to$\xrightarrow{p}$

MATH5412 - Advanced Probability Theory II

Aaron Wang
aswang@connect.ust.hk

Spring 2022

Abstract

These are notes for MATH5412 at HKUST, the second course in a two part graduate-level course taught by Bao Zhigang in Spring 2022. The main focus is as a continuation of MATH5411.

Contents

1 February 8th, 2022

This is the first lecture of this course. We will discuss a bit about the logistics of the class and an overview of the content.

1.1 Overview of the Course

Although the previous course, MATH5411, covered the first half of Durrett [Dur19], this course will not be following the second half, which is largely about stochastic processes and Brownian motion, as there is another course MATH5450, Stochastic Processes, which will cover this exactly. Instead, this course will mostly look at limiting theorems, similar to the last part of MATH5411, but relaxing the i.i.d. constraints.

In MATH5411, we considered $S_n = \sum_{i=1}^{n} X_i$, but we had three assumptions:

1. The second moment $\mathbf{E}[X^2]$ exists.
2. X_i is sequence of random variables in $\mathbb{R}$.
3. X_i are independent.

Now, in this course, we will attempt to relax these assumptions, and these three extensions are along completely different directions. Here is a brief overview of these three extensions:

1.1.1 Stable Law

From the central limit theorem, we know that if the second moment exists, S_n goes to a Gaussian distribution under an appropriate normalization. If the second moment does not exist, we have the **stable law**.

The stable law is not like Gaussian, which is universal in a sense, since as long as the second moment exists, S_n goes to a Gaussian distribution under a normalization. Once you don't have a second moment, the limiting distribution depends on the tail behavior, with different tail behaviors resulting in different limiting distributions. As such, we have a *class of distributions* when we don't have the second moment.

In the last course, besides considering S_n, we also discussed the sum of triangular array for a given sequence of random variable. In this case, the limiting distribution might not be Gaussian, with a typical example being Poisson convergence.

Example 1.1 (Poisson Convergence for Rare Events)
Let $Y_{n,m} \sim \text{Be}(p_n)$ where $p = p_n = \frac{c}{n}$ with $Y_{n,m}$ i.i.d. $1 \leq m \leq n$. Then the limiting distribution of $S_n = \sum_{m=1}^{n} Y_{n,m}$ approaches Poisson(c).

If we have a sum of triangular arrays where we the second moment does not exist, the possible limiting distribution is called the **infinitely divisible distribution**. It will contain the stable law as a special case.

To reiterate, in the case of triangular array, if we have the second moment, we'd get either a Gaussian or Poisson limiting distribution. For the case where we don't have the second moment, we would get a class of distribution called the infinitely divisible distribution.

> **Remark 1.2 —** This part will take 3-4 lectures. References for this section can be found in Chapter 3 of Durrett [Dur19].

1.1.2 Functional Limiting Theorem

In the previous course, we were only concerned about the weak convergence of random variables in $\mathbb{R}$. What if we want to do the same for *random vectors* in $\mathbb{R}^k$? Thinking even more broadly, we want to consider the weak convergence of *random functions*, or **random processes**. This leads to the second extension which is the **functional limiting theorem**.

One typical example where the functional limiting theorem is used is when considering *empirical process*.

Example 1.3 (Example of Needing the Weak Convergence of a Random Function)

Say we have $X \sim F$, with F unknown and we want to perform statistical inference, with a sample $X_1, \ldots, X_n \sim F$ i.i.d. We can construct the **empirical distribution** $F_n(t) = \frac{1}{n}\sum\limits_{i=1}^{n} \mathbb{1}(X_i \leq t)$ to approximate F.

We then want to figure out how well this approximation is by taking it's difference $F_n(t) - F(t)$. By the law of large number, we know for any fixed t, the difference $F_n(t) - F(t)$ goes to 0, since $\mathbf{E}[F_n(t) - F(t)] = 0$. We also know that the fluctuation is given by CLT if we multiply by $\sqrt{n}$, simply from the CLT for i.i.d. random variables.

However, we don't only want to consider this closeness for a fixed t, we want to measure closeness as a *whole function*. As such, we might introduce a distance between two functions, say the **Kolmogorov–Smirnov Statistics** $:= \sup_t |F_n(t) - F(t)|$. We know that this goes to zero by the *Glivenko–Cantelli theorem*[a], which was introduced in the previous course. The problem is if we want to use this statistic for hypothesis testing, then we need to know the precise distribution of this statistic under suitable normalization. It turns out the suitable normalization is $\sqrt{n}$. If we consider $X(t) = \sqrt{n}(F_n(t) - F(t))$, which is a random function, the statistic becomes $\sup_t |X(t)|$, which is still a random variable. However, to do this, we need to find the weak limit of the whole stochastic process. Eventially, $X(t)$ will go to the *Brownian bridge*[b].

[a] https://en.wikipedia.org/wiki/Glivenko%E2%80%93Cantelli_theorem
[b] https://en.wikipedia.org/wiki/Brownian_bridge

> **Remark 1.4 —** This part will also be quite short. References for this section can be found in Chapter 2 of Billingsley's *Convergence of Probability Measure* [Bil86].

1.1.3 Martingale and it's Limiting Theorem

Roughly speaking, a martingale can be thought of the sum of a random variable. This random variable, in martingale theory, are called the **martingale differences**, which are not necessarily independent. These martingale differences lie somewhere between uncorrelated and independent random variables, having more structure than uncorrelated variables, but are not as good as independent variables. As such, although they are not necessarily independent, they share many common features with independent random variables.

Remark 1.5 — This part will be a major part of this course. References can be found in Chapter 5 of [Dur19] and Hall and Heyde's *Martingale Limit Theory and its Application* [HH80].

1.1.4 Concentration (if time permits)

If time permits, we will also cover something called **martingale concentration**. Very roughly speaking, concentration can be thought as an analog to the law of large numbers. Recall for WLLN, we briefly described geometric concentration. The systematic discussion of concentration will mainly focus on the non-asymptotic part, but we will still be considering a function of a large number of random variables. These random variables may be independent or not, or even martingale differences. This section is not necessarily about the limiting part of probability theory, as it focuses on the non-asymptotic behavior.

Remark 1.6 — References for part will be taken from Vershynin's *High-Dimensional Probability* [Ver19].

1.2 Heavy Tail Limiting (Poisson) Convergence

Before introducing the stable law, we will quickly review the heavy tail limiting convergence from the last part of MATH5411. Heuristically, the stable law and the heavy tail convergence are very related.

As with Example 1.1, we consider a triangular array, $Y_{n,1}, \ldots, Y_{n,n} \sim \mathrm{Be}(p)$ i.i.d. with $p = p_n = \frac{\lambda}{n}$. We have

$$\sum_{m=1}^{n} Y_{n,m} \implies \mathrm{Poisson}(\lambda).$$

After that, we did a generalization to not require the elements in the triangular array to be i.i.d.

Theorem 1.7 (Poisson Convergence for non i.i.d. Bernoulli Random Variables)
For each n, let $X_{n,m}$, $1 \le m \le n$ be independent r.v. with $\Pr(X_{n,m} = 1) = 1 - \Pr(X_{n,m} = 0) = \beta_{n,m}$. If $\sum\limits_{m=1}^{n} \beta_{n,m} \to \lambda$ and $\max\limits_{m} \beta_{n,m} \to 0$, then, $S_n = \sum\limits_{m=1}^{n} X_{n,m} \implies \mathrm{Poisson}(\lambda)$.

Remark 1.8 — This is similar to Lindeberg's condition for CLT.

After this, we can extend to non-Bernoulli random variables, being able to take any non-negative integer value, as long as it is "almost" Bernoulli.

Theorem 1.9 (Poisson Convergence for non-Bernoulli Random Variables)
For each n, let $X_{n,m}$, $1 \leq m \leq n$ be independent r.v. with $\Pr(X_{n,m} = 1) = \beta_{n,m}$ and $\Pr(X_{n,m} \geq 2) = \epsilon_{n,m}$. If $\sum_{m=1}^{n} \beta_{n,m} \to \lambda$, $\max_m \beta_{n,m} \to 0$ and $\sum_{m=1}^{n} \epsilon_{n,m} \to 0$, then, $S_n = \sum_{m=1}^{n} X_{n,m} \implies \text{Poisson}(\lambda)$.

Now with this general result, we are able to solve a mathematical modelling problem.

Example 1.10 (Modelling Customer Arrival)
Suppose we open a bank and we want to know the number of arrivals $N([s,t])$ during a time duration $[s,t]$. To model, this we make the following assumptions:

(i) The number in disjoint intervals are independent

(ii) The distribution of $N(s,t)$ only depends on $t-s$, i.e. it is **time homogeneous**

(iii) $\Pr(N([0,h]) = 1) = \lambda h + o(h)$, and

(iv) $\Pr(N([0,h]) \geq 2) = o(h)$

Theorem 1.11
If (i) - (iv) in Example **??** hold, then $N([0,t])$ has an exact Poisson distribution with mean λt.

For this example, what we really care is not the Poisson convergence, rather the consequence of this mathematical modelling problem. For this example, we not only get the distribution for a fixed t, we get a stochastic process. If we let t run from 0 to infinity, we get what is called a **Poisson point process**.

Definition 1.12 (Poisson point process with rate λ)**.** A family of random variables $N_t = N([0,t])$, $t \geq 0$, satisfying:

1. If $0 = t_0 < t_1 < \ldots < t_n$ then $N_{t_k} - N_{t_{k-1}} = N([t_{k-1}, t_k])$ are all independent.
2. $N_t - N_s \sim \text{Poisson}(\lambda(t-s))$.

There are also a few other ways to characterize a Poisson point process, such as by the time of arrival. Thus, this process can be characterized by these points if it's counting function satisfy the properties in Definition **??**. We can also regard a Poisson point process as a random measure, leading to us being able to generalize a Poisson point process on a measure space.

Definition 1.13 (Poisson point process on a measurable space $(S, \mathscr{S}, \mu)$)**.** A random map $m : \mathscr{S} \to \{0, 1, \dots\}$ that for each ω is a measure on $\mathscr{S}$, and has the following property:
If $A_1, A_2, \dots, A_n$ are disjoint with $\mu(A_i) < \infty$ then:

1. $m(A_1), \dots, m(A_n)$ are independent.
2. $m(A_i) =^D \text{Poisson}(\mu(A_i))$.

1.3 Stable Law

We have:

$$X_1, X_2, \dots X_n \text{ i.i.d.} \quad S_n = \sum_{i=1}^{n} X_i$$

If $\mathbf{E}X_i = \mu$ and $\mathbf{Var}X_i = \sigma^2$, we have:

$$\frac{S_n - n\mu}{\sqrt{n}\sigma} \implies N(0,1)$$

Now, if $\mathbf{E}X_i^2 = \infty$, do we have a_n, b_n, Y s.t.:

$$\frac{S_n - b_n}{a_n} \implies Y \quad (Y \text{ nondegenerate})$$

Let us start with a simple case where everything about X_i is known.

Example 1.14

Consider $X_1, X_2, \dots$ i.i.d.

$$\Pr(X_1 > x) = \Pr(X_1 < -x) = \frac{x^{-\alpha}}{2}, \text{ for } x \geq_1, 0 < \alpha < 2$$

Density $f(x) = \alpha \frac{|x|^{-\alpha-1}}{2}$, $|x| > 1$ Note that this is:

- symmetric (indicates $b_n = 0$)
- $\mathbf{E}X_1^2 = 2 \int_1^\infty x \Pr(|x_1| > x) dx = \int_1^\infty x^{-\alpha+1} dx = \infty$

The solution is:

$$\mathbf{E}[e^{isS_n}] = \left[\underbrace{\mathbf{E}e^{isX_1}}_{\phi(s)} \right]^n = [1 - (1 - \phi(s))]^n$$

$$1 - \phi(s) = \int_1^\infty (1 - e^{ist}) \frac{\alpha}{2|x|^{\alpha+1}} dx + \int_{-\infty}^{-1} (1 - e^{isx}) \frac{\alpha}{2|x|^{\alpha+1}} dx$$

2 February 10th, 2022

2.1 Cont.

Recall from the last time, we found that most of the contribution is from the large points of scale $O(n^{1/\alpha})$.

Let us define an index set of large points:

$$I_n(\epsilon) = \{m \leq n : |X_m| > \epsilon n^{1/\alpha}\}$$

and define the sums:

$$\hat{S}_n(\epsilon) = \sum_{m \in I_n(\epsilon)} X_m = \sum_{m=1}^{n} X_m \mathbb{1}(|x_m| > \epsilon n^{1/\alpha})$$

$$\overline{S}_n(\epsilon) = S_n - \hat{S}_n(\epsilon) = \sum_{m=1}^{n} X_m \mathbb{1}(|X_m| \leq \epsilon \leq \epsilon^{1/\alpha})$$

Now we have two task:

- Show $\frac{\overline{S}_n(\epsilon)}{n^{1/\alpha}}$ is small if ϵ is small
- Find the limit of $\frac{\hat{S}_n(\epsilon)}{n^{1/\alpha}}$

Proof.

$$\mathbf{E}\left[\frac{\overline{S}_n(\epsilon)}{n^{1/\alpha}}\right]^2 = n^{-\frac{2}{\alpha}} \cdot n \cdot \mathbf{E}\left[\overline{X_1}(\epsilon)\right]^2, \quad \overline{X_i}(\epsilon) = X_i \mathbb{1}(|X_i| \leq \epsilon n^{1/\alpha})$$

$$\mathbf{E}[\overline{X_1}(\epsilon)]^2 = \int_0^\infty 2y \Pr(|\overline{X}_1(\epsilon)| \geq y\ dy) \leq \int_0^{\epsilon n^{1/\alpha}} 2y$$

Later we choose $\epsilon = \epsilon \to 0$ as $n \to \infty$. $\square$

Proof. Proof of (2).

$$\mathbf{E}\exp\left(it\frac{\hat{S}_n(\epsilon)}{n^{1/\alpha}}\right) = \sum_{m=0}^{n} \mathbf{E}\left[\exp\left(it\frac{\hat{S}_n(\epsilon)}{n^{1/\alpha}}\right)\Big||I_n(\epsilon)| = m\right]\Pr(|I_n(\epsilon)| = m)$$

We will use two facts:

1. $|I_n(\epsilon)|$ is Bin $\left(n, \frac{\epsilon^{-\alpha}}{n}\right) \sim$ Poisson$(\epsilon^{-\alpha})$. $\Pr(|X_n| > \epsilon n^{\frac{1}{\alpha}}) = \epsilon^{-\alpha}\frac{1}{n}$.

$\square$

References

[HH80] P. Hall and C.C. Heyde. *Martingale Limit Theory and Its Application*. Probability and mathematical statistics. Academic Press, 1980. ISBN: 9781483240244. URL: http://www.stat.yale.edu/~mjk56/MartingaleLimitTheoryAndItsApplication.pdf.

[Bil86] Patrick Billingsley. *Probability and Measure*. Second. John Wiley and Sons, 1986. URL: https://www.colorado.edu/amath/sites/default/files/attached-files/billingsley.pdf.

[Dur19] Rick Durrett. *Probability: Theory and Examples*. 2019. URL: https://services.math.duke.edu/~rtd/PTE/PTE5_011119.pdf.

[Ver19] Roman Vershynin. *High-Dimensional Probability*. 2019. URL: https://www.math.uci.edu/~rvershyn/papers/HDP-book/HDP-book.pdf.

Index