High Dimensional Probability Notes

Nong Minh Hieu[1]

[1] School of Physical and Mathematical Sciences, Nanyang Technological University (NTU - Singapore)

Contents

1 Random variables

1.1 Basic inequalities

First, we revisit the definition of a random variable as well as some basic inequalities that we learned in introductory statistics.

Definition 1.1 (Random variable). ─────────────────────────────────────

Let $(\Omega, \Sigma, \mathbb{P})$ be a probability space. A random variable X is defined as a mapping from the sample space Ω to $\mathbb{R}$:

$$X : \Omega \to \mathbb{R} \tag{1}$$

Σ is the σ-algebra containing the possible events (collection of subsets of Ω) and $\mathbb{P}$ is a probability measure that assigns events with probabilities:

$$\mathbb{P} : \Sigma \to [0, 1] \tag{2}$$

For a given probability space $(\Omega, \Sigma, \mathbb{P})$ and a random variable $X : \Omega \to \mathbb{R}$, we will use the following basic notations throughout this note:

- $\|X\|_{L^p}$ - The pth root of the pth moment of the random variable X.

$$\|X\|_{L^p} = (\mathbb{E}|X|^p)^{1/p}, \ p \in (0, \infty) \tag{3}$$

$$\|X\|_{L^\infty} = \operatorname{ess\,sup} |X| \tag{4}$$

- $L^p(\Omega, \Sigma, \mathbb{P})$ - The space of random variables X satisfying:

$$L^p(\Omega, \Sigma, \mathbb{P}) = \left\{ X : \Omega \to \mathbb{R} \,\middle|\, \|X\|_{L^p} < \infty \right\} \tag{5}$$

Some basic inequalities and identities:

- **1. Jensen's Inequality** - For a random variable X and a convex function $\varphi : \mathbb{R} \to \mathbb{R}$, we have:

$$\varphi(\mathbb{E}X) \leqslant \mathbb{E}\varphi(X) \tag{6}$$

- **2. Monotonicity of L^p norm** - For a random variable X:

$$\|X\|_{L^p} \leqslant \|X\|_{L^q}, \ 0 \leqslant p \leqslant q \leqslant \infty. \tag{7}$$

- **3. Minkowski's Inequality** - For $1 \leqslant p \leqslant \infty$ and two random variables X, Y in $L^p(\Omega, \Sigma, \mathbb{P})$ space:

$$\|X + Y\|_{L^p} \leqslant \|X\|_{L^p} + \|Y\|_{L^p}. \tag{8}$$

- **4. Holder's Inequality** - For $p, q \in [1, \infty]$ such that $1/p + 1/q = 1$. Then, for random variables $X \in L^p(\Omega, \Sigma, \mathbb{P})$ and $Y \in L^q(\Omega, \Sigma, \mathbb{P})$, we have:

$$|\mathbb{E}XY| \leqslant \|X\|_{L^p} \cdot \|Y\|_{L^q}. \tag{9}$$

- **5. Markov's Inequality** - For a non-negative random variable X and $t > 0$, we have:

$$\mathbb{P}(X \geqslant t) \leqslant \frac{\mathbb{E}X}{t}. \tag{10}$$

We can also generalize Markov's Inequality for p^{th} moment:

$$\mathbb{P}(|X| \geqslant t) \leqslant \frac{\mathbb{E}[|X|^p]}{t^p}, \forall t > 0, k \in [2, \infty). \tag{11}$$

- **6. Chebyshev's Inequality** - For a random variable X with mean μ and variance σ^2. Then, for any $t > 0$, we have:

$$\mathbb{P}(|X - \mu| \geqslant t) \leqslant \frac{\sigma^2}{t^2}. \tag{12}$$

- **7. Integral Identity** - Let X be a non-negative random variable, we have:

$$\mathbb{E}X = \int_0^\infty \mathbb{P}(X > t)dt. \tag{13}$$

Exercises

Exercise 1.1.1: Generalized Integral Identity

Let X be a random variable (not necessarily non-negative). Prove the following identity:

$$\mathbb{E}X = \int_0^\infty \mathbb{P}(X > t)dt - \int_{-\infty}^0 \mathbb{P}(X < t)dt. \tag{14}$$

Solution (Exercise 1.1.1). _____

For $x \in \mathbb{R}$, using the basic integral indentity, we have:

$$|x| = \int_0^\infty \mathbf{1}\{t < |x|\}dt$$

We consider the following cases:

- *When $x < 0 \implies x = -|x|$:*

$$x = -\int_0^\infty \mathbf{1}\{t < |x|\}dt = -\int_0^\infty \mathbf{1}\{t < -x\}dt = -\int_0^\infty \mathbf{1}\{-t > x\}dt = -\int_{-\infty}^0 \mathbf{1}\{t > x\}dt.$$

- *When $x \geqslant 0 \implies x = |x|$:*

$$x = \int_0^\infty \mathbf{1}\{t < |x|\}dt = \int_0^\infty \mathbf{1}\{t < x\}dt.$$

Therefore, for $x \in \mathbb{R}$, we can write:

$$x = \int_0^\infty \mathbf{1}\{t < x\}dt - \int_{-\infty}^0 \mathbf{1}\{t > x\}dt.$$

Therefore, for a random variable X not necessarily non-negative, we have:

$$\mathbb{E}X = \mathbb{E}\left[\int_0^\infty \mathbf{1}\{t < X\}dt - \int_{-\infty}^0 \mathbf{1}\{t > X\}dt\right]$$

$$= \mathbb{E}\int_0^\infty \mathbf{1}\{t < X\}dt - \mathbb{E}\int_{-\infty}^0 \mathbf{1}\{t > X\}dt$$

$$= \int_0^\infty \mathbb{E}\mathbf{1}\{t < X\}dt - \int_{-\infty}^0 \mathbb{E}\mathbf{1}\{t > X\}dt$$

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

$\square$.

> **Exercise 1.1.2: p^{th}-moments via tails**
>
> Let X be a random variable and $p \in (0, \infty)$. Show that:
>
> $$\mathbb{E}|X|^p = \int_0^\infty pt^{p-1}\mathbb{P}(|X| > t)dt. \qquad (15)$$

Solution (Exercise 1.1.2). ─────────────────────────────────

Let X be a random variable that is not necessarily non-negative. Using the integral identity, we have:

$$\mathbb{E}|X|^p = \int_0^\infty \mathbb{P}(u < |X|^p)du.$$

Let $t^p = u \implies pt^{p-1}dt = du$. Since we integrate u from $0 \to \infty$, we also integrate t from $0 \to \infty$ when changing the variables. Hence, we have:

$$\mathbb{E}|X|^p = \int_0^\infty \mathbb{P}(t^p < |X|^p)pt^{p-1}dt = \int_0^\infty \mathbb{P}(t < |X|)pt^{p-1}dt.$$

Hence, we obtained the desired identity. □.

1.2 Limit Theorems

1.2.1 Weak Law of Large Numbers

> **Theorem 1.1: Weak Law of Large Numbers (WLLN)**
>
> Let $X_1, \ldots, X_N$ be *i.i.d* random variables with mean μ. Consider the sum:
>
> $$S_N = X_1 + \cdots + X_N$$
>
> Then, the sample mean **converges to μ in probability** $(S_N/N \xrightarrow{p} \mu)$:
>
> $$\lim_{N\to\infty} \mathbb{P}\Big(|S_N/N - \mu| > \epsilon\Big) = 0, \ \forall \epsilon > 0 \qquad (16)$$

Proof (Weak Law of Large Numbers (**WLLN**)). ─────────────────────────────────

We split the proof into two sections corresponding to the assumptions of finite variance and non-finite variance.

1. **Finite variance case:** *Suppose that $\mathrm{Var} X_i = \sigma^2 < \infty$ for all $1 \leqslant i \leqslant N$. Let $\bar{X} = S_N/N$. Then, $\bar{X}$ is a random variable with the following mean and variance:*

$$\mathbb{E}\bar{X} = \mu \quad and \quad \mathrm{Var}\bar{X} = \frac{\sigma^2}{N}.$$

Hence, by the Chebyshev's inequality, we have:

$$\mathbb{P}\Big(|S_N/N - \mu| > \epsilon\Big) = \mathbb{P}\Big(|\bar{X} - \mu| > \epsilon\Big) \leqslant \frac{\sigma^2}{N\epsilon^2}.$$

Therefore, we have:

$$\lim_{N\to\infty} \mathbb{P}\Big(|S_N/N - \mu| > \epsilon\Big) \leqslant \lim_{N\to\infty} \frac{\sigma^2}{N\epsilon^2} = 0.$$

*Hence, we have $\lim_{N\to\infty} \mathbb{P}\Big(|S_N/N - \mu| > \epsilon\Big) = 0$ and we obtained (**WLLN**).*

2. ***Non-finite variance case***: *In this case, we rely on the Levy Continuity Theorem* (**LCT**), *which relies on the convergence of the characteristic function. For $n \geqslant 1$, define the sequence of random variable $Y_n = S_n/n$. Hence, we have:*

$$\begin{aligned}
\varphi_{Y_n}(t) &= \varphi_{S_n/n}(t) \\
&= \varphi_{S_n}(t/n) \\
&= \prod_{i=1}^{n} \varphi_{X_i}(t/n) = \left[\varphi_X(t/n)\right]^n,
\end{aligned}$$

Where $X = X_1 = \cdots = X_n$. By Taylor's expansion, we have:

$$\varphi_X(t/n) = 1 + \frac{it\mathbb{E}[X]}{n} + \mathcal{O}(1/n^2) = 1 + \frac{it\mu}{n} + \mathcal{O}(1/n^2).$$

Hence, we have:

$$\lim_{n \to \infty} \varphi_{Y_n}(t) = \lim_{n \to \infty} \left(1 + \frac{it\mu}{n} + \mathcal{O}(1/n^2)\right)^n = e^{it\mu}.$$

Therefore, by (**LCT**), *we have $Y_n \xrightarrow{p} \mu$.*

Remark 1.1 (Taylor expansion of Moment Generating and Characteristic Functions). ———
Given a random variable X. For reference, the following are the Taylor expansions of the Moment Generating Function $M_X(t)$ and the Characteristic Function $\varphi_X(t)$:

$$\begin{aligned}
M_X(t) &= \mathbb{E}[e^{tX}] = 1 + \sum_{n=1}^{\infty} \frac{t^n}{n!}\mathbb{E}[X^n], \\
\varphi_X(t) &= \mathbb{E}[e^{itX}] = 1 + \sum_{n=1}^{\infty} \frac{(it)^n}{n!}\mathbb{E}[X^n].
\end{aligned} \tag{17}$$

For the sake of my laziness, here are the Taylor expansion for the first three terms of both the MGF and the CF:

$$\begin{aligned}
M_X(t) &= 1 + t\mathbb{E}[X] + \frac{t^2}{2}\mathbb{E}[X^2] + \mathcal{O}(t^3), \\
\varphi_X(t) &= 1 + it\mathbb{E}[X] - \frac{t^2}{2}\mathbb{E}[X^2] + \mathcal{O}(t^3).
\end{aligned} \tag{18}$$

□.

Theorem 1.2: Levy Continuity Theorem (**LCT**)

Let $X_1, X_2, \ldots$ be *i.i.d* random variables. Then:

$$\forall t \in \mathbb{R} : \lim_{n \to \infty} \varphi_{X_n}(t) = \varphi_X(t) \iff X_n \xrightarrow{d} X, \tag{19}$$

for some random variable X. In a special case where $X = c$ for some $c \in \mathbb{R}$, we have:

$$\forall t \in \mathbb{R} : \lim_{n \to \infty} \varphi_{X_n}(t) = e^{itc} \iff X_n \xrightarrow{p} c. \tag{20}$$

Proof (Levy Continuity Theorem (**LCT**)). ————————————————————
The proof for (**LCT**) *can be found in Gut 2004, Section 9.1, Theorem 9.1 and Collorary 9.1* □.

1.2.2 Strong Law of Large Numbers

Theorem 1.3: Strong Law of Large Numbers (SLLN)

Let $X_1, \ldots, X_N$ be *i.i.d* random variables with mean μ. Consider the sum:

$$S_N = X_1 + \cdots + X_N$$

Then, the sample mean **converges to μ almost surely** $(S_N/N \xrightarrow{a.s} \mu)$:

$$\mathbb{P}\left(\limsup_{N \to \infty} |S_N/N - \mu| > \epsilon \right) = 0, \ \forall \epsilon > 0 \tag{21}$$

Proof (Strong Law of Large Numbers (**SLLN**)). ────────────────

*For the sake of simplicity, we will present the proof for (**SLLN**) with an additional assumption that $\mathbb{E}[|X_n|^4] < \infty, \forall n \geqslant 1$. The proof for the general case of (**SLLN**) (also called the Kolmogorov Strong Law) can be found in Gut 2004, Section 6, Theorem 6.1. For convenience, we assume the following:*

1. *$\mathbb{E}[|X_n|^4] = K < \infty$.*

2. *$\mathbb{E}[X_n] = 0$. For non-zero mean case, we can set $Y_n = X_n - \mu$ and repeat the same arguments made below.*

We aim to prove that $\mathbb{P}\left(\limsup_{N \to \infty} |S_N/N| > \epsilon\right) = 0$ for any $\epsilon > 0$. Firstly, use the Multinomial formula to expand $\mathbb{E}[S_n]$. The expansion will contain the terms in the following forms:

$$X_i^2, X_i^3 X_j, X_i^2 X_j^2, X_i^2 X_j X_k, X_i X_j X_k X_\ell,$$

where i, j, k, ℓ are distinct indices. By independence, we have:

$$\mathbb{E}[X_i^3 X_j] = \mathbb{E}[X_i^2 X_j X_k] = \mathbb{E}[X_i X_l X_k X_\ell] = 0.$$

As a result, we have the following remaining terms by the Multinomial formula:

$$\mathbb{E}[S_n^4] = \sum_{i=1}^n \mathbb{E}[X_i^4] + \binom{4}{2} \sum_{1 \leqslant i < j \leqslant n} \mathbb{E}[X_i^2 X_j^2]$$

$$= \sum_{i=1}^n \mathbb{E}[X_i^4] + 6 \underbrace{\sum_{1 \leqslant i < j \leqslant n} \mathbb{E}[X_i^2 X_j^2]}_{n(n-1)/2 \ terms}$$

$$= nK + 3n(n-1)\mathbb{E}[X_i^2 X_j^2].$$

By independence, we have $\mathbb{E}[X_i^2 X_j^2] = \mathbb{E}[X_i^2]\mathbb{E}[X_j^2]$ and for any $1 \leqslant i \leqslant n$. Furthermore, we have $\mathbb{E}[X_i^2] = \mathrm{Var}(X_i) + \mu^2 = \sigma^2 + \mu^2$. Therefore:

$$\mathbb{E}[S_n^4] = nK + 3n(n-1)(\sigma^2 + \mu^2) < nK + 3n^2(\sigma^2 + \mu^2).$$

Applying Markov's Inequality with the fourth moment, we have:

$$\mathbb{P}(|S_n/n| \geqslant \epsilon) = \mathbb{P}(|S_n| \geqslant n\epsilon)$$

$$\leqslant \frac{\mathbb{E}[S_n^4]}{n^4 \epsilon^4}$$

$$< \frac{K}{n^3 \epsilon^4} + \frac{3(\sigma^2 + \mu^2)}{n^2}.$$

6

Therefore, we have:

$$\sum_{n=1}^{\infty} \mathbb{P}(|S_n/n| \geqslant \epsilon) < \frac{K}{\epsilon^4} \sum_{n=1}^{\infty} n^{-3} + 3(\sigma^2 + \mu^2) \sum_{n=1}^{\infty} n^{-2} < \infty \tag{22}$$

*Finally, by the Borel-Cantelli Lemma (**BCL**), we have:*

$$\mathbb{P}\left(\limsup_{n \to \infty} |S_n/n| \geqslant \epsilon \right) = 0, \quad \forall \epsilon > 0.$$

□.

Theorem 1.4: Borel-Cantelli Lemma (BCL)

1. First Borel-Cantelli Lemma: Given a probability space $(X, \mathcal{S}, \mathbb{P})$ and a sequence $\{A_n\}_{n=1}^{\infty} \subset \mathcal{S}$. If $\sum_{n=1}^{\infty} \mathbb{P}(A_n) < \infty$, we have:

$$\mathbb{P}\left(\limsup_{n \to \infty} A_n \right) = 0. \tag{23}$$

2. Second Borel-Cantelli Lemma: On the other hand, if $\sum_{n=1}^{\infty} \mathbb{P}(E_n) = \infty$, we have:

$$\mathbb{P}\left(\limsup_{n \to \infty} A_n \right) = 1. \tag{24}$$

Proof (Borel-Cantelli Lemma (**BCL**)). ─────────────────────
We focus on proving the first Borel-Cantelli lemma. We define another sequence of $\mathcal{S}$-measurable sets $\{B_n\}_{n=1}^{\infty}$ such that:

$$B_n = \bigcup_{k=n}^{\infty} A_n.$$

Hence, we have $B_{\ell+1} \subset B_\ell$ for every $\ell \geqslant 1$. In other words, B_n is a decreasing sequence of $\mathcal{S}$-measurable sets. By continuity of measure, we have:

$$\mathbb{P}\left(\lim_{n \to \infty} B_n \right) = \lim_{n \to \infty} \mathbb{P}(B_n)$$

$$= \lim_{n \to \infty} \sum_{k=n}^{\infty} \mathbb{P}(A_n) \quad \textit{(By additivity)}$$

$$= \sum_{i=1}^{\infty} \mathbb{P}(A_i) - \lim_{n \to \infty} \sum_{k=1}^{n} \mathbb{P}(A_n)$$

$$= 0.$$

Furthermore, we have:

$$\mathbb{P}\left(\lim_{n \to \infty} B_n \right) = \mathbb{P}\left(\lim_{n \to \infty} \bigcup_{k=n}^{\infty} \right) = \mathbb{P}\left(\bigcap_{n=1}^{\infty} \bigcup_{k=n}^{\infty} A_n \right) = \mathbb{P}\left(\limsup_{n \to \infty} A_n \right).$$

Hence proved the first Borel-Cantelli Lemma. To prove the second Borel-Cantelli Lemma, we prove the following:

$$1 - \mathbb{P}\left(\limsup_{n \to \infty} A_n \right) = \mathbb{P}\left(\left\{ \limsup_{n \to \infty} A_n \right\}^c \right)$$

$$= \mathbb{P}\left(\liminf_{n \to \infty} A_n^c \right) = 0.$$

□.

1.2.3 Central Limit Theorem

> **Theorem 1.5: Central Limit Theorem (CLT)**
>
> Let $X_1, \ldots X_n$ be a sequence of *i.i.d* random variables with expected value μ and finite variance σ^2. Then, we have:
>
> $$\frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} \xrightarrow{d} \mathcal{N}(0,1) \quad as \quad n \to \infty, \tag{25}$$
>
> where $\bar{X}_n = S_n/n$ and $\mathcal{N}(0,1)$ is the standard normal distribution.

Proof (Central Limit Theorem (**CLT**)). _____

We prove this via the Characteristic Function. Let $\bar{Z}_n = \frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}}$, notice that:

$$\bar{Z}_n = \frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{X_i - \mu}{\sigma},$$

Let $Z_i = X_i - \mu$ for $1 \leqslant i \leqslant n$ and suppose $Z = Z_1 = \cdots = Z_n$, we have:

$$
\begin{aligned}
\varphi_{\bar{Z}_n}(t) = \varphi_{\sum_{i=1}^{n} Z_i}\left(\frac{t}{\sqrt{n}}\right) &= \left[\varphi_Z\left(\frac{t}{\sqrt{n}}\right)\right]^n \\
&= \left[1 + \frac{it\mathbb{E}[Z]}{\sqrt{n}} - \frac{t^2}{2n}\mathbb{E}[Z^2] + \mathcal{O}(1/n)\right]^n \quad \text{(Taylor's Expansion)} \\
&= \left[1 - \frac{t^2}{2n} + \mathcal{O}(1/n)\right]^n.
\end{aligned}
$$

The final equality comes from the fact that $\mathbb{E}[Z] = 0$ and $\mathbb{E}[Z^2] = \mathbb{E}[Z]^2 + \mathrm{Var}(Z) = 1$. Finally, we have:

$$\lim_{n \to \infty} \varphi_{\bar{Z}_n}(t) = \lim_{n \to \infty} \left[1 - \frac{t^2}{2n} + \mathcal{O}(1/n)\right]^n = e^{-t^2/2}.$$

*Since $e^{-t^2/2}$ is the Characteristic Function of the standard normal distribution, by (**LCT**), we have $\bar{Z}_n \xrightarrow{d} \mathcal{N}(0,1)$.* □.

1.3 Convergence of Random Variables

B List of Definitions

B   Important Theorems

C   Important Corollaries

D   Important Propositions

↩ *Return to contents*

E   References

References

Durrett, Rick (2010). *Probability: Theory and Examples*. 4th. USA: Cambridge University Press. ISBN: 0521765390.

Gut, Allan (2004). *A Graduate Course in Probability*. Graduate Text in Mathematics.

undefinedinlar, Erhan (2011). *Probability and Stochastics*. Springer New York. ISBN: 9780387878591. DOI: 10.1007/978-0-387-87859-1. URL: http://dx.doi.org/10.1007/978-0-387-87859-1.

Wikipedia (2023). *Vitali set — Wikipedia, The Free Encyclopedia*. http://en.wikipedia.org/w/index.php?title=Vitali%20set&oldid=1187241923. [Online; accessed 24-December-2023].

↩ *Return to contents*