope with $\text{diam}(P) = 1$. Let m be the number of vertices of P . Let $\varepsilon > 0$. We can cover P with m^k balls of radius ε for $k \geq \lceil 1/\varepsilon^2 \rceil$.

Proof. Take T to be the vertex set of P . $|T| = m$. Note that for any $x \in P$, $x \in \text{conv}(T)$. By Theorem 0.0.2, taking $k \geq \lceil 1/\varepsilon^2 \rceil$, we can find $x_1, \dots, x_k \in T$ such that

$$\left\| x - \frac{1}{k} \sum_{j=1}^k x_j \right\| \leq \frac{1}{\sqrt{k}} \leq \varepsilon$$

The number of ball centres obtained from selecting a set of k points out of m with repetition is bounded by m^k (possibly repeating orders). Hence we have an ε -cover sufficient to cover P . □

Exercise 0.0.5 (Binomial coefficient inequality). Show that for $1 \leq r \leq n$

$$\left(\frac{n}{r} \right)^r \leq \binom{n}{r} \leq \sum_{k=0}^r \binom{n}{k} \leq \left(\frac{en}{r} \right)^r$$

Answer. For the first inequality, consider

$$\frac{\left(\frac{n}{r} \right)^r}{\binom{n}{r}} = \frac{\frac{n}{r} \cdot \frac{n}{r} \cdot \dots \cdot \frac{n}{r}}{\frac{n}{r} \cdot \frac{n-1}{r-1} \cdot \dots \cdot \frac{n-r+1}{1}} \leq 1 \cdot 1 \cdot \dots \cdot 1 = 1$$

The second inequality follows immediately. To justify the last inequality, write

$$\begin{aligned}
\left(\frac{en}{r}\right)^r &= e^r \cdot \left(\frac{n}{r}\right)^r = \sum_{k=0}^{\infty} \frac{r^k}{k!} \cdot \left(\frac{n}{r}\right)^r && \text{(Maclaurin series for } e^x\text{)} \\
&\geq \sum_{k=0}^r \frac{r^k}{k!} \cdot \left(\frac{n}{r}\right)^r \\
&= \sum_{k=0}^r \frac{n^k \cdot n^{r-k}}{k! \cdot r^{r-k}} \\
&\geq \sum_{k=0}^r \frac{n^k}{k!} && \text{(by } n \geq r\text{)} \\
&\geq \sum_{k=0}^r \binom{n}{k}
\end{aligned}$$

□

Exercise 0.0.6 (Improved covering). Show that in the setting of Corollary 0.0.4, for $k \geq \lceil 1/\varepsilon^2 \rceil$

$$(C + C\varepsilon^2 m)^k$$

balls suffice for a suitable constant C .

Answer. We can give a tighter bound than given in the proof of Corollary 0.0.4 on the number of ball centres obtained from selecting a set of k points out of m with repetition (since computing the mean of k is order-invariant with respect to input points). By the “stars-and-bars”[†] argument, this quantity is given by

$$\binom{m+k-1}{k-1}$$

Note that $\min\{k-1, m\} = k-1 \leq \min\{k, m-1\}$, so looking at row $m+k-1$ of Pascal’s triangle

$$\binom{m+k-1}{k-1} \leq \binom{m+k-1}{k}$$

Then, using Exercise 0.0.5

$$\binom{m+k-1}{k} \leq \left(\frac{e(m+k-1)}{k}\right)^k = \left(e \frac{k-1}{k} + e \frac{1}{k} m\right)^k \leq (e + e\varepsilon^2 m)^k$$

□

[†][https://en.wikipedia.org/wiki/Stars_and_bars_\(combinatorics\)](https://en.wikipedia.org/wiki/Stars_and_bars_(combinatorics))

1 Preliminaries on random variables

1.1 Basic quantities

The **expectation** of a random variable X is denoted as $\mathbb{E}X$, and **variance** is denoted as $\text{Var}(X) = \mathbb{E}(X - \mathbb{E}X)^2$. (We note that the expectation operator \mathbb{E} can be directly defined as the Lebesgue integral of the random variable (measurable function) $X : \Omega \rightarrow \mathbb{R}$ in the probability space $(\Omega, \mathcal{M}, \mathbb{P})$.)

The **moment generating function** of X is given by

$$M_X(t) = \mathbb{E}e^{tX} \quad \text{for all } t \in \mathbb{R}$$

The **p-th moment** of X is given by $\mathbb{E}X^p$. We also let $\|X\|_p = (\mathbb{E}X^p)^{\frac{1}{p}}$ denote the **p-norm** of X . For $p = \infty$, we have

$$\|X\|_\infty = \text{ess sup } X$$

recalling that the **essential supremum** of a function f is the "smallest value γ such that $\{\omega \in \Omega : |f(\omega)| > \gamma\}$ has measure 0".

From this, we can define the **L^p spaces**[†], given a probability space $(\Omega, \mathcal{M}, \mathbb{P})$

$$L^p = \{X : \|X\|_p < \infty\}$$

Results from measure and integration theory tell us that the $(L^p, \|\cdot\|_p)$ are complete. In the case of L^2 , we have that with the inner product

$$\begin{aligned} \langle X, Y \rangle &= \int_{\Omega} XY(\omega) \mu(\omega) \\ &= \mathbb{E}XY \end{aligned}$$

$(L^2, \langle \cdot, \cdot \rangle)$ is a Hilbert space. In this case we can express the **standard deviation** of X as $\sqrt{\text{Var}(X)} = \|X - \mathbb{E}X\|_2$, and the **covariance** of random variable X and Y as

$$\text{Cov}(X, Y) = \mathbb{E}(X - \mathbb{E}X)(Y - \mathbb{E}Y) = \langle X - \mathbb{E}X, Y - \mathbb{E}Y \rangle$$

In this setting, considering random variables as vectors in L^2 , the covariance between X and Y can be interpreted as the *alignment* between the vectors $X - \mathbb{E}X$ and $Y - \mathbb{E}Y$.

1.2 Some classical inequalities

We say $f : \mathbb{R} \rightarrow \mathbb{R}$ is **convex** if

$$f((1-t)x + ty) \leq (1-t)f(x) + tf(y) \quad \text{for all } x, y \in \mathbb{R} \text{ and } t \in [0, 1]$$

Jensen's inequality states that for any random variable X and a convex function f , we get

$$f(\mathbb{E}X) \leq \mathbb{E}(f(X))$$

A corollary of Jensen's inequality implies that[§]

$$\|X\|_p \leq \|X\|_q \quad \text{for all } 1 \leq p \leq q \leq \infty$$

Minkowski's inequality asserts that the triangle inequality holds for the L_p spaces

$$\|X + Y\|_p \leq \|X\|_p + \|Y\|_p \quad \text{for all } X, Y \in L^p$$

In L^2 , we have the **Cauchy-Schwarz inequality**, which states that $|\mathbb{E}XY| \leq \mathbb{E}|XY| \leq \|X\|_2\|Y\|_2$. **Holder's inequality** additionally asserts that for $1/p + 1/q = 1$

$$|\mathbb{E}XY| \leq \|XY\|_1 \leq \|X\|_p\|Y\|_q$$

[†]A technical note is that the objects of L_p are actually equivalence classes of functions $[X]$ with equality almost everywhere, otherwise $\|\cdot\|_p$ is only a semi-norm.

[§]For $q < \infty$, the result follows by applying Jensen's inequality for $f(x) = x^{\frac{q}{p}}$. Otherwise, $\|X\|_\infty = \gamma = (\mathbb{E}\gamma^p)^{\frac{1}{p}} = \|\gamma\|_p \geq \|X\|_p$.

which also holds for $p = 1, q = \infty$.

The **cumulative distribution function** of X is defined as

$$F_X(t) = \mathbb{P}\{X \leq t\} = \mathbb{P}(X^{-1}(-\infty, t]) \quad \text{for all } t \in \mathbb{R}$$

and we refer to $\mathbb{P}\{X > t\} = 1 - F_X(t)$ as the **tail** of X .

Lemma 1.2.1 (Integral identity). Let $X \geq 0$ be a random variable. Then

$$\mathbb{E}X = \int_0^\infty \mathbb{P}\{X > t\} dt$$

with left side $= \infty$ iff right side $= \infty$.

Exercise 1.2.2 (Generalization of integral identity). Show that Lemma can be extended to be valid for any X

$$\mathbb{E}X = \int_0^\infty \mathbb{P}\{X > t\} dt - \int_{-\infty}^0 \mathbb{P}\{X < t\} dt$$

Answer. For not necessary non-negative X , $\mathbb{E}X := \mathbb{E}X^+ - \mathbb{E}X^-$ when they exist and are both $< \infty$, where

$$X^+ = \begin{cases} X & \text{if } X \geq 0 \\ 0 & \text{otherwise} \end{cases} \quad X^- = \begin{cases} -X & \text{if } X \leq 0 \\ 0 & \text{otherwise} \end{cases}$$

Applying Lemma 1.2.1 to the terms yields the result. For the second term

$$\mathbb{E}X^- = \int_0^\infty \mathbb{P}\{X^- > t\} dt = \int_0^\infty \mathbb{P}\{X < -t\} dt = \int_{-\infty}^0 \mathbb{P}\{X < t\} dt$$

□

Exercise 1.2.3 (p -th moment via the tail). Let X be a random variable and $0 < p < \infty$. Show that

$$\mathbb{E}|X|^p = \int_0^\infty pt^{p-1} \mathbb{P}\{|X| > t\} dt$$

whenever the right side is $< \infty$.

Answer. On the right side, substitute $u = t^p$, so $du = pt^{p-1} dt$ and

$$\int_0^\infty pt^{p-1} \mathbb{P}\{|X| > t\} dt = \int_0^\infty \mathbb{P}\{|X| > u^{\frac{1}{p}}\} du = \int_0^\infty \mathbb{P}\{|X|^p > u\} du = \mathbb{E}|X|^p$$

where the last equality comes from applying Lemma 1.2.1 to the random variable $|X|^p \geq 0$.

□

Proposition 1.2.4 (Markov's inequality). Let $X \geq 0$ with $\mathbb{E}X < \infty$. Then for $t > 0$

$$\mathbb{P}\{X \geq t\} \leq \frac{\mathbb{E}X}{t}$$

Proof.

□

Corollary 1.2.5 (Chebyshev's inequality). Let X have $\mathbb{E}X < \infty$ and $\text{Var}(X) < \infty$. Then for $t > 0$

$$\mathbb{P}\{|X - \mathbb{E}X| > t\} \leq \frac{\text{Var}(X)}{t^2}$$

Exercise 1.2.6. Give a proof of Chebyshev's inequality using Markov's inequality.

Answer. The random variable $|X - \mathbb{E}X|^2$ is well-defined (by $\mathbb{E}X < \infty$), non-negative, with finite expectation. Applying Markov's inequality with $t^2 > 0$ yields

$$\mathbb{P}\{|X - \mathbb{E}X| \geq t\} = \mathbb{P}\{|X - \mathbb{E}X|^2 \geq t^2\} \leq \frac{\text{Var}(X)}{t^2}$$

□

1.3 Limits theorems