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** $F_n(t) = \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}(X_i \leq t)$ 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 $X_{n,m}$, $1 \leq m \leq n$ be independent r.v. with $\Pr(X_{n,m} = 1) = 1 - \Pr(X_{n,m} = 0) = \beta_{n,m}$. If $\sum_{m=1}^{n} \beta_{n,m} \to \lambda$ and $\max_m \beta_{n,m} \to 0$, then, $S_n = \sum_{m=1}^{n} X_{n,m} \implies \text{Poisson}(\lambda)$.

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 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. $m(A_i) \overset{D}{=} \text{Poisson}(\mu(A_i))$.

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.

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

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:

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

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:

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

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:

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

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

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

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

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

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:

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

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.