Math remark
fondation for analysis and probability

X

ElegantLaTeX Program

Update: March 7, 2025

Contents

1 Genenralisation of several inequalities

1.1 Jense,Hölder,Minkeoski's inequalities

The description of the Jense inequality depends on the properties of the convex function, it is a strong inequalities which can be applied in many places. We should pay a little attention to the outline of the section:

$$\text{Jense} \Rightarrow \text{Hölder} \Rightarrow \text{Minkeoski}$$

Theorem 1.1 (Jense)

Suppose $\varphi : I \to \mathbb{R}$ is a convex function defined on an interval, $(X, \mathcal{E}, \mu)$ is a probability space and $f \in L^1(X)$ with $imf \subset I$, then $\int_X f d\mu \in I$ and $\varphi \circ f$ is integrable such that

$$\varphi(\int_X f d\mu) \leq \int_X \varphi \circ f d\mu$$

The equality holds iff f is constant almost everywhere.

Proof. ∎

Remark There are many other forms of the Jense's inequality, we take some examples.

- **Finite forms**: the motivation of the inequality comes from the definition of the convex function, i.e. the real-valued function satisfies

$$f(tx + (1 - t)y) \leq tf(x) + (1 - t)f(y)$$

for any defined x, y and $t \in [0, 1]$, here a question about the distribution of the weights appears, which is the core of the convex function. we can generalize the inequalities by appling the weights to the n different points in interval such that $\sum_i w_i = 1$, then we can conclude the inequality:

$$f(\sum_i w_i x_i) \leq \sum_i w_i f(x_i)$$

notice that the defined domain usually is a convex set, which ensures the effectivity of $f(\sum_i w_i x_i)$.

- **Expectation:** By a simple change of the notation, the Jense inequality in a probaility space can be wriiten as the form:

$$\varphi(E[X]) \leq E[\varphi(X)]$$

Applying a classic convex function $t \mapsto t^2$ we can get the important inequality in probability:

$$E^2[X] \leq E[X^2]$$

- **Concave:** some function like $t \mapsto lnt$ is a concave function, the Jense's inequality can be just changed the order of the inequality. The reason is simple, if f is a concave function , then $-f$ will be a convex function.

The classic proof of the Hölder's inquality covers the inequality of Young:

$$ab \leq \frac{a^p}{p} + \frac{b^q}{q}$$

for any $a, b \geq 0$ and $p, q > 1$ such that $1/p + 1/q = 1$. The complete proof can be found in **[Rudin 1 Ex 6.10]**. I don't choose the proof here for a comparison of strength of the different inequality. Here the proof is elegant and given by Mon.Mardare in TD, and a similar proof via Jense can be found in **[SU TD1 EX15-16]**.

Theorem 1.2 (Hölder)

Suppose that $p, q > 1$ and $1/p + 1/q = 1$, for any two mesurable functions $f, g : (X, \mathcal{E}, \mu) \to \mathbb{C}$, we have

$$\int_X |fg| d\mu \leq (\int_X |f|^p d\mu)^{1/p} (\int_X |g|^q d\mu)^{1/q}$$

the equality holds iff $|f| = c|g|$ almost everywhere (u-p.p) for some constant c.

Proof. Let $u = \frac{|f|}{\|f\|_p}$ and $v = \frac{|g|}{\|g\|_q}$, then notice that $\|u\|_p = \|v\|_q = 1$. We know $t \mapsto lnt$ is a concave function on $(0, +\infty)$, so we can estimate by Jense's inequality

$$ln(uv) = ln(u+v) = \frac{1}{p}lnu^p + \frac{1}{q}lnv^q$$
$$\leq ln(\frac{1}{p}u^p + \frac{1}{q}v^q)$$

by the monotone of the function, we have $uv \leq \frac{1}{p}u^p + \frac{1}{q}v^q$, hence we can conclude

$$\frac{\int_X |fg|d\mu}{(\int_X |f|^p d\mu)^{1/p}(\int_X |g|^q d\mu)^{1/q}} = \int_X uv d\mu$$
$$\leq \frac{1}{p}\int_X u^p d\mu + \frac{1}{q}\int_X v^q d\mu$$
$$= \frac{1}{p}\|u\|_p^p + \frac{1}{q}\|v\|_q^q = 1$$

finally, we discuss the equality. By Jense, we know the equality holds iff $u = v$. Notice that if $u = 0$ or $v = 0$ almost everywhere, then the inequality can be reduced to $0 \leq 0$, it holds, otherwise we can get that $|f| = \frac{\|f\|_p}{\|g\|_q}|g|$, so we finish the proof. ∎

Corollary 1.2.1 (Cauchy-Schwartz)

For any two square integrable function $f, g \in L^2_{\mathbb{C}}(\mathbb{R}^n)$, we have

$$|\int_{\mathbb{R}^n} f\overline{g} dl|^2 \leq \int_{\mathbb{R}^n} |f|^2 dl \cdot \int_{\mathbb{R}^n} |g|^2 dl$$

The equality holds iff $|f| = c|g|$ almost everywhere (u-p.p) for some constant c.

Proof. Although it is the special case of Hölder when $p = q = 2$, but usually in a Hilbert space we have the beautiful form as following

$$| < u, v > | \leq \|u\|\|v\|$$

the inequality has a good geometric intuition, and the proof of it is very beautiful and elementry. Notice $< u, v > \in \mathbb{C}$, so there exists $z \in \mathbb{C}$ such that $|z| = 1$ and $z < u, v >= | < u, v > |$, and we let $p(t) =< tzu + v, tzu + v >$ defined on $\mathbb{R}$, then

$$p(t) = t^2 z\overline{z} < u, u > + tz < u, v > + t\overline{z} < v, u > + < v, v >$$
$$= t^2 z\overline{z} < u, u > + tz < u, v > + t\overline{z < u, v >} + < v, v >$$
$$= t^2\|u\|^2 + 2| < u, v > |t + \|v\|^2$$

so $p(t)$ can be arranged to be a quadratic polynomial with respect to real value t, and $p(t) = \|tzu + v\|^2 \geq 0$, so we have suiffsant and necessary condition that

$$\Delta = 4| < u, v > |^2 - 4\|u\|^2\|v\|^2 \geq 0$$

which is the inequality we hope to get, and $\Delta = 0$ happens iff the polynomial satisfies $p(t) = (t\|u\| + \|v\|)^2 = 0$. ∎

Theorem 1.3 (Minkeoski)

For any $p \geq 1$, suppose that $f, g : (X, \mathcal{E}, \mu) \to \mathbb{C}$ are two mesurable functions, then we have

$$(\int_X |f + g|^p d\mu)^{1/p} \leq (\int_X |f|^p d\mu)^{1/p} + (\int_X |g|^p d\mu)^{1/p}$$

The equality holds iff $|f| = c|g|$ almost everywhere (u-p.p) for some constant c.

Proof. If ∎

Review some basic inequalities (discrete)...

1.2 Markov, Tchebychev, Cantelli's inequalities

This section covers some basic inequalities in properties, They are always very useful when estimation. And in this section ,we always use $(\Omega, \mathcal{T}, P)$ to denote a probability space

Theorem 1.4 (Markov)

Suppose that $Y : \Omega \to \mathbb{R}^+$ is a random variable, then for any $t > 0$, we have

$$P(Y \geq t) \leq \frac{E(Y)}{t}$$

Proof. We notice that $t \cdot \mathbb{1}_{\{Y \geq t\}} \leq Y$, it is easily to prove. If $x \in \{Y \geq t\}$, then $t \cdot \mathbb{1}_{\{Y \geq t\}}(x) = t \leq Y(x)$, otherwise $t \cdot \mathbb{1}_{\{Y \geq t\}}(x) = 0 \leq Y(x)$, so by monotone of the expectation

$$E[t \cdot \mathbb{1}_{\{Y \geq t\}}] = tE[t \cdot \mathbb{1}_{\{Y \geq t\}}] = tP(Y \geq t)$$

∎

Theorem 1.5 (Tchebychev)

Suppose that $Y : \Omega \to \mathbb{R}$ is a random variable in $L^2(\Omega)$ (square integrable), then for any $t > 0$ we have

$$P(|Y - E[Y]| \geq t) \leq \frac{V(Y)}{t^2}$$

Proof. We apply Markov to prove it, we can estimate

$$
\begin{aligned}
P(|Y - E[Y]| \geq t) &= P((Y - E[Y])^2 \geq t^2) \\
&\leq \frac{E[(Y - E(Y))^2]}{t^2} \\
&= \frac{V(Y)}{t^2}
\end{aligned}
$$

∎

Theorem 1.6 (Cantelli)

Suppose that $Y : \Omega \to \mathbb{R}$ is a random variable in $L^2(\Omega)$ (square integrable), then for any $t > 0$ we have

$$
P(Y - E[Y] \geq t) \leq \frac{V(Y)}{V(Y) + t^2}
$$

$$
P(|Y - E[Y]| \geq t) \leq \frac{2V(Y)}{V(Y) + t^2}
$$

Proof. The proof of the inequality need to use inequality: for any $a, b > 0$,

$$
P(X \geq a) \leq \frac{E[(X + b)^2]}{(a + b)^2} \tag{1}
$$

We write $P(X \geq a) = E[\mathbb{1}_{\{X \geq a\}}]$, then we just need to prove $\mathbb{1}_{\{X \geq a\}} \leq (\frac{X+b}{a+b})^2$. For any $w \in \{X \geq a\}$, we have $(\frac{X+b}{a+b})^2(w) \geq (\frac{a+b}{a+b})^2 = 1 = \mathbb{1}_{\{X \geq a\}}(w)$, otherwise will conclude $(\frac{X+b}{a+b})^2 \geq 0$, so the inequality 1 is valid.

For proove the Cantelli, we let $Z = \frac{Y - E[Y]}{\sigma(Y)}$, then by above inequality

$$P(Y \geq t) \leq \frac{E[(Y+b)^2]}{(t+b)^2}$$
$$= \frac{E[Y^2] + 2bE[Y] + b^2 E[1]}{(t+b)^2}$$
$$= \frac{1 + b^2}{(t+b)^2}$$

We can put $b = 1/t$, then $\frac{1+b^2}{(t+b)^2} = \frac{1}{1+t^2}$, and then rewrite the inequality we get $P(Y - E[Y] \geq t) \leq \frac{V(Y)}{V(Y)+t^2}$. and notice that

$$P(Y - E[Y] \leq -t) = P((-Y) - E[-Y] \geq t)$$

and $V(-Y) = V(Y)$ implies the same upper bound, so we finish the proof. ∎

The importance of the three inequalities is the estimation of the upper bound of some probability, if we can know the the expected value, we can get an upper bound

by Markov; if we also know the variance, a more precise upper bound can be given by Tchebychev and Cantelli.

Cantelli's inequality is a work based on the Tchebychev's inequality, if we rewrite the Tchebychev's inequality by take $t = \sigma(Y)t$, then we have

$$P(|X - \mu| \leq \sigma t) \geq \frac{1}{t^2}$$

Which always appears in the some statistical object like confidence interval, but it is a two-side interval estimation. If we rewrite Cantelli's inequality similarly:

$$P(X - \mu \leq \sigma t) \geq \frac{1}{1 + t^2}$$

$$P(|X - \mu| \leq \sigma t) \geq \frac{2}{1 + t^2}$$

We can find that Cantelli's inequality can do one-side estimation and it provides a worse bound then Tchebychev's when $t \geq \sigma$.

One appliaction of the Tchebychev is to prove the law of large theorem, suppose that $X_1, X_2, ..., X_n$ are the sequence of random variable with same distribution, that means

$$E[\frac{X_1 + ... + X_n}{n}] = \mu \quad V[\frac{X_1 + ... + X_n}{n}] = \frac{\sigma^2}{n}$$

then using the Tchebycheve's inequality we have

$$P(|\frac{X_1 + ... + X_n}{n} - \mu| \geq c) \leq \frac{\sigma^2}{nc^2}$$

Hence as $n \to \infty$, then the probability that the average of the random variable is not the exceptation μ will be zero, that what law of large number want to say: when the number of sample is enough big, then the mean value of the sample is nearly the real mean value.

2 Classic anaylsis theory and technic

2.1 Power series and analytic function

This section conatains Three parts:

- overview for power series and the important result
- real analytic function and Abel's theorem
- the superiority of the holomorphic function

Power series is an special case of the series of the function, it gives by the follwoing sequence:

$$f_n : D \to \mathbb{C}, \quad z \mapsto a_n z^n$$

more general, we can write it as the form $a_n(z - z_0)^n$, but notice that we can just change variable by $u = z - z_0$, it is a translation, so clearly a diffeomorphism. In above statement D denote the domain where the series will be well-defined, i.e. convergent, so here needs a clear definition.

Lemma 2.1 (Abel's Lemma)

Given a complex-valued sequence $(a_n)_{n \in \mathbb{N}}$, if there exists $z_0 \in \mathbb{C}$ such that $(a_n z_0^n)_{n \in \mathbb{N}}$ is bounded, then we can conclude that

-If $|z| < |z_0|$, then power series $\sum_{n \in \mathbb{N}} a_n z^n$ converge.

-If $|z| < |z_0|$, then power series $\sum_{n \in \mathbb{N}} a_n z^n$ diverge.

Proof. The bounded sequnce implies a real number M such that $a_n z_0^n \le M$, so

$$|a_n z^n| = |a_n z_0^n \cdot \frac{z^n}{z_0^n}| \le M |\frac{z}{z_0}|^n$$

so clearly the power series converge if $|z| \le |z_0|$ by M-test.

If $|z| > |z_0|$, then $(a_n z^n)_{n \in \mathbb{N}}$ is not bounded, so immediately it cant converge to zero, so the power series diverge. ∎

The lemma inspire us that the domaine is nearly a disk, although we can not determine the convergence of the boundary, but we can formally define the **radius of convergence** of a power series $\sum_{n \in \mathbb{N}} a_n z^n$ is given by

$$R = \sup\{r \geq 0 | (a_n r^n)_{n \in \mathbb{N}} \text{ is bounded}\}$$

and we call $D = D(0, R)$ the **domaine of convergence**. A more simple form of the raduis of convergence is given by Cauchy and d'Alembert:

$$1/R = \limsup |a_n|^{1/n} = \lim |\frac{a_{n+1}}{a_n}|$$

and we convented here $1/0 = \infty$ and $1/\infty = 0$.

The convergence of the power series provides a good case of function, analytic function, which is the theme of the next section. We put function

$$f : \overline{D(0, r)} \to \mathbb{C}, \quad dz \mapsto \sum_{n \in \mathbb{N}} a_n z^n$$

where $r \in (0, R)$, and we suppose here $R > 0$. The function will be well-defined and continous since $z \mapsto a_n z^n$ is always uniformly continous on a compact set, so we

can conclude that f is well-defined and continous on domain convergence when we take $r \to R$. We will take about the derivable of the power series, it is a important property.

Theorem 2.2 (Derivable)
The power series $f(z) = \sum_{n\in\mathbb{N}} a_n z^n$ defined on the domain $D(0, R)$ with raduis $R > 0$, then the derivative of f is also a power sereies with the same raduis of convergence and has the form

$$f'(z) = \sum_{n\in\mathbb{N}} n a_n z^{n-1}$$

Proof. ∎

An immediate result is that power series is a holomorphic function, and another result is an inversion of the process, that means if we know a function definitely is converged by some power series in some domain, then we can write each coefficient in terms of the derivative of the function, usually in calculus we call it the Taylor's Formula.

Corollary 2.2.1 (Taylor's series)
If $f(z)$ is the sum of the power series $\sum_{n\in\mathbb{N}} a_n z^n$ on the domaine $D(0, R)$ with $R > 0$,

then we have

$$a_n = \frac{f^{(n)}(0)}{n!}.$$

Proof. Manipulate above theorem to f k times we get

$$f^{(k)}(x) = \sum_{n \in \mathbb{N}} n(n-1) \cdots (n-k+1) a_n z^{n-k}$$

and we put $z = 0$, then we get $n! a_n = f^{(k)}(0)$, that is what we want. ∎

we can also conclude the **analytic continuation** about the power series, although it can be got direct from holomorphic function, but the proof here is easy and it gives us the connection.

Corollary 2.2.2 *Suppose that $\sum_{n \in \mathbb{N}} a_n z^n$ and $\sum_{n \in \mathbb{N}} b_n z^n$ are two power series with the positive raduis R_1 and R_2, if there exists an voisinage of zero V such that $\sum_{n \in \mathbb{N}} a_n z^n = \sum_{n \in \mathbb{N}} b_n z^n$, then $a_n = b_n$ for any $n \in \mathbb{N}$.*

Proof. We Let $f(z) = \sum_{n \in \mathbb{N}} a_n z^n$ and $g(z) = \sum_{n \in \mathbb{N}} b_n z^n$ on domaine, then we have

$$0 = (f - g)(z) = \sum_{n \in \mathbb{N}} (a_n - b_n)^n z^n$$

on the subset of the domaine V, so we applying above theorem we can get that

$$a_n - b_n = n!(f - g)(0) = 0$$

Which is the result we get, ∎

After some basic description of power series, we can define the **analytic function**,

Definition 2.1 *Let $\mathbb{K} = \mathbb{R}$ or $\mathbb{C}$, U is a open set in $\mathbb{K}$, a function $f : U \to \mathbb{K}$ is an analytic function if for any $z_0 \in U$, the function can be converged by a power series in a small voisinage of z_0.*

formally speaking, for any $z_0 \in U$, there exists $r > 0$ and a sequence $(a_n)_{n \in \mathbb{N}}$ such that in the domaine $D(z_0, r) \subset U$ we have

$$f(z) = \sum_{n \in \mathbb{N}} a_n (z - z_0)^n$$

Remark a power series defined on a domain is clearly an analytic function, the choice of r can be formulated by $\min\{|z_0|, R - |z_0|\}$, where R is the raduis of convergence.

The classic discussion about the analytic function is that whether analytic is a smooth function(C^∞), and conversely a smooth function is a analytic function? In

the time of Abel, analysis is focused on $\mathbb{R}$, and a smooth function is usually considered to do expansion by Talyor's Formula, so the question in $\mathbb{R}$ is equivalent to the question that for each smooth function whether its Taylor sereies converges to itself or not. In conclusion, the relation show as folloing in $\mathbb{R}$

$$\text{analytic function} \subset C^{\infty} \text{smooth function}$$

Proposition 2.3 (infinitely differentiable)

For any analytic function defined on an open set of $\mathbb{K}$, it is infinitely differentiable and its derivative is still an analytic function.

Proof. We let $f : U \to \mathbb{K}$, and take $a \in U$, the analytic function implies a positive $r > 0$ and a seuqnce (a_n), such that $f(z) = \sum a_n (z-a)^n$ for any $z \in D(a, r)$, then applying theorem 2.2 we can conclude that

$$f^{(k)} = k! a_k$$

for any $k \in \mathbb{N}$, hence it is infinitely differentiable. What's more, for any $z \in D(a, r)$, $f'(z) = \sum n a_n (z-a)^{n-1}$ by the same theorem, which gurantee that the derivative is still an analytic function. ∎

Now we will give an example that a real-value function is smooth but not analytic to finish our conclusion.

Example 2.1 *we consider the following real function*

$$f : \mathbb{R} \to \mathbb{R}, \quad x \mapsto \begin{cases} e^{-1/x^2} & x \neq 0 \\ 0 & x = 0 \end{cases}$$

Notice that $x \mapsto e^{-1/x^2}$ is a C^∞ function for any $x \neq 0$, and we can prove by recurrence that its n-times derivative has the form

$$f^{(n)}(x) = \frac{P_n(x)}{x^{3n}} e^{-1/x^2}$$

where $P_n \in \mathbb{R}[x]$. Then calculate the limite by changing $u = 1/x$, we know

$$\lim_{x \to 0} \frac{e^{-1/x^2}}{x^{3n}} = \lim_{u \to \infty} \frac{u^{3n}}{e^{u^2}} = 0$$

so for any $n \in \mathbb{N}$, $f^{(n)}(0)$ tends to zero, so we expend the function by adding $f^{(n)}(0) = 0$ such that $f \in C^\infty(\mathbb{R})$.

Suppose that f is analytics at $x = 0$, then we can write the Talyor's series at the voinsinage of zero

$$f(x) = \sum_{n \in \mathbb{N}} \frac{f^{(n)}(0)}{n!} x^n = 0$$

which means the function equals to null in the voisinage, clearly absurd since the series is not convergent to itself. Hence, f can not be analytic at zero although it is a smooth function.

Another work of Abel is about the convergence on the boundary of domaine of convergence, we give it here the version in $\mathbb{R}$.

Theorem 2.4 (Abel's Theorem)
Let $f(z) = \sum_{n \in \mathbb{N}} a_n z^n$ be a power sereies defined on a domaine $(-R, R)$ and $R > 0$. If the power series converges at $z = R$(or $z = -R$), then the f is continous (one-sid) at $z = R$(or $z = -R$). Or simply, the theorem is equivalent to porve that

$$\lim_{r \to 1^-} \sum_{n \in \mathbb{N}} a_n r^n = \sum_{n \in \mathbb{N}} a_n$$

if $\sum_{n \in \mathbb{N}} a_n$ is a covergent series.

Proof. the rigorous proof needs to use the **summation by parts formula**:

$$\sum_{n=M}^{N} a_n b_n = a_N B_N - a_M B_{M-1} - \sum_{n=M}^{N-1} (a_{n+1} - a_n) B_n$$

where a_n and b_n are all finite, and $B_n = \sum_{k=1}^{n} b_k$ with convention $B_0 = 0$. ∎

Finally i will talk about holomorphic function. Loosely speaking, homomorphic function is just the differnetiable functions in $\mathbb{C}$, why do we not just call them "a differentaible function" and give them a special name? The important reason is the crazy regularity of the complex-valued function. By **Cauchy's integral formula**, for any open set containing an open disc D, if a function f is holomorphic in the open set, then for any $z \in D$

$$f(z) = \frac{1}{2\pi i} \int_{\partial D} \frac{f(\zeta)}{\zeta - z} d\zeta$$

then by caculation

$$\frac{f(z+h) - f(z)}{h} = \frac{1}{2\pi i} \int_{\partial D} \frac{f(\zeta)}{(\zeta - z)(\zeta - z - h)} d\zeta$$

By the continuity of the parametric integration at $h = 0$, we can conclude that f has derivatice and $f(z) = \frac{1}{2\pi i} \int_{\partial D} \frac{f(\zeta)}{(\zeta - z)^2} d\zeta$, hence we can prove by recurrence that f is infinitely times differnetiable and with the form

$$f^{(n)}(z) = \frac{n!}{2\pi i} \int_{\partial D} \frac{f(\zeta)}{(\zeta - z)^{n+1}} d\zeta \tag{2}$$

That is an amazing result and also a fundamental theorem for complex analysis, so clearly, any holomorphic function is a C^∞ function, also the most important points is that the high-order derivative only depends on the original function. We recall the

theorem 2.2, then we can complete the connection between analytic function and a smooth function in $\mathbb{C}$:

$$\text{analytic function} \iff \text{holomorphic function}$$

Theorem 2.5 (Stein 4.4)

If f is holomorphic function in an open set $\Omega \subset \mathbb{C}$, then f is an analytic function in Ω.

Proof. By definition, we just need to verify that in any $z_0 \in \Omega$, there exists a voisinage such that the Talyor's sereies $\sum_{n \in \mathbb{N}} \frac{f^{(n)}(z_0)}{n!}(z - z_0)^n$ converges to f.

Too see that we notice in Cauchy integration formula

$$\frac{1}{\zeta - z} = \frac{1}{\zeta - z_0} \frac{1}{1 - u}$$

by taking $u = \frac{z - z_0}{\zeta - z_0}$, then for any $|u| < 1$, we can rewrite formula as the form

$$f(z) = \frac{1}{2\pi i} \int_{\partial D} \frac{f(\zeta)}{\zeta - z_0} \sum_{n \in \mathbb{N}} u^n d\zeta = \sum_{n \in \mathbb{N}} \frac{(z - z_0)^n}{2\pi i} \int_{\partial D} \frac{f(\zeta)}{(\zeta - z_0)^{n+1}} d\zeta$$

Here interchange $\int / \sum$ can be done since the holomorphic function is continous. and take equation 2 into above equality then we can get the result we hope. ∎

Remark That means for any holomorphic function, we can do expanasion casually.

2.2 Limit superior and limit inferior

A direct question from **Bolzano-Weierstrass theorem** is that how to determiner the range of the vlaue which the convergent subsequence converges to? Gnenerally, the sequence can be classified into three types: convergent, divergent and oscillation.

We take a real sequence $(a_n)_{n \in \mathbb{N}}$ and define a set

$$A_k = \{a_{k+p} | p \in \mathbb{N}\}$$

it means a set which cantains the terms of sequence except $a_0, ..., a_{k-1}$, then we define two new sequence by

$$v_n = \sup A_n, \quad w_n = \inf A_n$$

Notice that $A_{k+1} \subset A_k$ for any k, so clearly we have

$$w_n \leq w_{n+1} \leq a_{n+1} \leq v_{n+1} \leq v_n$$

for any n. Here w_n is an increasing sequence and v_n is a decreasing sequence, **so highly likely they will be convergent.** If $(a_n)_{n \in \mathbb{N}}$ is a convergent sequence, clearly as $n \to \infty$, v_n and w_n will converge to the limit of a_n, but if $(a_n)_{n \in \mathbb{N}}$ is oscillated, the convergence will change. See the example below:

Example 2.2 *We define a seuqnce as following*

$$a_n = \begin{cases} \frac{1}{n} & n \equiv 0 \mod 3 \\ \frac{1}{n} + 1 & n \equiv 1 \mod 3 \\ \frac{1}{n} + 2 & n \equiv 2 \mod 3 \end{cases}$$

we calculate the first 6 terms of v_n and w_n

$$w_1 = w_2 = ... = w_6 = 0$$
$$v_1 = v_2 = 3/2, v_3 = v_4 = v_5 = 1/5 + 2, v_6 = 1/7 + 2$$

so we can conclude with a little work that w_n converges to 0, and v_n converges to 2.

In above example, we easily guess that all possible values the subsequence will converge to are $0, 1, 2$, so w_n and v_n seem to give the superimum and infimum respectively, hence we formally define the limit superior and limit inferior of a sequence $(a_n)_{n \in \mathbb{N}}$ by

$$\limsup a_n := \inf_{k \geq 0} (\sup A_k)$$

$$\liminf a_n := \sup_{k \geq 0} (\inf A_k)$$

we conclude the properties as folloing

Proposition 2.6 *For a general real number $\mathbb{R} \cup \{\pm\infty\}$, we define $(a_n)_{n\in\mathbb{N}}$ is a squence of $\mathbb{R}$, then for any subsequence $(a_{n_k})_{k\in\mathbb{N}}$ which converges to L, then*

$$\liminf a_n \leq L \leq \limsup a_n.$$

In another words, all adherence points of the sequence will be in $[\liminf a_n, \limsup a_n]$.

Proof. ∎

Here we give an application in probability theory about sequence of the events.

Theorem 2.7 (Borel-Cantelli Lemma) *Suppose $(A_i)_{i\in\mathbb{N}}$ is a sequence of events in a probability space $(\Omega, \mathcal{T}, P)$, then*
(a) If $\sum_{n\in\mathbb{N}} P(A_i) < +\infty$, then the probability that infinitely many of them occur is zero.
(b) If $\sum_{n\in\mathbb{N}} P(A_i) = +\infty$ and the events are independent, then the probability that infinitely mant of them occur is 1.

we need to translate the language into the meusure theory. here "the probability that infinitely many of them occur" can be equivalent to the statement below: almost

all $w \in \Omega$ will be in infinitely many of the events $\{A_i\}$. we use the limit superior and inferior to formalize the statement

$$\limsup A_i = \cap_{N \in \mathbb{N}} \cup_{i \geq N} A_i$$

here we fistly prove that

$$\limsup A_i = \{w \in \Omega | w \text{ is in infinitely many of events}\}$$

for any $w \in \limsup A_i$, $w \in \cup_{i \geq 0} A_i$, so there exists $x_0 \in \mathbb{N}$ such that $w \in A_{x_0}$. Then we consider $w \in \cup_{i \geq x_0 + 1} A_i$, so there exists $x_1 \geq x_0 + 1$ such that $w \in A_{x_1},...,$ by recurence we can conclude a strictly incrasing map of integer $n \mapsto x_n$, then we can conclude that $w \in \cap_{n \in \mathbb{N}} A_{x_n}$.

For another hand, if w in in infinitely many of evnets, then we label all the events where it occurs, and it will be countable, so there exists a correspondence with $\mathbb{N}$ and we denote it by f, then $w \in \cap_{n \in \mathbb{N}} A_{f(n)} \subset \limsup A_i$.

So we can conclude Borel-Cantelli Lemma by the meusure theory as following:
(a) $\sum_{n \in \mathbb{N}} P(A_i) < +\infty \implies P(\limsup A_i) = 0$.
(b) $\sum_{n \in \mathbb{N}} P(A_i) = +\infty \wedge \text{independence} \implies P(\limsup A_i) = 1$.

So finally we finish the proof of theorem:

Proof. we notice that

$$w \mapsto \#\{i \in \mathbb{N} | w \in A_i\}$$

is a mesurable function by discrete mesure, and then we notice that

$$\int_\Omega \#\{i \in \mathbb{N} | w \in A_i\} dP = \int_\Omega \sum_{n \in \mathbb{N}} \mathbb{1}_{A_i} dP$$
$$= \sum_{n \in \mathbb{N}} P(A_i)$$

so if $\sum_{n \in \mathbb{N}} P(A_i) < +\infty$, then $\#\{i \in \mathbb{N} | w \in A_i\}$ must be almost finite, i.e. almost all $w \in \Omega$ occur in finite many of events, so $\limsup A_i = 0$.

If $\sum_{n \in \mathbb{N}} P(A_i) = +\infty$, by independence we have

$$P(\cup_{i=N}^M A_i^c) = \prod_{i=N}^M 1 - P(A_i^c) \leq e^{-\sum_{i=N}^M P(A_i)}$$

as $M \to \infty$ we can know that $P(\cap_{i \geq N} A_i^c) = 0$ for any N, then by De Morgan's Law

$$P(\limsup A_i) = P((\cup_{N \in \mathbb{N}} \cap_{i \geq N} A_i^c)^c) = 1 - 0 = 1$$

∎

2.3 Theorem of transfert: algebra of the random variable

We firstly talk about the proof of the theorem of transfert, and then apply it into The probability to get the general form of experience, and finally see some examples of the caculation of random variable via the theorem.

Notice that in this section $\mathbb{R}^+ = [0, +\infty]$ for genenralisation.

Theorem 2.8 (Transfert) *Suppose that $(X, \mathcal{E}, \mu)$ is a measure space and $(Y, \mathcal{T})$ is a measurbale space. Given a measurbale function $f : X \to Y$, then*
(1) The function invites a natural measure $v : \mathcal{T} \to \mathbb{R}^+$ by

$$v(B) = \mu(f^{-1}(B)) = \mu \circ f^{-1}(B)$$

we call it the image measure of f such that $(Y, \mathcal{T}, v)$ is a measure space.
*(2) For any $v-$integrable **(or postive measurbale)** function $h : Y \to \mathbb{C}$, we have*

$$\int_Y h dv = \int_X h \circ f d\mu$$

Proof. (1) Just verify. (2) By «technic» lemma, for any positive measurbale function

h there exists $B_n \in \mathcal{T}$ and $c_n \in \mathbb{R}^+$ such that $h = \sum_{n \in \mathbb{N}} c_n 1_{B_n}$, so

$$
\begin{aligned}
\int_Y h\, dv &= \sum_{n \in \mathbb{N}} c_n \int_Y 1_{B_n}\, dv \\
&= \sum_{n \in \mathbb{N}} c_n v(B_n) \\
&= \sum_{n \in \mathbb{N}} c_n \mu(f^{-1}(B_n)) \\
&= \sum_{n \in \mathbb{N}} c_n \int_X 1_{f^{-1}(B_n)}\, d\mu \\
&= \sum_{n \in \mathbb{N}} c_n \int_X 1_{B_n} \circ f\, d\mu \\
&= \int_X \left(\sum_{n \in \mathbb{N}} c_n 1_{B_n}\right) \circ f\, d\mu = \int_X h \circ f\, d\mu
\end{aligned}
$$

Then some easy work to generalize any measurbale function... ∎

With the basic theorem we can translate the theorem into the probability, we notice that the image measure is actually the probability distribution of a random variable, hence $Y = \mathbb{R}$ and $\mathcal{T} = \mathcal{B}(\mathbb{R})$. And the intergation of a measure function under a probability measure is just the expectation.

Let us see the example we

Theorem 2.9 (Transfert in probability) *Given a probability space $(\Omega, \mathcal{T}, P)$ and P_X is a probability distribution of a random variable X, then for any postive borel function g, we have*

$$E[g(X)] = \int_\Omega g(X)dP = \int_\mathbb{R} gdP_X$$

Furthermore if X is continous with density $f_X : \mathbb{R} \to \mathbb{R}$, then

$$E[g(X)] = \int_\mathbb{R} g \cdot f_X dl$$

where l is the lebesgue measure.

Corollary 2.9.1 *For a continous random variable X with density f_X, the expactation*

$$E[X] = \int_\mathbb{R} t f_x(t) dt$$

Remark For a abstract probability space, it contains events which can not directly be described by number, and random variable gives a correspondence between the probability of the events and value in $\mathbb{R}$, that is important. So theorem of transfert just let us can opreate the expectation of a random variable directly from a integration of a real-valued function, so it gives a correspondence in the forms of integration.

Here is some appliaction of theorem of transfert:

Proposition 2.10 (convolution) *Suppose that X, Y are two independent random variable, then the distribution of the random variable $X + Y$ is the convolution of P_X and P_Y:*

$$P_{X+Y}(A) = \int_{\mathbb{R}} P_X(A - y) dP_Y(y)$$

If f_X and f_Y are their density functions, then the density of $X + Y$ is also the convolution of f_X and f_Y:

$$f_{X+Y}(t) = \int_{\mathbb{R}} f_X(t - y) f_Y(y) dy$$

Proof. For any $A \in \mathcal{B}(\mathbb{R})$, we have

$$
\begin{aligned}
(P_X * P_Y)(A) &= \int_{\mathbb{R}} P_X(A - y) dP_Y(y) \\
&= \int_{\mathbb{R}} (\int_{\mathbb{R}} 1_{A-y}(x) dP_X(x)) dP_Y(y) \\
&= \int_{\mathbb{R}} (\int_{\mathbb{R}} 1_A(x + y) dP_X(x)) dP_Y(y) \\
&= \int_{\mathbb{R}^2} 1_A(x + y) dP_{(X,Y)}(x, y)
\end{aligned}
$$

Notice that if we take $g(x, y) = x + y$, then clearly $1_A(x + y) = 1_A \circ g(x, y)$, so $(P_X * P_Y)(A) = E[1_A(X + Y)] = P(X + Y \in A) = P_{X+Y}(A)$. For the density, we take a borel function h, then

$$
\begin{aligned}
E[h(X + Y)] &= \int_{\mathbb{R}} h \, dP_{X+Y} \\
&= \int_{\mathbb{R}} h \, d(P_X * P_Y) \\
&= \int_{\mathbb{R}^2} h(x + y) f_X(x) f_Y(y) dx dy \\
&= \int_{\mathbb{R}^2} h(z) f_X(z - y) f_Y(y) dy dz \\
&= \int_{\mathbb{R}} h(z) (\int_{\mathbb{R}} f_X(z - y) f_Y(y) dy) dz
\end{aligned}
$$

Hence we can conclude our result by the theorem of transfert. ∎

3 Complex analysis

3.1 Holomorphic Function

Complex-valued function is a function defined on $\mathbb{C}$, formally it is

$$f : \mathbb{C} \to \mathbb{C}, \quad z \mapsto f(z)$$

$\mathbb{C}$ can be seen same as $\mathbb{R}^2$ in view of $\mathbb{R}-$vector space, so there exists a correspondence with two real variable functions.

For any z, there exists $u, v \in \mathbb{R}$ such that $f(z) = u + iv$, then

$$u = \frac{1}{2}(f(z) + \overline{f(z)}), \quad v = \frac{1}{2i}(f(z) - \overline{f(z)})$$

so by taking all z, we can get two functions from $\mathbb{C}$ to $\mathbb{R}$, and we just identify $z \in \mathbb{R}$ with $(x, y) \in \mathbb{R}^2$ by $z = x + iy$, then

$$f(z) = f(x + iy) = f(x, y) = u(z) + iv(z) = (u(x, y), v(x, y))$$

Here we should notice that $\mathbb{R}^2$ equips the norm $\| \cdot \|_2$ to make it compatiable.

Notice that $\mathbb{C}$ natually has the norm structure, so it makes sense to define its differntiation directly from the topology:

Definition 3.1 *Let Ω be an open set of $\mathbb{C}$, a function $f : \Omega \to \mathbb{C}$ is called holomorphic at $z_0 \in \Omega$ if the limit*

$$\lim_{h \to 0} \frac{f(z_0 + h) - f(z_0)}{h}$$

exists. We use $H(\Omega)$ to denote the set of all holomorphic functions on Ω.

Example 3.1 *An important example that a complex-valued function is not holomorphic is $z \mapsto \overline{z}$. We observe the function at $z = 0$*

$$\frac{\overline{z + h} - \overline{z}}{h} = \frac{\overline{h}}{h} = \frac{re^{-a}}{re^{a}} = e^{-2a}$$

so the limit does not exist. By if we rewrite the function in real variables by $(x, y) \mapsto (x, -y)$, then clearly the jacobi matrix is $\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$, which means the function is always differentiable in $\mathbb{R}^2$, so here difference appears and we can realize that the differentiation of complex-value function is a little different from.

Although the two types of differentiation is not the same thing, is there some connection between them? Let us firstly condiser a set of differential opreators:

$$\begin{cases} x = \frac{z + \overline{z}}{2} \\ y = \frac{z - \overline{z}}{2i} \end{cases} \implies \begin{cases} \frac{\partial}{\partial z} = \frac{\partial}{\partial x}\frac{\partial x}{\partial z} + \frac{\partial}{\partial y}\frac{\partial y}{\partial z} = \frac{1}{2}(\frac{\partial}{\partial x} - i\frac{\partial}{\partial y}) \\ \frac{\partial}{\partial \overline{z}} = \frac{\partial}{\partial x}\frac{\partial x}{\partial \overline{z}} + \frac{\partial}{\partial y}\frac{\partial y}{\partial \overline{z}} = \frac{1}{2}(\frac{\partial}{\partial x} + i\frac{\partial}{\partial y}) \end{cases}$$

Suppose f is a complex-value function holomorphic at 0 with $f(0) = 0$, then

$$f(z) = L(z) + o(z) = ax + by + o(z) = \frac{a - ib}{2}z + \frac{a + ib}{2}\overline{z} + o(z)$$

with L a $\mathbb{C}-$linear map with $a, b \in \mathbb{C}$ and $z = x + iy$, then by above opreators we know that

$$\frac{f(z)}{z} = \frac{\partial f}{\partial z}(0) + \frac{\partial f}{\partial \overline{z}}(0) \cdot \frac{\overline{z}}{z} + \frac{o(z)}{z}$$

if the limit exists as z tends to zero, them the term having $\frac{\overline{z}}{z}$ must vanish, that means $\frac{\partial f}{\partial \overline{z}} = 0$, that is the necessary condition that f is holomorphic at $z = 0$.

Furthermore, if we merely view $\mathbb{C}$ as $\mathbb{R}^2$ to write function as the form of linear combination by two real-value functions $f = u + iv$, then we can conclude the Jacobien matrix

$$J_f(x, y) = \begin{pmatrix} \frac{\partial u}{\partial x} & \frac{\partial u}{\partial y} \\ \frac{\partial v}{\partial x} & \frac{\partial v}{\partial y} \end{pmatrix}$$

Then $L(z)$ can be viewed as vector (x, y) opreates on $J_f(x, y)$, and notice that $L(z) = k \cdot z$ form some $k \in \mathbb{C}$ and $\mathbb{C}$ is isomorphic to a subspace of $M_2(\mathbb{R})$ with the form $\begin{pmatrix} a & -b \\ b & a \end{pmatrix}$, so immediately the jacobi matrix has the same form and we get a relation

: **Cauchy-Riemann Equation:**

$$\begin{cases} \frac{\partial u}{\partial x} = \frac{\partial v}{\partial y} \\[2mm] \frac{\partial u}{\partial y} = -\frac{\partial v}{\partial x} \end{cases}$$

Let us reflect the procedure above, the different things is that L must be a $\mathbb{C}-$linear map instead of $\mathbb{R}-$linear map, which is crucial.

Lemma 3.1 *Let $V = Lin(\mathbb{R}^2, \mathbb{C})$ be the set of all linear maps from $\mathbb{R}^2$ to $\mathbb{C}$, then*
(1) V is a $\mathbb{R}-$vector space with $dx : (x, y) \mapsto x$ and $dy : (x, y) \mapsto y$ as the base.
(2) $dz = dx + idy$ and $d\bar{z} = dx - idy$ compose another base of V
(3) V is a $\mathbb{C}-$vector space if and only if for any $L \in V$ under the base dz and $d\bar{z}$, the coefficient of $d\bar{z}$ is zero.

Proof. (1) For any $L \in V$, we have $L(x, y) = x \cdot L(1, 0) + y \cdot L(0, 1)$ notice that dx and dy actually are in V, and $dx(x, y) = x$, $dy(x, y) = y$, then $L = L(1, 0)dx + L(0, 1)dy$ with $L(1, 0), L(0, 1) \in \mathbb{C}$.

(2) we take $c_1 dz + c_2 d\bar{z} = 0$, then $(c_1 + c_2)dx + i(c_1 - c_2)dy = 0$, which implies $c_1 = c_2 = 0$, so d_z and $d\bar{z}$ are linearly independent.

(3) Take $f = adz + bd\overline{z} \in V$ and $u \in \mathbb{C}^2$

$$f(iu) - if(u) = adz(iu) + bd\overline{z}(iu) - iadz(u) - ibd\overline{z}(u)$$
$$= -2bu_y$$

The result must be zero since the map in linear over $\mathbb{C}$, then $b = 0$. ∎

Let us consider differential

$$df : U \to Lin(\mathbb{R}^2, \mathbb{C})$$

notice that in differnetial $L(1,0) = Df(a,b)(1,0) = f_x(a,b)$, so we have that

$$df = f_x dx + f_y dy$$

Nowe if we expect to write df as the combination of dz and $d\overline{z}$, so

$$df = \mu dz + \lambda d\overline{z} = (\mu + \lambda)dx + i(\mu - \lambda)dy$$

with caculation we have that

$$\mu = \frac{f_x - if_y}{2}, \quad \lambda = \frac{f_x + if_y}{2}$$

Which implies the differential opreators like above.

So we can conclude the fundamental theorem of holomorphic function.

Theorem 3.2 *Suppose that $f = u + iv$ is a complex function with u, v two functions of class $C^1(\mathbb{R}^2, \mathbb{R})$, then following statements are equivalent.*
(1) f is holomorphic on Ω.
(2) $\frac{\partial f}{\partial \bar{z}} = 0$ on Ω.
(3) Cauchy-Riemann Equation is satisfied.

Proof. (2) f is holomorphic, then for any $z_0 \in \Omega$, $df(z_0)$ must be a $\mathbb{C}-$ linear map in $Lin(\mathbb{R}^2, \mathbb{C})$, so the coefficient of $d\bar{z}$ must be zero by above lemma, and notice that the coefficient is just $\frac{\partial f}{\partial \bar{z}}(z_0)$.

(3) Just calculate

$$
\begin{aligned}
0 = \frac{\partial f}{\partial \bar{z}} &= \frac{1}{2}\left(\frac{\partial f}{\partial x} + i\frac{\partial f}{\partial y}\right) \\
&= \frac{1}{2}\left(\frac{\partial u + iv}{\partial x} + i\frac{\partial u + iv}{\partial y}\right) \\
&= \frac{1}{2}\left[\left(\frac{\partial u}{\partial x} - \frac{\partial v}{\partial y}\right) + i\left(\frac{\partial u}{\partial y} + \frac{\partial v}{\partial x}\right)\right]
\end{aligned}
$$

∎

4 probaility

4.1 Independence

Definition 4.1 *Given a probability space* $(\Omega, \mathcal{T}, P)$

- *Two events* $A, B \in \mathcal{T}$ *are called independent if*

$$P(A \cap B) = P(A)P(B)$$

- *Two random variable* X, Y *are called independent if for any* $A, B \in \mathcal{B}(\mathbb{R})$

$$P(X \in A; Y \in B) = P(X \in A)P(Y \in B)$$

Remark Independence is a simple concept in probability but not in mathematics. Notice that in measure product of Borel set, we have $\mathcal{B}(\mathbb{R}^2) = \mathcal{B}(\mathbb{R}) \otimes \mathcal{B}(\mathbb{R})$, and equivalently we can rewrite the definition of the independence of two random variable by

$$P((X, Y) \in A \times B) = P(X \in A)P(Y \in B)$$

that means it invites a new probability measure with the properties

$$P_{(X,Y)}(A \times B) = P_X(A)P_Y(B)$$

where (X, Y) is a natural random variable defined on Ω^2 by $(a, b) \mapsto (X(a), X(b))$. Clearly we can deduced that independence is the trival condition that measure product is well-defined in probability space.

Theorem 4.1 *Two random variable X and Y are independent if and only if $P_{(X,Y)} = P_X \otimes P_Y$, in another word, they define the same measure in space $(\mathbb{R}^2, \mathcal{B}(\mathbb{R}^2))$.*

Proof. Firstly, P_X and P_Y are two natural probability measure on $\mathbb{R}$, so from the construction of the measure product, $P_X \otimes P_Y$ exists. So it is a measure on space $(\mathbb{R}^2, \mathcal{B}(\mathbb{R}^2))$. We put $S = \{A \times B | A, B \in \mathcal{B}(\mathbb{R})\}$. it is the pi-system on $\mathbb{R}^2$. We notice that on S, $P_{(X,Y)} = P_X \otimes P_Y$ if and only if the two random variables are independent, so by the **theorem of unique exitension of the measure** we can know that they are actually same in $\mathbb{R}^2$. ∎

Several result can be concluded from the theorem, it refers to some probability value with respect two independence variable.

Corollary 4.1.1 *Suppose that X, Y are two independent random variable in $(\Omega, \mathcal{T}, P)$, then*
(1) For any two borel functions $f, g : \mathbb{R} \to \mathbb{R}$ such that $f(X), g(Y)$ are $P-$integrable,

$$E[f(X)g(Y)] = E[f(X)]E[g(Y)]$$

(2) If X and Y are continous with density f_X and f_Y, then random variable (X,Y) admets a density $f_X \cdot f_Y$.

(3) $Cov(X,Y) = 0$

(4) $Var(X + Y) = Var(X) + Var(Y)$

Proof. (1) Using Fubini's theorem

$$E[f(X)g(Y)] = \int_{\mathbb{R}^2} f(a)g(b)dP_X(a)dP_Y(b)$$
$$= (\int_{\mathbb{R}} fdP_X) \cdot (\int_{\mathbb{R}} gdP_Y)$$
$$= E[f(X)]E[g(Y)]$$

(2)For any $A, B \in \mathcal{B}(\mathbb{R})$

$$P_{(X,Y)}(A \times B) = P_X(A)P_Y(B)$$
$$= (\int_{\mathbb{R}} f_X(t) \cdot 1_A(t)dt) \cdot (\int_{\mathbb{R}} f_Y(u) \cdot 1_B(u)du)$$
$$= \int_{\mathbb{R}^2} 1_{A \times B}(t, u)(f_X(t)f_Y(u))dl(t, u)$$

so clearly $(t, u) \mapsto f_X(t) f_Y(u)$ is the density of (X, Y) since the class $A \times B$ generates $\mathcal{B}(\mathbb{R}^2)$.

(3) and (4) are immediate afte caculation. ∎

4.2 Convergence

4.3 Law of large number

Definition 4.2 *A sequence of random variable $(X_n)_{n \in \mathbb{N}}$ is called identically distributed if they share the same distribution, that means for any $i, j \in \mathbb{N}$:*

$$P_{X_i} = P_{X_j}$$

Usually we use "id." to denote the random variable.

There are a criteria to determine if the random variables are identically distributed after changing variable:

Proposition 4.2 *If X_1 and X_2 are two identcially distributed random variable, and $f : \mathbb{R} \to \mathbb{R}$ is a positive borel function(or integrable), then $f(X_1)$ and $f(X_2)$ are still identically distributed.*

Proof. we notice that by the theorem of transfert

$$E[f(X_i)] = \int_{\mathbb{R}} f dP_{X_i}$$

idectical distribution implies $E[f(X_1)] = E[f(X_2)]$, then we take $f = 1_A$ with A a borel se, immediate we get that $P_{X_1}(A) = P_{X_2}(A)$ for any borel set A, so they are equal. ∎

Now we can conclude the important theorem in probability theory:

Theorem 4.3 (Strong Law)

Suppose that $(X_i)_{i \in \mathbb{N}}$ is a sequence of independent, idectically distributed and integrable random variables, then for almost every $w \in \Omega$

$$\frac{S_n(w)}{n} = \frac{(X_1 + ... + Xn)(w)}{n} \xrightarrow[n \to \infty]{} E[X_0]$$

Or we can say the sample mean $\frac{S_n}{n}$ converges to zero almost surely.

Corollary 4.3.1 *For any borel set A, we have a almost surely convergence*

$$\frac{\#\{i \le n | X_i(w) \in A\}}{n} \xrightarrow[n \to \infty]{} P(X_0 \in A)$$

Theorem 4.4 (Weak Law)

Suppose that $(X_i)_{i \in \mathbb{N}}$ is a sequence of independent, idectically distributed and square integrable random variables, then for any $\epsilon > 0$,

$$P(|\frac{S_n}{n} - E(X_0)| > \epsilon) \xrightarrow[n \to \infty]{} 0$$

Or we can say the sample mean $\frac{S_n}{n}$ converges t0 $E[X_0]$ by probability.

Proof. The proof is immediate via the strong law by convergence. squeare integrable implies integrable by Cauchy-Schwartz inequalit, then the seuqence converges almost surely to $E[X_0]$, clearly convegent by probability. ∎