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 **2**
- 1.1 Overview of the Course . . . 2
- 1.2 Heavy Tail Limiting (Poisson) Convergence . . . 4
- 1.3 Stable Law . . . 6

2 February 10th, 2022 **10**
- 2.1 Stable Law Continued . . . 10
- 2.2 Proof of Stable Law . . . 13

3 February 17th, 2022 **15**
- 3.1 Proof of Stable Law Continued . . . 15
- 3.2 Infinitely Divisible Distribution . . . 18
- 3.3 Functional Limit Theorems . . . 19

4 February 22nd, 2022 **20**
- 4.1 Functional Limit Theorems Continued . . . 20
- 4.2 Finite-Dimensional Distribution . . . 22
- 4.3 Weak Convergence in $(C, \mathscr{C})$ . . . 24

5 February 24th, 2022 **25**
- 5.1 Relative Compactness of $\{\mu_n\}$ . . . 25

Index **26**

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, $Y_{n,1}, \dots, Y_{n,n} \sim \mathrm{Be}(p)$ 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 $\mathbb{P}(X_{n,m} = 1) = 1 - \mathbb{P}(X_{n,m} = 0) = \beta_{n,m}$. If $\sum_{m=1}^{n} \beta_{n,m} \to \lambda$ and $\max_m \beta_{n,m} \to 0$, then, $S_n = \sum_{m=1}^{n} X_{n,m} \implies \mathrm{Poisson}(\lambda)$.

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 $\mathbb{P}(X_{n,m} = 1) = \beta_{n,m}$ and $\mathbb{P}(X_{n,m} \geq 2) = \epsilon_{n,m}$. 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) $\mathbb{P}(N([0,h]) = 1) = \lambda h + o(h)$, and

(iv) $\mathbb{P}(N([0,h]) \geq 2) = o(h)$

Theorem 1.11
If (i) - (iv) in Example 1.10 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 1.12. 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 $m : \mathscr{S} \to \{0, 1, \ldots\}$ that for each ω is a measure on $\mathscr{S}$, and has the following property:
If $A_1, A_2, \ldots, A_n$ are disjoint with $\mu(A_i) < \infty$ then:

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

where $\mu(A) := \mathbf{E}[m(A)]$ is the mean measure of m.

1.3 Stable Law

Now that we have review Poisson point processes, let us move onto stable law. Consider:

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

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 $\mathbf{E}X_i^2 = \infty$, we want to ask if we have a_n, b_n, Y such that:

$$\frac{S_n - b_n}{a_n} \implies Y \tag{1}$$

Where Y is nondegenerate (if it is, then it would be trivial). a_n is basically the typical size of the fluctuation of S_n. In the case where the second moment exist, we know that this is of order $\sqrt{n}$. If we don't have the second moment, which are so called **heavy tailed random variables**, then these variables are more likely to take on large values, meaning that a_n should intuitively be larger than $\sqrt{n}$. How much larger depends on the explicit tail behavior of X_i. In a very special case, will be Gaussian, but in most cases it will not.

Similar to with the CLT, we eventually want to remove the assumptions about the distribution. However, let us first start with a specific special case where we know the explicit distribution of X_i. For this, we will present two solutions, the first does not have anything to do with Poisson point process, the but second will relate it to this heuristic.

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

$$\mathbb{P}(X_1 > x) = \mathbb{P}(X_1 < -x) = \frac{x^{-\alpha}}{2}, \text{ for } x \geq 1,\ 0 < \alpha < 2.$$

The density function is thus given by:

$$f(x) = \alpha \frac{|x|^{-\alpha-1}}{2}, \quad |x| > 1$$

Note that this density function is symmetric (indicating $b_n = 0$). In addition, computing the second moment using the tail sum formula, we have:

$$\mathbf{E}X_1^2 = 2\int_1^\infty x\mathbb{P}(|X_1| > x)dx = \int_1^\infty x^{-\alpha+1}dx = \infty$$

since $\alpha < 2$.

Remark 1.14 — Note that if $\alpha = 2$, then $\mathbf{E}X_1^2 = 0$. However, this is a very different case, and as such we do not consider it. In this case, the CLT holds with normalization $\sqrt{n \log n}$. See Theorem 1.12.3 from [UZ11].

Remark 1.15 — As mentioned above, since this is symmetric, we can have $b_n = 0$. This suggests that we might have non-zero b_n for non-symmetric cases.

Solution 1

We will try to compute limiting distribution using the Levy's continuity theorem by finding the limit of the characteristic function.

Theorem 1.16 (Levy's Continuity Theorem)
Suppose we have:

- a sequence of random variables $\{X_n\}_{n=1}^{\infty}$, not necessarily sharing a common probability space,
- the sequence of corresponding characteristic functions $\{\varphi_n\}_{n=1}^{\infty}$, where $\varphi_n(t) = \mathbf{E}\left[e^{itX_n}\right], \ \forall t \in \mathbb{R}, \ \forall n \in \mathbb{N}$,

If the sequence of characteristic functions converges pointwise to some function $\varphi_n(t) \to \varphi(t)\ \forall t \in \mathbb{R}$, then the following statements are equivalent:

- $X_n \xrightarrow{d} X$ for some random variable X.
- $\{X_n\}_{n=1}^{\infty}$ is tight: $\lim_{x\to\infty}\left(\sup_n \mathrm{P}\left[\,|X_n| > x\,\right]\right) = 0$;
- $\varphi(t)$ is the characteristic function of some random variable X;
- $\varphi(t)$ is a continuous function of t;
- $\varphi(t)$ is continuous at $t = 0$.

We have:

$$\mathbf{E}[e^{isS_n}] = \mathbf{E}[e^{is\sum_{i=1}^n X_i}] = \left[\mathbf{E}[e^{isX_1}]\right]^n$$

Now we need to choose a normalization such that this does boil down to a characteristic function of a single point mass. In other words, we want this to be of the form $(1 + O(\frac{1}{n}))^n$. In this case, we choose $\varphi(s) = \mathbf{E}[e^{isX_1}$, such that:

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

such that $1 - \varphi(s) \sim O(\frac{1}{n})$. We have:

$$\begin{aligned}
1 - \varphi(s) &= \mathbf{E}[e^{isX_1}] \\
&= \int_1^{\infty} (1 - e^{isx}) \frac{\alpha}{2|x|^{\alpha+1}} dx + \int_{-\infty}^{-1} (1 - e^{isx}) \frac{\alpha}{2|x|^{\alpha+1}} dx \\
&= \alpha \int_1^{\infty} \frac{1 - \cos(sx)}{x^{\alpha+1}} dx \qquad (2)
\end{aligned}$$

For the case where $s \geq 0$, with a change of variables, we have:

$$1 - \varphi(s) = \alpha \int_1^\infty \frac{1 - \cos(sx)}{x^{\alpha+1}} dx = s^\alpha \alpha \int_s^\infty \frac{1 - \cos(u)}{u^{\alpha+1}} du \tag{3}$$

Going back to $\frac{S_n}{a_n} = Y$, with a_n roughly greater than $\sqrt{n}$. Now we absolve a_n into s, meaning that s should be really small, eventually going to zero. In the integral on the RHS of Equation 3, the singularity at infinity can be ignored, due to the $u^{\alpha+1}$ in the denominator. For the singularity at zero, note that $1 - \cos(u) \sim u^2$ when u close to zero, meaning the integrand is almost $u^{1-\alpha}$. Since $\alpha < 2$, this is also integrable. Thus, as n go to infinity, the integral goes to a constant, giving us $1 - \varphi(s) = s^\alpha C_\alpha$. Choosing $s = \frac{t}{n^{1/\alpha}}$, with fixed $t \in \mathbb{R}$. This gives us:

$$\mathbf{E}\left[\exp\left\{it\frac{S_n}{n^{1/\alpha}}\right\}\right] = \left[1 - \frac{1}{n}|t|^\alpha C_\alpha\right]^n \to e^{-C_\alpha |t|^\alpha}.$$

This means that $\frac{S_n}{n^{1/\alpha}} \to Y$ with characteristic function $e^{-C_\alpha |t|^\alpha}$. This is one case of a stable law, with a specific example.

Note that in this case, Gaussian is also a special case of the stable law if we choose $\alpha = 2$. If we choose $\alpha = 1$, we would get Cauchy. In the general case, we do not have a explicit formula for the distribution or density function, so we often just express in terms of the characteristic function. For some asymptotic analysis of the density function, refer to [Nol20].

As expected, this scaling is larger than $\sqrt{n}$ since $\alpha < 2$. This solution is simple because we have the explicit distribution.

Solution 2

Before presenting the solution, we will do some preliminary analysis on the solution found above.

The reason why we expect the normalization to be larger than $\sqrt{n}$ is because of the large tails. This motivates us to look more into these tail behavior. We believe that there is major contribution to S_n from the larger parts random variables. Investigating this way will allow us to remove any explicit assumptions of the distribution besides the tail.

Starting from Equation 2, plugging in the value of s, we want to find a b such that the contribution from the small parts:

$$\int_1^{n^b} \frac{1 - \cos(\frac{t}{n^{1/\alpha}})}{x^{\alpha+1}} dx \ll O\left(\frac{1}{n}\right).$$

We can see that this occurs if $b < 1/\alpha$, using by using Taylor expansion. This rough analysis tells us that the contribution from parts of the distribution before $n^{1/\alpha}$ are not important for the limiting distribution. In other words, only the parts greater than $n^{1/\alpha}$ are relevant to our analysis. Now let us look at the behaviour when it is on the order of $n^{1/\alpha}$.

By definition, for any $b > a > 0$, if $a^{1/\alpha} > 1$, then we want to consider the scale:

$$\begin{aligned}\mathbb{P}(an^{1/\alpha} < X_1 < bn^{1/\alpha}) &= \mathbb{P}(X_1 > an^{1/\alpha}) - \mathbb{P}(X_1 > bn^{1/\alpha}),\\ &= \frac{1}{2}(a^{-\alpha} - b^{-\alpha}) \cdot \frac{1}{n}\end{aligned}$$

since this is the scale where it starts to contribute to the limiting distribution. Let us study the indicator function:

$$\mathbb{1}\left(\frac{X_1}{n^{1/\alpha}} \in (a,b)\right) \sim \text{Be}\left(\frac{1}{2}(a^{-\alpha} - b^{-\alpha}) \cdot \frac{1}{n}\right).$$

This parallels the Poisson convergence. If we denote the counting measure of this indicator function, we have:

$$N_n((a,b)) = \sum_{i=1}^{n} \mathbb{1}\left(\frac{X_1}{n^{1/\alpha}} \in (a,b)\right) \implies N(a,b) = \text{Poisson}\left(\frac{1}{2}(a^{-\alpha} - b^{-\alpha})\right).$$

Note that we are counting the number of X_i that are of the order of $n^{1/\alpha}$. Note that a Poisson r.v. is of order 1, meaning that number of X_i of order $n^{1/\alpha}$ is of constant order. This tells us that the major contribution of S_n in the heavy tail case comes from a constant number of points.

> **Remark 1.17** — This is in contrast to the CLT, in which if you decompose it into small and large parts, both have significant contributions.

More generally, we can define this constant measure as:

$$N_n(A) \quad \forall A \subset \mathbb{R} \setminus (-\delta, \delta), \quad \delta n^{1/\alpha} > 1$$

then:

$$\mathbb{P}\left(\frac{X_1}{n^{1/\alpha}} \in A\right) = \int_A \frac{\alpha}{2|x|^{\alpha+1}} dx \cdot \frac{1}{n}$$

If we think of $N_n(A)$ as a random measure, we have:

$$N_n(a) \implies N(A) \sim \text{Poisson}\left(\int_A \frac{\alpha}{2|x|^{\alpha+1}} dx\right) = \text{Poisson}(\mu(A))$$

meaning that N is a Poisson point process on $(\mathbb{R} \setminus (-\delta,\delta), \mathscr{B}(\mathbb{R} \setminus (-\delta,\delta)), \mu(\cdot))$.

This means that S_n will converge to the sum of points in this Poisson point process. Thus, the stable law is just the distribution of points in the Poisson point process.

> **Remark 1.18** — This Poisson point process is no longer homogeneous, since the measure is not Lebegue measurable. Most of the points in S_n will go to zero, besides the finite heavy tail ones.

2 February 10th, 2022

2.1 Stable Law Continued

Recall from the last time, we concluded that most of the contribution for S_n is from the large points of scale $O(n^{1/\alpha})$ and that this is of constant order. Let us define an index set of large points:

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

and define the sums:

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

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

Intuitively speaking $\hat{S}_n(\epsilon)$ represents the sum of large points and $\overline{S}_n(\epsilon)$ represents the sum of small points.

Remark 2.1 — Later on, ϵ will be chosen to be as small as possible. Later we will it to go to zero along with n, e.g. $1/\log n$, since we might exclude relevant points. For now we will consider it fixed.

Now we have two task, to show

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

Proof. of 1.

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

$$\begin{aligned}
\mathbf{E}\left[\left(\overline{X_1}(\epsilon)\right)^2\right] &= \int_0^\infty 2y\mathbb{P}(|\overline{X}_1(\epsilon)| > y)\, dy \\
&\le \int_0^{\epsilon n^{1/\alpha}} 2y\mathbb{P}(|X_1| > y) dy \\
&= \int_0^1 2y\, dy + \int_1^{\epsilon n^{1/\alpha}} \cdot 2y y^{-\alpha} dy \le \frac{2\epsilon^{2-\alpha}}{2-\alpha} n^{\frac{2}{\alpha}-1}
\end{aligned}$$

This gives us:

$$\mathbf{E}\left[\left(\frac{\overline{S}_n(\epsilon)}{n^{1/\alpha}}\right)^2\right] \le \frac{2\epsilon^{2-\alpha}}{2-\alpha}, \quad 0 < \alpha < 2$$

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

Proof. of 2.

Note that $\hat{S}_n(\epsilon)$ is a sum of a random number of r.v. We will find the characteristic function using the total law of expectation:

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

Now, we need to find these two terms. We will start with finding $\mathbb{P}(|I_n(\epsilon)| = m)$. We will use two facts:

1. $|I_n(\epsilon)| = \sum_{m=1}^{n}\mathbb{1}(|X_m| > \epsilon n^{1/\alpha})$ is $\text{Bin}\left(n, \frac{\epsilon^{-\alpha}}{n}\right) \sim \text{Poisson}(\epsilon^{-\alpha})$, giving us $\mathbb{P}(|X_n| > \epsilon n^{\frac{1}{\alpha}}) = \epsilon^{-\alpha}\frac{1}{n}$.
2. The conditional distribution of $\hat{S}_n(\epsilon)\Big||I_n(\epsilon)| = m$ equals the distribution of the sum of m i.i.d. r.v. with c.d.f. F_ϵ defined as:
$$1 - F_\epsilon(x) = \mathbb{P}\left(\frac{X_1}{n^{1/\alpha}} > x\Big|\frac{|X_1|}{n^{1/\alpha}} > \epsilon\right).$$
i.e. F_ϵ is the conditional distribution of $\frac{X_1}{n^{1/\alpha}}$ given $\frac{|X_1|}{n^{1/\alpha}} > \epsilon$.

Proof.

$$\begin{aligned}\mathbb{P}(\hat{S_n}(\epsilon) \in B\big||I_n(\epsilon)| = m) &= \frac{\mathbb{P}(\hat{S_n}(\epsilon) \in B, |I_n(\epsilon)| = m)}{\mathbb{P}(|I_n(\epsilon)| = m)} \\ &= \frac{\binom{n}{m}\mathbb{P}\left(\sum_{i=1}^{m} X_i \in B, |X_1| > \epsilon n^{1/\alpha}, \ldots, |X_m| > \epsilon n^{1/\alpha}\right)}{\binom{n}{m}\mathbb{P}\left(|X_1| > \epsilon n^{1/\alpha}, \ldots, |X_m| > \epsilon n^{1/\alpha}\right)}\end{aligned}$$

□

For our distribution, we have:

$$1 - F_\epsilon(x) = \frac{x^{-\alpha}}{2\epsilon^{-\alpha}}, \quad x \geq \epsilon.$$

i.e. F_ϵ is the c.d.f. of ϵX_1, meaning that the characteristic function of F_ϵ is $\varphi(\epsilon t)$. Consequently:

$$\begin{aligned}\mathbf{E}\left[\exp\left\{it\hat{S}_n(n^{1/\alpha})\right\}\right] &= \sum_{m=0}^{n}\binom{n}{m}\left(\frac{\epsilon^{-\alpha}}{n}\right)^m\left(1 - \frac{\epsilon^{-\alpha}}{n}\right)^{n-m}[\varphi(\epsilon t)]^m \\ &\to \sum_{m=0}^{\infty}\exp(-\epsilon^{-\alpha})\cdot(-\epsilon^{-\alpha})^m\frac{[\varphi(\epsilon t)]^m}{m!} = \exp\left\{-\epsilon^{-\alpha}(1 - \varphi(\epsilon t))\right\}\end{aligned}$$

using Poisson approximation for binomial and DCT.

Recall earlier that we have an approximation for $1 - \varphi(\epsilon t) = C_\alpha\epsilon^\alpha|t|^\alpha$ if $\epsilon \to 0$, giving us:

$$\mathbf{E}\left[\exp\left\{it\hat{S}_n(n^{1/\alpha})\right\}\right] = \exp(-C_\alpha|t|^\alpha),$$

which is the same as Solution 1. Note that we need to choose $\epsilon = \epsilon_n \downarrow 0$. For more details see Lemma 3.7.1 of [Dur19]. □

From this solution, we can see that only the tail part matters. Now, we will try to generalize this solution.

Definition 2.2 (slowly varying function). $L : \mathbb{R} \to \mathbb{R}$ is a slowly varying function if it satisfies:

$$\lim_{x\to+\infty} \frac{L(tx)}{L(x)} = 1,$$

for any fixed $t > 0$.

Example 2.3

$\log x$, $\log\log x$, $\log\sqrt{x}$ are slowly varying functions, but any power function x^t is not.

Theorem 2.4 (stable law)

Suppose $X_1, X_2 \ldots$ are i.i.d. with distribution satisfying:

(i) $\lim\limits_{x\to+\infty} \mathbb{P}(X_1 > x)/\mathbb{P}(|X_1| > x) = \theta \in [0,1]$ (tails may not be significant)

(ii) $\mathbb{P}(|X_1| > x) = x^{-\alpha}L(x)$, $\alpha < 2$, and L is slowly varying (general total tail)

Let $S_n = \sum\limits_{i=1}^{n} X_i$, $a_n = \inf\{x : \mathbb{P}(|X_1| > x) \le \frac{1}{n}\}$, $b_n = n\mathbf{E}\left[X_1 \mathbb{1}(|X_1| \le a_n)\right]$, then as $n \to \infty$:

$$\frac{S_n - b_n}{a_n} \implies Y,$$

for a non-degenerate r.v. Y.

Remark 2.5 — θ in Theorem 2.4 indicates the relative heaviness between the right and left tail. If θ close to 1, the right tail is dominant, if $\theta \approx \frac{1}{2}$ then both tails are roughly equal.

We want to choose a_n s.t. $\mathbb{P}\left(\frac{X_1}{a_n} \in (\alpha,\beta)\right) \sim \frac{1}{n}$ since $\frac{S_n}{a_n} = \sum\limits_{i=1}^{n} \frac{X_1}{a_n}$, and we want the number of large points to be a constant order random variable. A natural choice is $\mathbb{P}(|X_1| \ge a_n) \sim \frac{1}{n}$, which gives us the quantile of $\frac{1}{n}$, i.e. $a_n = \inf\{x : \mathbb{P}(|X_1| > x) \le \frac{1}{n}\}$.

Remark 2.6 — We could have used ca_n for any constant c. In this case, we just choose $c = 1$.

For choosing b_n, we can choose $b_n = n\mathbf{E}\left[X_1 \mathbb{1}(|X_1| \le ca_n)\right]$ for any constant as well. This is because $b_n = n\mathbf{E}[\underbrace{X_1}_{a_n} \underbrace{\mathbb{1}(|X_1| \le ca_n)}_{\mathbb{P}(\cdot)\sim 1/n}] \sim a_n$, meaning that the limit would differ by a constant factor.

Remark 2.7 — The reason why we can truncate to of order a_n instead of something much larger say a_n^2 is because with high probability there are no such points.

2.2 Proof of Stable Law

Claim 2.8.

$$n\mathbb{P}(|X_1| > \alpha_n) \to 1, \quad n \to \infty$$

Proof. omitted. □

For the tail behavior, we get:

$$\begin{aligned}
&n\mathbb{P}(|X_1| > x\alpha_n) \to \theta x^{-\alpha}, \quad n \to \infty, x > 0 \\
\sim &n\mathbb{P}(|X_1| > \alpha_n) \cdot \theta = n(xa_n)^{\alpha} L(xa_n) \cdot \theta \\
\sim &n(xa_n)^{\alpha} L(a_n) \cdot \theta = nx^{-\alpha}\mathbb{P}(|X_1| > a_n) \cdot \theta \sim x^{-\alpha} \cdot \theta
\end{aligned}$$

meaning that a constant in front of a_n does not affect the convergence.

This also tells us that if we use compute the counting measure:

$$N_n((x, \infty)) = \sum_{m=1}^{n} \mathbb{1}\left(\frac{X_m}{a_n} > x\right) \implies \text{Poisson}(\theta x^{-\alpha}).$$

More generally $N_n(A)$ converges to a Poisson point process $N(A)$ with mean measure

$$\mathbf{E}N(A) = \mu(A) = \int_{A\cap(0,\infty)} \theta\alpha|x|^{-(\alpha+1)}dx + \int_{A\cap(-\infty,0)} (1-\theta)\alpha|x|^{-(\alpha+1)}dx.$$

Now we will decompose the points into large and small parts. Let us define index set:

$$I_n(\epsilon) = \{m \leq n : |X_m| > \epsilon a_n\}$$

and define the following:

$$\begin{aligned}
\hat{S}_n(\epsilon) &= \sum_{m\in I_n(\epsilon)} X_m \quad \text{(sum of large points)} \\
\overline{\mu}(\epsilon) &= \mathbf{E}[X_m \mathbb{1}(|X_n| \leq \epsilon a_n)] = \mathbf{E}\overline{X_m}(\epsilon) \\
\hat{\mu}(\epsilon) &= \mathbf{E}[X_m \mathbb{1}(\epsilon a_n < |X_n| \leq a_n)] \\
\overline{S}_n(\epsilon) &= (S_n - b_n) - (\hat{S}_n(\epsilon) - n\hat{\mu}(\epsilon)) \\
&= \sum_{m=1}^{n} \left(\overline{X_m}(\epsilon) - \overline{\mu}(\epsilon)\right) \text{ (centered sum of small points)}
\end{aligned}$$

Remark 2.9 — Unlike the special case, we need to subtract by b_n, since it is no long symmetric.

Remark 2.10 — Note from the definition of b_n, we truncate $\hat{\mu}(\epsilon)$ instead of going to infinity.

Now we have once again have two tasks:

1. Show $\frac{\overline{S}_n(\epsilon)}{a_n}$ is small if ϵ is small.
2. Find the limit of $\frac{\hat{S}_n(\epsilon)-n\hat{\mu}(\epsilon)}{a_n}$.

Proof. of 1.

$$\begin{aligned}\mathbf{E}\left[\left(\frac{\overline{S}_n(\epsilon)}{a_n}\right)^2\right] &\le n\mathbf{E}\left[\left(\frac{\overline{X_1}(\epsilon)}{a_n}\right)^2\right] \\ &\le \int_0^\epsilon 2y\mathbb{P}(|\overline{X}_1(\epsilon)| > ya_n)\, dy \\ &= \underbrace{n\mathbb{P}(|X_1| > a_n)}_{\to 1} \int_0^\epsilon 2y \underbrace{\frac{\mathbb{P}(|X_1| \ge ya_n)}{\mathbb{P}(|X_1| > a_n)}}_{y^{-\alpha}} dy \\ &\to \int_0^\epsilon 2yy^{-\alpha} dy = \frac{2}{2-\alpha}\epsilon^{2-\alpha} \to 0 \text{ if } \epsilon \to 0\end{aligned}$$

$\square$

Proof. of 2.
Let us first consider trying to compute the characteristic function of $\frac{S_n}{a_n}$, since we can add the constant part later. We have the following:

(i) $|I_n(\epsilon)| \to \text{Poisson}(\epsilon^{-\alpha})$

(ii) Given $|I_n(\epsilon)| = m$, $\frac{\hat{S_n}(\epsilon)}{a_n}$ has the same distribution as the sum of m i.i.d. r.v. with c.d.f. F_ϵ, which is again the conditional distribution of $\frac{X_1}{a_n}$ given $\frac{|X_1|}{a_n} \ge \epsilon$. This time, we need to distinguish the left and right tails:

$$\begin{aligned}1 - F_n^\epsilon(x) &= \mathbb{P}\left(\frac{X_1}{a_n} > x \middle| \frac{|X_1|}{a_n} > \epsilon\right) \to \theta\frac{x^{-\alpha}}{\epsilon^{-\alpha}} \\ F_n^\epsilon(-x) &= \mathbb{P}\left(\frac{X_1}{a_n} < -x \middle| \frac{|X_1|}{a_n} > \epsilon\right) \to (1-\theta)\frac{|x|^{-\alpha}}{\epsilon^{-\alpha}}\end{aligned}$$

Let $\Psi_n^\epsilon(t) \to \Psi^\epsilon(t)$ be the c.f. of F_n^ϵ:

$$\Psi_n^\epsilon(t) \to \Psi^\epsilon(t) = \int_\epsilon^\infty e^{itx}\theta\epsilon^\alpha \alpha x^{-(\alpha+1)}dx + \int_{-\infty}^{-\epsilon} e^{itx}(1-\theta)\epsilon^\alpha|x|^{-(\theta+1)}dx$$

Note that these tails only hold when $|x| > \epsilon$, as the density would be zero otherwise. This will be continued next lecture.

$\square$

3 February 17th, 2022

3.1 Proof of Stable Law Continued

Let $\Psi_n^\epsilon(t)$ be the c.f. of F_n^ϵ. We have:

$$\Psi_n^\epsilon(t) \to \Psi^\epsilon(t) = \int_\epsilon^\infty e^{itx}\theta\epsilon^\alpha x^{-(\alpha+1)}dx + \int_{-\infty}^{-\epsilon} e^{itx}(1-\theta)\epsilon^\alpha |x|^{-(\alpha+1)}dx$$

Now, we have:

$$\begin{aligned}
\mathbf{E}\exp\left(it\frac{\hat{S}_n(\epsilon)}{a_n}\right) &= \sum_m \mathbf{E}\exp\left(it\frac{\hat{S}_n(\epsilon)}{a_n}\middle| |I_n(\epsilon)| = m\right)\cdot \mathbb{P}\left(|I_n(\epsilon) = m\right) \\
&\sim \sum_{m=0}^\infty [\Psi^\epsilon(t)]^m \frac{(\epsilon^{-\alpha})^m e^{-\epsilon^{-\alpha}}}{m!} \\
&= \exp\left(-\epsilon^{-\alpha}(1-\Psi^\epsilon(t))\right) \\
&= \exp\left[\int_\epsilon^\infty (e^{itx}-1)\theta\alpha x^{-(\alpha+1)}dx + \int_{-\infty}^{-\epsilon}(e^{itx}-1)(1-\theta)\alpha|x|^{-(\alpha+1)}dx\right]
\end{aligned}$$

Remark 3.1 — The approximation should be justified by DCT, since we need to justify the convergence of the total sum.

Note that for the above case, ϵ is fixed. In the general case, we need to send $\epsilon \downarrow 0$. However, when $x \to 0$, $e^{itx}-1 \sim itx$, and $x \cdot x^{-(\alpha+1)} = x^{-\alpha}$ is not integrable around 0 if $\alpha \geq 1$.

Remark 3.2 — When $\theta \neq \frac{1}{2}$, this singularity appears, which does not happen when we consider the special case.

As such, we need to consider the centered sum $\exp\left(-it\frac{n\hat{\mu(\epsilon)}}{a_n}\right)$, with:

$$\hat{\mu(\epsilon)} = \mathbf{E}X_1\mathbb{1}(\epsilon a_n < |X_1| \leq a_n).$$

As seen previously, from the assumption of the tail behavior and slowly varying function, we have:

$$\begin{aligned}
&\mathbb{P}\left(x < \frac{X_1}{a_n} \le y\right) = \frac{1}{n}\theta(x^{-\alpha} - y^{-\alpha}) \\
\implies & n\hat{\mu(\epsilon)}a_n \to \int_{\epsilon}^{1} x\theta\alpha x^{-(\alpha+1)}dx + \int_{-1}^{-\epsilon} x(1-\theta)\alpha|x|^{-(\alpha+1)}dx \\
\implies & \mathbf{E}\exp\left(it\frac{\hat{S_n(\epsilon)} - n\hat{\mu}(\epsilon)}{a_n}\right) \to \exp\Bigg[\int_{1}^{\infty}(e^{itx}-1)\theta\alpha x^{-(\alpha+1)}dx \\
&+ \int_{\epsilon}^{1}(e^{-tx} - 1 - itx)\theta\alpha x^{-(\alpha+1)}dx \\
&+ \int_{-1}^{-\epsilon}(e^{itx} - 1 - itx)(1-\theta)\alpha|x|^{-(\alpha+1)}dx \\
&+ \int_{-\infty}^{-1}(e^{itx}-1)(1-\theta)\alpha|x|^{-(\alpha+1)}dx\Bigg]
\end{aligned}$$

which is integrable.

Simplifying, and sending $\epsilon \downarrow 0$, we get:

$$\exp\left[itc + \int_0^\infty (e^{itx} - 1 - \frac{itx}{1+x^2})\theta\alpha x^{-(\alpha+1)}dx + \int_{-\infty}^{0}(e^{itx} - 1 - \frac{itx}{1+x^2})(1-\theta)\alpha|x|^{-(\alpha+1)}dx\right]. \tag{4}$$

Definition 3.3 (**stable law**)**.** Distribution with characteristic function of the form 4.

Remark 3.4 (Alternative representation) —

$$\exp\left[itc - b|t|^\alpha(1 + i\kappa\operatorname{sgn}(t)w_alpha(t))\right]$$

with:

$$k = 2\theta - 1 \in [-1,1], \quad w_alpha(t) = \begin{cases}\tan(\frac{\pi\alpha}{2}), & \alpha \neq 1 \\ \frac{2}{\pi}\log|t| & \alpha = 1\end{cases}$$

for $0 < \alpha \le 2$. See (Brenman. 1968, page 204-206)

Example 3.5

If $\alpha = 2$, the stable law becomes Gaussian.

Example 3.6

If $\alpha = 1$, $c = 0$, $\kappa = 0$, we get the Cauchy distribution.

Example 3.7

If $\alpha = \frac{1}{2}$, $c = 0$, $\kappa = 1$, $b = 1$, we get density function:

$$(2\pi y^3)^{-1/2} \exp(-1/2y), \quad y \geq 0.$$

Remark 3.8 — The density function are not known except for the above 3 cases.

Theorem 3.9

Y is stable law $\iff$ Y is the weak limit of $\frac{\sum_{i=1}^{n} X_i - b_n}{a_n}$ for a given sequence of i.i.d. X_i's.

Example 3.10

Let $X_1, X_2, \dots$ be i.i.d. with a density function that is symmetric about 0 and continuous and positive at 0. We claim:

$$\frac{1}{n} \sum_{i=1}^{n} \frac{1}{X_i} \implies \text{a Cauchy distribution } (\alpha = 1, \kappa = 0).$$

Proof. Consider when $x \to \infty$:

$$\mathbb{P}\left(\frac{1}{X_1} > x\right) = \mathbb{P}(0 \leq X_1 < \boldsymbol{x}^{-1}) = \int_0^{\boldsymbol{x}^{-1}} f(y)dy = \frac{f(0)}{x}$$

Similarly, for the left tail:

$$\mathbb{P}\left(\frac{1}{X_1} < -x\right) = \frac{f(0)}{x}.$$

In addition, we have $\theta = \frac{1}{2}$ by assumption (of symmetry), giving us $b_n = 0$. Thus:

$$\mathbb{P}\left(\left|\frac{1}{X_1}\right| > a_n\right) = \frac{2f(0)}{a_n} = \frac{1}{n} \implies a_n = 2f(0) \cdot n$$

Thus:

$$\frac{1}{n} \sum_{i=1}^{n} X_i \implies \text{Cauchy}.$$

□

Remark 3.11 — Whenever we prove with stable law, we check the tail behavior.

Note that the centralization constant is not necessary if $\alpha < 1$.

Consider $X_1, X_2, \ldots$ i.i.d. with exact distribution:

$$\mathbb{P}(X_1 > x) = \theta x^{-\alpha} \quad \mathbb{P}(X_1 < -x) = (1-\theta)x^{-\alpha}, \quad 0 < \alpha < 2, |x| \geq 1.$$

In this case, we know that $a_n = n^{1/\alpha}$. Meanwhile, we have:

$$\begin{aligned} b_n &= n\mathbf{E}X_1\mathbb{1}(|X_n| < a_n) \\ &= n\int_1^{n^{1/\alpha}} (2\theta - 1)\alpha x^{-\alpha} dx \sim \begin{cases} cn & \alpha > 1 \\ cn\log n & \alpha = 1 \\ cn^{1/\alpha} & \alpha < 1 \end{cases}. \end{aligned}$$

Note that if $\alpha < 1$, we don't need to subtract by b_n to have convergence, but we will have a different limit if we do/don't.

Remark 3.12 — If $\alpha > 1$, the constant cn.

3.2 Infinitely Divisible Distribution

As we mentioned previously, the stable law is the limit of $\frac{\sum_{i=1}^n X_i - b_n}{a_n}$ for a given sequence of i.i.d. X_i's.

On the other hand, the **infinitely divisible distribution** is the limit of $\frac{\sum_{i=1}^n X_{n,i} - b_n}{a_n}$ for triangular array with i.i.d. $X_{n,i}$'s for each n.

Example 3.13

Gaussian $\in$ stable law, Poisson $\in$ infinitely divisible law

Here we won't derive the infinitely divisible distributions, but we will state some results. If interested, consult the textbook.

Example 3.14 (Poisoon as an infinitely divisible distribution)

Poisson is the limit of triangular array of Bernoulli r.v. $X_{n,1}, \ldots, X_{n,n}$ with:

$$\mathbb{P}(X_{n,i} = 1) = 1 - \mathbb{P}(X_{n,i} = 0) = \frac{\lambda}{n}$$

Note that the c.f. of Poisson(λ) is $\exp(\lambda(e^{it} - 1))$ which is not a stable law.

Theorem 3.15 (**Levy-Khinchin Theorem**)

Z has an infinitely divisible distribution $\iff$ its c.f. is of the form:

$$\varphi(t) = \exp\left[ict - \frac{\sigma^2 t^2}{2} + \int \left(e^{itx} - 1 - \frac{itx}{1+x^2}\right)\mu(dx)\right]$$

where μ is a measure (not necessarily probability measure) with:

$$\mu(\{0\}) = 0, \qquad \int \frac{x^2}{1+x^2}\mu(dx) < \infty.$$

Example 3.16 (Examples of infinitely divisible distributions)
If we consider:

1. Gaussian, $\mu = 0$ measure.
2. Poisson, we have:
$$c = \int \frac{x}{1+x^2}\mu(dx), \quad \sigma^2 = 0, \quad \mu(\{1\}) = \lambda \text{ (single point mass)}$$
3. all stable law: $\sigma^2 = 0$.
4. Compound Poisson:
Let $\xi_1, \xi_2, \ldots$ be i.i.d. and $N(\lambda)$ be an independent Poisson(λ) with c.f.:
$$\varphi(t) = \mathbf{E}\exp(it\xi_1) = \int \exp(itx)\mu_\xi(dx).$$
Let $Z = \xi_1 + \ldots + \xi_{N(\lambda)}$ is infinitely divisible:
$$\mathbf{E}\exp(itZ) = \exp(-\lambda(1-\varphi(t))) = \exp\left[\lambda \int (e^{itx} - 1)\right]$$

This is the end of this chapter about stable law.

3.3 Functional Limit Theorems

Our aim for this chapter is to study the weak convergence in the space $C[0,1]$, which is the space of all continuous functions supported on $[0,1]$.

Remark 3.17 — The choice of considering on $[0,1]$ is just for convenience. We can do on other set as long as they are compact.

The weak convergence of a function means that as $n \to \infty$, $X_n(t) \to X(t), t \in [0,1]$ in distribution. First, we will consider the weak convergence on a much simpler space, namely $\mathbb{R}^k$.

We denote a **random vector** as:
$$\vec{X} = (X^1, \ldots, X^k) : (\Omega, \mathscr{F}) \to (\mathbb{R}^K, \mathscr{B})$$
so that $X^{-1}(B) \in \mathscr{F}, \forall B \in \mathscr{B}(\mathbb{R}^k)$.

Consider a random vector sequence $X_n = (X_n^1, \ldots, X_n^k)$, $n = 1, \ldots,$ with c.d.f. $F_n : \mathbb{R}^k \to [0,1]$:
$$F_n(\vec{x}) = \mathbb{P}(X_n^1 \le x_1, \ldots, X_n^k \le x_n).$$

Definition 3.18 (Convergence of a random vector sequence)**.** We say that F_n converges to F weakly if $F_n(x) \to F(x)$ at all continuity point of F, denoted by $F_n \implies F$ Further we say X_n converges to X weakly (in distribution) if $F_n \implies F$, denoted by $X_n \implies X$.

Definition 3.19 (Alternative definition of $X_n \implies X$)**.** We say $X_n \implies X$ if for any bounded continuous function: $f : \mathbb{R}^k \to \mathbb{R}$, $\mathbf{E}f(X_n) \to \mathbf{E}f(X)$.

Definition 3.20 (tightness)**.** We say a sequence of probability measure μ_n on $(\mathbb{R}^k, \mathscr{B}(\mathbb{R}^k))$ is **tight** if for any $\epsilon > 0, \exists M = M_\epsilon > 0$, s.t.:

$$\mu_n([-M, M]^k) \geq 1 - \epsilon.$$

Definition 3.21 (**characteristic function of random vector**)**.** The c.f. of $X = (X^1, \ldots, X^k)$ is defined as:

$$\varphi_X(t) = \mathbf{E} \exp(it \cdot X) = \mathbf{E} \exp\left(i \sum_{a=1}^{k} t_a X_a\right), \quad t = (t_1, \ldots, t_k) \in \mathbb{R}^k$$

Theorem 3.22 (inversion formula)

If $A = [a_1, b_1] \times \ldots \times [a_k, b_k]$ with $\mu(\partial A) = 0$:

$$\mu(A) = \lim_{T \to \infty} (2\pi)^k \int_{[-T,T]^k} \prod_{j=1}^{k} \left(\frac{e^{-is_j a_j} - e^{is_j b_j}}{is_j}\right) \varphi(s) ds$$

Theorem 3.23 (continuity theorem)

Let $X_n, 1 \leq n \leq \infty$ be a random vectors with c.f. φ_n, then:

$$X_n \implies X_\infty \iff \varphi_n(t) \to \varphi_\infty(t)$$

for any given $t \in \mathbb{R}$.

4 February 22nd, 2022

4.1 Functional Limit Theorems Continued

Theorem 4.1 (**Cramer-Wold device**)

A sufficient condition for $X_n \implies X_\infty$ is that:

$$\theta \cdot X_n \implies \theta \cdot X_\infty$$

for any given $\theta \in \mathbb{R}^k$.

Proof. According to the continuity theorem, we just need to show that $\varphi_n(t) \to \varphi_\infty(t)$. Thus, if we choose $t = \theta$, then this is true for any t. $\square$

Theorem 4.2 (multivariate CLT)
Let $X_1, X_2, \dots$ be i.i.d. random vectors in $\mathbb{R}^k$ with

$$\mathbf{E}X_1 = \mu = \begin{pmatrix} \mu_1 \\ \vdots \\ \mu_k \end{pmatrix}, \quad \Gamma_{ij} = \mathbf{E}(X_n(i) - \mu_i)(X_n(j) - \mu_j),$$

If $S_n = \sum_{i=1}^{n} X_i$, then:

$$\frac{1}{\sqrt{n}}(S_n - n\mu) \implies \chi$$

which is the multivariate Gaussian $N(0, \Gamma)$.

Definition 4.3 (**multivariate Gaussian**)**.** The j.p.d.f. is:

$$(2\pi)^{-\frac{k}{2}} \det(\Gamma)^{-\frac{1}{2}} \exp(-\frac{1}{2}X^T\Gamma X)$$

and the c.f. is:

$$\mathbf{E}e^{itX} = \exp(-\frac{1}{2}t^T\Gamma t)$$

Proof. WLOG, we assume $\mu = 0$ by considering $X_n' = X_n - \mu$. Let $\theta \in \mathbb{R}^k$, meaning $\theta \cdot X_n$ is a r.v. We have:

$$\mathbf{E}\theta \cdot X_n = 0, \quad \mathbf{E}(\theta \cdot X_n)^2 = \theta^T\Gamma\theta$$

By 1D CLT:

$$\frac{\sum_{i=1}^{n} \theta \cdot X_n}{\sqrt{n}} \implies \theta \cdot \chi$$

By the Cramer-Wold device, we have:

$$\frac{S_n}{\sqrt{n}} \implies \chi.$$

□

Now, we will return to a functional space. For this, let us review some common functional spaces:

Definition 4.4 ($C\left[0,1\right]$ Functional Space)**.** Similar to random variables, we define the space and metric:

space: all continuous functions on $[0, 1]$.

metric/topology: uniform topology $\rho(x, y) = \sup_{0\le t\le 1} |x(t) - y(t)|$.

Borel σ-field: generated by open sets.

For notation, we have: $C = C[0, 1]$, $\mathscr{C} = \mathscr{B}(C[0, 1])$.

Definition 4.5 (random function in C)**.** Given a probability space $(\Omega, \mathscr{F}, \mathbb{P})$, we say a map $X : \Omega \to C$ a random function if:

$$X^{-1}(A) = \{\omega : X(\omega) \in A\} \in \mathscr{F}, \quad \forall A \in \mathscr{C}.$$

Remark 4.6 — Some other possible notations include $X_t(\omega)$, $X(t;\omega)$, $t \in [0,1]$.

For a fixed ω, we call $X(\cdot;\omega)$ the **trajectory/sample path/realization of X**.

We can first think of $X(t)$ as a measurable map:

$$(\Omega, \mathscr{F},) \overset{X(t)}{\to} (C, \mathscr{C})$$

To consider it as a probability measure, we have the push forward/induced measure, giving us the distribution of $X(t)$.

Definition 4.7 (Distribution of $X(t)$)**.** If we think the random map as the map between two probability space:

$$(\Omega, \mathscr{F}, \mathbb{P}) \overset{X(t)}{\to} (C, \mathscr{C}, \underbrace{\mathbb{P} \circ X^{-1}}_{\mu_X})$$

with:

$$\mu_X(A) = \mathbb{P} \circ X^{-1}(A) = \mathbb{P}(X^{-1}(A)), \quad A \in \mathscr{C}.$$

This defines a probability measure in the new measurable space.

This is a bit more abstract than probability measures on $\mathbb{R}$ or $\mathbb{R}^k$, since for the simpler spaces, we have much simpler ways to identify them. For example, for $\mathbb{R}$, using the Stieltjes measure function, if we have a Stieltjes distribution, we don't need to know the value of the distribution on all Borel sets, we only need on all half-lines. For $\mathbb{R}^k$, we can use the joint distribution function. Naturally, we ask if we can identify the μ_X in $C[0,1]$. In other words, we don't want to find μ_X on all A, but only on some special class of events. To do this, we will first introduce the finite-dimensional distribution on μ_X.

4.2 Finite-Dimensional Distribution

The general idea is to pick a few time points on the sample path and consider their joint distributions.

Definition 4.8 (natural projection)**.** The natural projection $\Pi_{t_1,\ldots t_k} : C \to \mathbb{R}^k$:

$$\Pi_{t_1,\ldots t_k} = X(t) \to (X(t_1), \ldots X(t_n)), \quad t \in [0,1]$$

with $\mathbb{R}^k$ being equipped with the usual Euclidean metric.

Since this is a continuous map, it is also measurable. With this, we can induce a probability measure.

$$(C, \mathscr{C}, \mu_X) \overset{\Pi_{t_1,\ldots t_k}}{\to} (\mathbb{R}^k, \mathscr{B}(\mathbb{R}^k), \mu_X \circ \Pi^{-1}_{t_1,\ldots t_k})$$

Meaning $\mu_X \circ \Pi^{-1}_{t_1,\ldots t_k}$ is a finite dimension distribution, with:

$$\mu_X \circ \Pi^{-1}_{t_1,\ldots t_k}(B_1 \times B_2 \times \ldots \times B_k) = \mu_X(\Pi^{-1}_{t_1,\ldots t_k}(B_1 \times B_2 \times \ldots \times B_k))$$

Remark 4.9 — Since $\mu_X \circ \Pi^{-1}_{t_1,\dots t_k}$ is on $\mathbb{R}^k$, we only need to consider it on rectangles.

As such, given μ_X, we can get the probability distribution. However, what we're more interested in is the reverse, i.e. given $\mu_X \circ \Pi^{-1}_{t_1,\dots t_k}$, fix a μ_X on C. Another view of this finite-dimensional distribution is considering:

$$(\Omega, \mathscr{F}, \mathbb{P}) \overset{X_{t_1,\dots,t_k}=(X(t_1),\dots,X(t_k))}{\to} (\mathbb{R}^k, \mathscr{B}(\mathbb{R}^k), \mathbb{P} \circ X^{-1}_{t_1,\dots,t_k})$$

Now, we have:

$$\mathbb{P} \circ X^{-1}_{t_1,\dots,t_k}(B_1 \times B_2 \times \dots \times B_k) = \mathbb{P}((X(t_1),\dots,X(t)) \in B_1 \times \dots B_k)$$
$$\implies \mathbb{P}\{\omega : (X(t_1;\omega)\dots X(t_k;\omega)) \in B_1 \times \dots \times B_k\}$$

This gives us a way to get the finite dimensional distribution even if we don't know μ_X. Now our goal is to choose arbitrary fixed points t_k and fix the μ_X.

Suppose we have a collection of distributions $\nu_{t_1,\dots,t_k}$ on $\mathbb{R}^n$ for all $t_1,\dots,t_k \in [0,1]$, and all $k \in \mathbb{N}$, we want to know when we can regard them as the class of finite-dimensional distribution of a measure ν on $(C, \mathscr{C})$ and whether they can uniquely describe ν. Let us first find some requirements for this to be a finite-dimensional distribution:

1. If $\nu_{t_1,\dots,t_k}$'s are indeed finite-dimensional distributions of some $\mathcal{B}$, then:
$$\nu_{t_1,\dots,t_k}(B_1 \times \dots \times B_k) = \mathbb{P}((X(t_1),\dots X_(t_k)) \in B_1 \times \dots B_k)$$
for some random function $X(t)$ with distribution ν. Then:
$$\nu_{t_1,\dots,t_k}(B_1 \times \dots \times B_k) = \nu_{\Pi(t_1),\dots,\Pi(t_k)}(B_{\Pi(t_1)} \times \times B_k),$$
where Π is any permutation of 1 to k.
2. We should also be able to add to k and maintain consistency:
$$\nu_{t_1,\dots,t_k}(B_1 \times \dots \times B_k) = \nu_{t_1,\dots,t_k,t_{k+1}}(B_1 \times \dots \times B_k \times \mathbb{R}).$$

It turns out that these two consistency conditions are enough.

Theorem 4.10 (Kolmogorov extension theorem)
Let $T \in [0,1]$ be some interval. For any finite sequence of distinct time $t_1,\dots t_k \in T$. Let $\mu_{t_1,\dots t_k}$ be a probability measure on $\mathbb{R}^k$. Suppose these measures satisfy the two consistency conditions:

1. $\nu_{\Pi(t_1),\dots,\Pi(t_k)}(B_{\Pi(t_1)} \times \times B_k) = \nu_{t_1,\dots,t_k}(B_1 \times \dots \times B_k)$
2. $\nu_{t_1,\dots,t_k,t_{k+1}}(B_1 \times \dots \times B_k \times \mathbb{R}) = \nu_{t_1,\dots,t_k}(B_1 \times \dots \times B_k)$,

Then there exists some probability space $(\Omega, \mathscr{F}, \mathbb{P})$ and a random function $X : (\Omega, \mathscr{F}) \to (C, \mathscr{C})$ such that:

$$\nu_{t_1,\dots,t_k}(B_1 \times \dots \times B_k) = \mathbb{P}((X(t_1)\dots X(t_k)) \in B_1 \times \dots \times B_k)$$

and the distribution of $X(t)$, i.e. ν is uniquely determined by $\nu_{t_1,\dots t_k}$.

Proof. For proof, see Billingsley [Bil86]. In the book it is called Kolmogorov existence theorem. ☐

This means that it is enough to know all the finite-dimensional distributions.

4.3 Weak Convergence in $(C, \mathscr{C})$

let us consider random functions: $X_1(t), \ldots, X_n(t) \in C$ with distributions $\mu_1, \ldots, \mu_n$.

Definition 4.11 (**weak convergence**)**.** We say $\mu_n \implies \mu$ or $X_n(t) \implies X(t)$ for some $X(t)$ with distribution μ if:

$$\mathbf{E}f(X_n) \to \mathbf{E}f(X)$$

for any bounded and continuous functional. Here bounded means $\sup_{X \in C} |f(X)| \leq M$ for some M.

Theorem 4.12 (**continuous mapping theorem**)

If $h : C \to \mathbb{R}$ is continuous. Then $X_n \implies X$ implies $h(X_n) \implies h(X)$, with X_n being a random function.

Proof. Let $g : \mathbb{R} \to \mathbb{R}$ be any bounded continuous function. Then, $g \circ h$ is again a bounded continuous functional. If $X_n \implies X$, then:

$$\mathbf{E}g(h(X_n)) \implies \mathbf{E}g(h(X))$$

by definition of. Thus, $h(X_n) \implies h(X)$. □

Example 4.13

If $X_n \implies X$ in C, then $\sup_{t \in [0,1]} X_n(t) \implies \sup_{t \in [0,1]} X(t)$ using triangular inequality.

Example 4.14

If $X_n \implies X$ in C, then $\int_0^1 X_n(t)dt \implies \int_0^1 X(t)dx$.

Now, our task is to prove weak convergence in C. We first need prove the finite-dimensional convergence and tightness.

Definition 4.15 (finite-dimensional convergence)**.** We say that the finite-dimensional convergence holds for $X_n \implies X$ if for any given $t_1, t_2, \ldots t_k$ and any $k \in \mathbb{N}$, there is:

$$(X_n(t_1), \ldots, X_n(t_k)) \implies (X(t_1), \ldots, X(t_k)).$$

Remark 4.16 — Finite-dimensional convergence is not convergence-determining in C. For example, let us consider:

$$Z_n(t) = nt\mathbb{1}(t \in [0, 1/n]) + (2 - nt)\mathbb{1}(t \in [1/n, 2/n])$$

we define random function:

$$X_n(t; \omega) = Z_n(t), \quad \forall \omega \in \Omega$$

hence, $\mathbb{P}(X_n(t) = Z_n(t)) = 1$. Let $X(t; \omega) = 0$, $\forall \omega, \forall t$. For any given

$t_1, \dots t_k \in [0,1]$:

$$(X_n(t_1), \dots, X_n(t_k)) \implies (X(t_1), \dots, X(t_k))$$

However, $X_n(t) \not\Longrightarrow X(t)$, otherwise:

$$\sup_{t\in[0,1]} X_n(t) = 1 \implies \sup_{t\in[0,1]} X(t) = 0$$

by continuous mapping theorem.

Note that for any fixed n, the finite-dimensional distribution of $(X_n(t_1), \dots, X_n(t_k))$'s can be used to determine the distribution of $X_n(t)$. However, here $t_1, \dots, t_k$ can also be in the interval $(0, 2/n]$. But in finite-dimensional convergence, $t_1, \dots t_k$ are given at the beginning, i.e. independent of n, meaning it can jump out of the interval.

5 February 24th, 2022

5.1 Relative Compactness of $\{\mu_n\}$

Definition 5.1. definition

We say $\{\mu_n\}$ is relatively compact if for any subsequence of μ_n, say μ_{n_k}, one can find a further subsequence $\mu_{n_{k_i}}$ which converges weakly.

Lemma 5.2

If $\mu_n \implies \mu$ in the finite-dimensional sense and if $\{\mu_n\}$ is relatively compact, then $\mu_n \implies \mu$ ($\mu_n \to \mu$ weakly).

Proof. finite-dimensional convergence $\implies \mu_n \circ$ $\square$

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.

[UZ11] V.V. Uchaikin and V.M. Zolotarev. *Chance and Stability: Stable Distributions and their Applications*. Modern Probability and Statistics. De Gruyter, 2011. ISBN: 9783110935974. URL: https://citeseerx.ist.psu.edu/viewdoc/download?doi=%2010.1.1.117.6505&rep=rep1&type=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.

[Nol20] J.P. Nolan. *Univariate Stable Distributions: Models for Heavy Tailed Data*. Springer Series in Operations Research and Financial Engineering. Springer International Publishing, 2020. ISBN: 9783030529154. URL: https://link.springer.com/content/pdf/10.1007/978-3-030-52915-4.pdf.

Index

characteristic function of random vector, 20
continuous mapping theorem, 24
Cramer-Wold device, 20

empirical distribution, 3

functional limiting theorem, 3

heavy tailed random variables, 6

infinitely divisible distribution, 2, 18

Kolmogorov extension theorem, 23
Kolmogorov–Smirnov Statistics, 3

Levy's Continuity Theorem, 7
Levy-Khinchin Theorem, 18

martingale concentration, 4
martingale differences, 4
multivariate CLT, 21
multivariate Gaussian, 21

Poisson point process, 5
Possion convergence, 4

random processes, 3
random vector, 19

slowly varying function, 12
stable law, 2, 12, 16

tight, 20
time homogeneous, 5
trajectory/sample path/realization of X, 22

weak convergence, 24