Notes for High-Dimensional Probability Second Edition by Roman Vershynin

Gallant Tsao

July 1, 2025

Contents

1 A Quick Refresher on Analysis and Probability

1.1 Convex Sets and Functions

Definition 1.1.1. A subset $K \subseteq \mathbb{R}^n$ is a convex set if, for any pair of points in K, the line segment connecting these two points is also contained in K, i.e.

$$\lambda x + (1-\lambda)y \in K \quad \forall x, y \in K, \lambda \in [0,1].$$

Let $K \in \mathbb{R}^n$ be a convex subset. A function $f : K \to \mathbb{R}$ is a convex function if

$$f(\lambda x + (1-\lambda)y) \leq \lambda f(x) + (1-\lambda)f(y) \quad \forall x, y \in K, \lambda \in [0,1].$$

f is concave if the inequality above is reversed, or equivalently, if $-f$ is convex.

1.2 Norms and Inner Products

Definition 1.2.1. The Euclidean norm of a vector $x \in \mathbb{R}^n$ is

$$\|x\|_2 = \left(\sum_{i=1}^n x_i^2\right)^{1/2}.$$

Definition 1.2.2. The inner product (dot product) of two vectors $x, y \in \mathbb{R}^n$ is

$$\langle x, y\rangle = x^T y.$$

Definition 1.2.3. For $p \in [1, \infty]$, the ℓ^p norm of a vector $x \in \mathbb{R}^n$ is

$$\|x\|_p = \left(\sum_{i=1}^n |x_i|^p\right)^{1/p} \quad \text{for } p \in [1,\infty), \;\; \|x\|_\infty = \max_{i=1,\dots,n} |x_i|.$$

Theorem 1.2.4 (Minkowski's inequality)**.** For any vector $x, y \in \mathbb{R}^n$,

$$\|x+y\|_p \leq \|x\|_p + \|y\|_p.$$

It follows that the ℓ^p norm defines a norm on $\mathbb{R}^n$ for every $p \in [1, \infty]$.

Theorem 1.2.5 (Cauchy-Schwartz inequality)**.** For all vectors $x, y \in \mathbb{R}^n$,

$$|\langle x, y\rangle| \leq \|x\|_2 \|y\|_2.$$

Theorem 1.2.6 (Hölder's inequality)**.** For all vectors $x, y \in \mathbb{R}^n$,

$$|\langle x, y\rangle| \leq \|x\|_p \|y\|_{p'} \text{ if } \frac{1}{p} + \frac{1}{p'} = 1$$

where p, p' are called conjugate exponents.

1.3 Random Variables and Random Vectors

We'll do a brief review of some important concepts about random variables first:

Definition 1.3.1. The expectation (mean) of a random variable X is

$$\mathbb{E}[X] = \sum_{k=-\infty}^{\infty} k p_X(k) = \int_{-\infty}^{\infty} x f_X(x)\, dx.$$

Its variance is

$$\mathrm{Var}(X) = \mathbb{E}[(X - \mathbb{E}[X])^2] = \mathbb{E}[X^2] - (\mathbb{E}[X])^2.$$

The expectation is linear:

$$\mathbb{E}[a_1 X_1 + \cdots + a_n X_n] = a_1 \mathbb{E}[X_1] + \cdots + a_n \mathbb{E}[X_n].$$

For variance this is not the case. However, if the random variables are independent (or even uncorrelated):

$$\mathrm{Var}(a_1 X_1 + \cdots + a_n X_n) = a_1^2 \mathrm{Var}(X_1) + \cdots + a_n^2 \mathrm{Var}(X_n).$$

The simplest example of a random variable is the *indicator* of a given event E, which is

$$\mathbf{1}_E(x) = \begin{cases} 1 & \text{if } x \in E, \\ 0 & \text{if } x \notin E. \end{cases}$$

Its expectation is given by

$$\mathbb{E}[\mathbf{1}_E] = P(E).$$

Definition 1.3.2. The moment generating function (mgf) of a random variable X is given by

$$M_X(t) = \mathbb{E}[e^{tX}], t \in \mathbb{R}.$$

Definition 1.3.3. For $p > 0$, the pth moment of a random variable X is $\mathbb{E}[X^p]$, and the pth absolute moment is $\mathbb{E}[|X|^p]$. By taking the pth root of the absolute moment, we get the L^p norm of a random variable:

$$\|X\|_{L^p} = (\mathbb{E}[|X|^p])^{1/p}, \text{ and } \|X\|_\infty = \operatorname{ess\,sup} |X|,$$

where esssup denotes the essential supremum.
The normed space consisting of all random variables on a given probability space that have finite L^p norm is called the L^p space:

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

Definition 1.3.4. The standard deviation of a random variable X is

$$\sigma = \sqrt{\mathrm{Var}(X)} = \|X - \mathbb{E}[X]\|_{L^2}.$$

The covariance of two random variables X and Y is

$$\mathrm{Cov}(X, Y) = \mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])] = \langle X - \mathbb{E}[X], Y - \mathbb{E}[Y] \rangle_{L^2}.$$

Definition 1.3.5. A random vector $X = (X_1, \ldots, X_n)$ is a vector whose all n coordinates X_i are random variables. Its expected value is

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

Its covariance matrix is

$$\mathrm{Cov}(X) = \mathbb{E}[(X - \mathbb{E}[X])(X - \mathbb{E}[X])^T].$$

which is a $n \times n$ matrix whose (i, j)-th entry is $\mathrm{Cov}(X_i, X_j)$.

1.4 Union Bound

Lemma 1.4.1 (Union bound)**.** For any events $E_1, \ldots, E_n$, we have

$$P\left(\bigcup_{i=1}^{n} E_i\right) \leq \sum_{i=1}^{n} P(E_i).$$

Proof. If the event $\cup_{i=1}^{n}$ occurs, at least of the events E_i has to occur. Therefore their respective indicator random variables satisfy

$$\mathbf{1}_{\cup_{i=1}^{n} E_i} \leq \mathbf{1}_{E_i}.$$

Taking expectations and using the linearity of expectation completes the proof. □

Example 1.4.2 (Dense random graphs have no isolated vertices)**.** Consider the $G(n, p)$ graph from the Erdos-Renyi model, with $n \geq 2$. Show that if $p \geq 4 \ln n/n$ then there are no isolated vertices with probability at least $1 - 1/n$.

Proof. Call the vertices $1, \ldots, n$ and let E_i denote the event that vertex has no neighbors. This means that none of the other $n-1$ vertices are neighbors with vertex i, and these $n-1$ events are independent and have probability $1-p$ each. Thus $P(E_i) = (1-p)^{n-1}$.
Therefore, by union bound, we have

$$\begin{aligned} P\left(\bigcup_{i=1}^{n} E_i\right) &\leq \sum_{i=1}^{n} P(E_i) \\ &= n(1-p)^{n-1}. \end{aligned}$$

□

1.5 Conditioning

Definition 1.5.1. Given a probability space, the <u>conditional probability</u> of an event E given an event F is

$$P(E|F) = \frac{P(E \cap F)}{P(F)}.$$

1.6 Probabilistic Inequalities

Theorem 1.6.1 (Jensen's Inequality)**.** For any random variable X and a convex function $f : \mathbb{R} \to \mathbb{R}$,

$$f(\mathbb{E}[X]) \leq \mathbb{E}[f(X)].$$

This also holds for any random vector taking values in $\mathbb{R}^n$ and any convex function $f : \mathbb{R}^n \to \mathbb{R}$.

In particular, since any norm on $\mathbb{R}^n$ is convex, we get

$$\|\mathbb{E}[X]\| \leq \mathbb{E}[\|X\|].$$

Theorem 1.6.2 (Inequalities for random variables)**.** Minkowski inequality states that for any $p \in [1, \infty]$ and any random variables $X, Y \in L^p$,

$$\|X + Y\|_{L^p} \leq \|X\|_{L^p} + \|Y\|_{L^p}.$$

1.7 Limit Theorems

Theorem 1.7.1 (Strong law of large numbers)**.** Let $X_1, X_2, \ldots$ be a sequence of i.i.d. random variables with mean μ. Let $S_N = X_1 + \cdots + X_N$. Then as $N \to \infty$,

$$\frac{S_N}{N} \to \mu \text{ almost surely.}$$

Definition 1.7.2. A random variable X is a standard normal random variable, denoted $X \sim N(0,1)$, if its density is

$$f_X(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}, x \in \mathbb{R}.$$

X has mean zero and variance 1.
More generally, X as a normal distribution with mean μ and variance σ^2 if its density is

$$f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, x \in \mathbb{R}.$$

Theorem 1.7.3 (Lindeberg–Lévy CLT)**.** Let $X_1, X_2, \ldots$ be a sequence of i.i.d. random variables with mean μ and variance σ^2. Consider the sum $S_N = X_1 + \cdots + X_N$. Normalize this sum so that it has zero mean and unit variance:

$$Z_N := \frac{S_N - \mathbb{E}[S_N]}{\sqrt{\mathrm{Var}(S_N)}} = \frac{1}{\sigma\sqrt{N}} \sum_{i=1}^{N} (X_i - \mu).$$

Then as $N \to \infty$,

$$Z_N \to N(0,1) \text{ in distribution,}$$

meaning the CDF of Z_N converges pointwise to the CDF of $N(0,1)$.

Example 1.7.4 (Bernoulli and binomial distributions)**.** When $X_i \sim \mathrm{Ber}(p)$, $S_N \sim \mathrm{Binom}(N,p)$. In particular, theorem 1.7.3 gives us

$$\frac{S_N - Np}{\sqrt{Np(1-p)}} \to N(0,1) \text{ in distribution.}$$

The special case above is called the de Moivre-Laplace theorem.

There is also a version of the CLT used for the Poisson distribution, when $p \to 0$ for the Bernoulli random variables:

Definition 1.7.5. A random variable X has the Poisson distribution with parameter $\lambda > 0$, denoted $X \sim \mathrm{Pois}(\lambda)$, if

$$P(X = k) = e^{-\lambda} \frac{\lambda^k}{k!}, \ k \in \mathbb{N}_0.$$

Theorem 1.7.6 (Poisson limit theorem)**.** Consider independent random variables $X_{N,i}, p_{N,i}$ for $N \in \mathbb{N}$ and $1 \leq i \leq N$. Let

$$S_N = X_{N,1} + \cdots + X_{N,N}.$$

Assume that as $N \to \infty$,

$$\max_{i \leq N} p_{N,i} \to 0 \text{ and } \mathbb{E}[S_N] = \sum_{i=1}^{N} p_{N,i} \to \lambda < \infty.$$

Then as $N \to \infty$,
$$S_N \to \mathrm{Pois}(\lambda) \text{ in distribution.}$$

To approximate the Poisson distributions, we often have to deal with factorials. Here are a few useful tools for approximations:

Lemma 1.7.7 (Stirling approximation)**.**

$$n! = \sqrt{2\pi n}\left(\frac{n}{e}\right)^n (1 + o(1)) \text{ as } n \to \infty.$$

In particular, for $X \sim \mathrm{Pois}(\lambda)$,

$$P(Z = k) = \frac{e^{-\lambda}}{\sqrt{2\pi k}}\left(\frac{e\lambda}{k}\right)^k (1 + o(1)) \text{ as } k \to \infty.$$

Of course, there are also non-asymptotic results:

Lemma 1.7.8 (Bounds on the factorial)**.** For any $n \in \mathbb{N}$, we have

$$\left(\frac{n}{e}\right)^n \le n! \le en\left(\frac{n}{e}\right)^n.$$

Proof. For the lower bound, we use the Taylor series for e^x and drop all terms except the nth one, which gives
$$e^x \ge \frac{x^n}{n!}.$$
Substitute $x = n$ and rearranging gives the inequality.
For the upper bound,

$$\ln(n!) \le \sum_{k=1}^{n} \ln k \le \int_1^n \ln x \; dx + \ln n = n(\ln n - 1) + 1 + \ln n.$$

Exponentiating and rearranging gives the upper bound. □

Remark 1.7.9 (Gamma function)**.** The gamma function extends the notion of the factorial to all real numbers, even to all complex numbers with positive real part. It is defined as

$$\Gamma(z) := \int_0^\infty t^{z-1} e^{-t} \; dt.$$

Repeated integration by parts gives

$$\Gamma(n + 1) = n!, \; n \in \mathbb{N}_0.$$

Stirling approximation (lemma 1.7.7) is also valid for the gamma function:

$$\Gamma(z) = \sqrt{2\pi z}\left(\frac{z}{e}\right)^z (1 + o(1)) \text{ as } z \to \infty.$$