Math remark
fondation for analysis and probability

X

ElegantLaTeX Program

Update: January 1, 2026

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 $\text{im} f \subset I$, then $\int_X f d\mu \in I$ with

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

The equality holds if f is constant almost everywhere.

Proof. Let $a = \inf I \in \mathbb{R} \cup \{-\infty\}$ and $b = \sup I \in \mathbb{R} \cup \{+\infty\}$, then

$$a = \int_X a d\mu \leq \int_X f d\mu \leq \int_X b d\mu = b$$

if $\int_X f d\mu = a$ or b, then $\int_X (f-a) d\mu$ or $\int_X (f-b) d\mu$ is zero, it implies that $f = a$ or $f = b$ almost everywhere, so a or b is in I, which means $\int_X f d\mu \in I$, so the left of the inequality is well-defined.

By the convexity of the function φ, there exists $m \in \mathbb{R}$ such that for any $x \in I$

$$\varphi(x) \geq \varphi(y) + m(x - y)$$

with y an interior point of I, then apply it to the integral by taking $y = \int_X f d\mu$

$$\int_X \varphi \circ f d\mu \geq \int_X \varphi(y) + m(f(t) - y) d\mu(t)$$
$$= \varphi(y) = \varphi\left(\int_X f d\mu\right)$$

and if y is not interoir point i.e. f is a constant almost evry where, then we can take equality. ∎

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

- **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 domaine 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 $u = \dfrac{|f|}{\|f\|_p}$ and $v = \dfrac{|g|}{\|g\|_q}$, 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 $uv \leq \dfrac{1}{p}u^p + \dfrac{1}{q}v^q$, hence we can conclude

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

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 $|f| = \dfrac{\|f\|_p}{\|g\|_q}|g|$, so we finish the proof. ∎

Corollary 1.2.1 (Cauchy-Schwartz)

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

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

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 $Z = \dfrac{Y - E[Y]}{\sigma(Y)}$, then by above inequality

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

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

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 domaine 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 $\sum\limits_{n \in \mathbb{N}} a_n z^n$ converge.

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

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

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

so clearly the power series converge if $|z| \leq |z_0|$ 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. $\blacksquare$

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 $\sum\limits_{n \in \mathbb{N}} a_n z^n$ 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 domaine 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 domaine* $D(0, R)$ *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 domaine, 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. $\blacksquare$

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 $u = \dfrac{z - z_0}{\zeta - z_0}$, 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 expansion casually.

2.2 Anayltic continuation

Notice the corollary 2.2.2 we conclude in the previous section, it is an important properties, it claims that the local value of the power series determines the global vaule. In general, two analytic function f and g, if they are equal in a local voinsinage V, then for any $x \in V$, there exists a open ball $B(x, r)$ such that their power series expansions are equal. The most important things is that V is not compact, so our open cover $B(x, r)$ will bring the properties by extending, so f and g are actually equal!

The statement above is a little abstract, it is better to draw a picture from a voinsinage V, then see that if we want to let $f = g$ we just need to consider the point nearly the ∂V...

In this section we will prove the analytic continuation, and three other important theorems in complex anaylsis, although they have various methods to porve, but here our inspiration is continuation:

$$\text{Analytic continuation} \implies \text{Isolated zero} \implies \text{Open mapping} \implies \text{Maximum modulus}$$

Theorem 2.6 (analytic continuation)

*Given an **open and connect** set U in complex plane, and $f, g : U \to \mathbb{C}$ are two holomorphic (analytic) functions. If they are equal in a voinsinage of U, then $f = g$ in whole U.*

Proof. We let $S = \{z \in U | f^{(n)}(z) = g^{(n)}(z) \text{ for any} n \in \mathbb{N}\}$, and clearly $S = \bigcup_{n \in \mathbb{N}} \{z \in U | f^{(n)}(z) = g^{(n)}(z)\}$. It is easy to verify that S is the intersection of infinite closed sets, so S is closed.

S is also open. We take any $z_0 \in S$, we notice that S is a subset of the voinsinage where f and g are equal, then by corollary 2.2.2, it implies that they have the same seires in a voinsinage of z_0 contained in S, then clearly $f^{(n)}(z) = a_n = b_n = g^{(n)}(z)$ in the voinsinage, so that S must be open. What's more S is not empty, and U is connect, so $U = S$.

■

 Remark We notice that the proof is based on the topological property of the domain U, so the properties of holomorphic is speacial.

One directly result of analytic continuation is the isolated zero theorem, it claims that the zero of a holomorphic function must be discrete:

Theorem 2.7 (isolated zero) *Suppose that U is **an open and connect set** in complex plane, and $f \in H(U)$ is a non-zero holomorphic function, then*

- If $z \in U$ such that $f(z_0) = 0$, then f can be expended in a neighborhood $\Omega \subset U$:

$$f(z) = (z - z_0)^n g(z)$$

where g is a non-vanishing holomorphic function in Ω.
- Any zero z_0 of f, there exists a neighborhood such that z_0 is the unique zero in the neighborhood.
- The set of all zeros of f is discrete.

Proof. If z_0 is a zero, then there exists a expanasion of power series near the z_0 with

$$f(z) = \sum_{k \in \mathbb{N}} a_k (z - z_0)^k, \quad a_0 = 0$$

So we just take $n = \min\{i \in \mathbb{N} | a_i \neq 0\}$, then

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

So we just take $g(z) = \sum_{k \in \mathbb{N}} a_{n+k} z^k$. It is well-defined since in a small reigon near z_0 and does not contain z_0 we have surely $g(z) = \dfrac{f(z)}{(z - z_0)^n}$, then by analytic continuation g is holomorphic on a samll neighborhood with z_0 in it, and $g(z_0) = a_n \neq 0$. Hence by continuity, there exists a positive $r > 0$ such that

$$||f(z_0)| - |a_n|| \leq |f(z_0) - a_n| < |a_n|/2$$

for any $z \in B(z_0, r)$, which implies $|f(z_0)| > |a_n|/2 > 0$ in the reigon, so it is non-vanishing. Just take the intersection of the previous neighborhood and $B(z_0, r)$, then we get the Ω we hope.

The second statement is immediate from the first statement. For the third, if there exists a limit point a in the set of zeros, then for any small neighborhood with a in it, it will conatins infinite other zeros, which contradicts with the second statement. ∎

With the help of the isolated zero theorem, we can generalize the analytic continuation, locally two holomorphic function are just equal in a non-discrete set:

Theorem 2.8 (strong analytic continuation) *Given an **open and connect** set U in complex plane, and $f, g : U \to \mathbb{C}$ are two holomorphic (analytic) functions. If they are equal in a set $S \subset U$, and U admet an accumulation point, then $f = g$ in U.*

Proof. Just take $f - g$ as the new holomorphic function in U, notice it will be zero in S and S is not discrete, so $f - g$ must be zero by isolated zero. ∎

Remark An direct result if two holomorphic functions are equal in a non-constant convergent sequence, then they will behave equal globally.

The other two result is the maximum modulus principle and open mapping theorem.

Theorem 2.9 (open mapping)

*A non-constant holomorphic function defined on **an open and connect set** must be an open map.*

Proof. Take any $z_0 \in U$, and we will prove that $f(z_0) \in f(U)$ must be an interior point. By isolated zero theorem, there exists $B = B(z_0, r) \subset U$ such that in it we have

$$f(z) - f(z_0) = (z - z_0)^n g(z)$$

so f can be written as the composition of the following holomorphic functions:
- $u : z \mapsto (z - z_0)g^{1/n}(z)$ from B to $\mathbb{C}$.
- $v : z \mapsto z^n$ from $\mathbb{C}$ to $\mathbb{C}$.
- $t : z \mapsto z + f(z_0)$ from $\mathbb{C}$ to $\mathbb{C}$.

so $f = t \circ v \circ u$, we just need to verify the three maps are open, it can be proved via **the local inversion theorem: if the derivative is not zero at some point, then locally there exists a diffeomorphism, so the image of the point will be an interior.**

For u we can compute $u'(z) = g^{1/n}(z) + \dfrac{(z - z_0)}{n} g'(z) g^{\frac{1-n}{n}}(z)$, clearly $u'(z_0) = g^{1/n}(z_0) \neq 0$, and u is holomorphic so class of C^1, so locally there exists a diffeomorphism between neighborhoods U_{z_0} and $V_{u(z_0)}$, so $u(z_0)$ is a interior point of $f(B)$.

For v the proof is similar if we take $z_0 \neq 0$ since $v'(z) = nz^{n-1}$. In the case of $z_0 = 0$, we take a enough samll positive $r > 0$, then $|z| < r$ implies $|z|^n < r^n$, so $B(v(0), r) \subset B(0, r^n) = v(B(0, r))$, so v is open. Finally, t is clearly open. Thus after composition, f is open. ∎

Theorem 2.10 (maximum modulus)

*If f in a non-constant holomorphic function on **an open and connect set** U, then $|f|$ can not attain a maximum in U.*

-Weakly, if $K \subset U$ is a compact set, then $|f|_K|$ attain the maximum in the boundary:

$$\sup_{z \in K} |f(z)| \leq \sup_{\partial K} |f(z)|$$

Proof. Assume that f did attain a maximum in $z_0 \in U$, then we take a small voinsinage of x_0, called V, by open mapping theorem $f(V)$ must be open with $f(z_0)$ in it, so there exists $r > 0$ such that $B(f(z_0), r) \subset f(D)$. We take $\delta = sign(Re(f(z_0)))$, then $f(z_0) + \delta \dfrac{r}{2} \in f(D)$, we denote it $f(w_0)$, then $|f(w_0)| > |f(z_0)|$ clearly, which contradicts with the assumption.

The weakly statement is immediate since $|f|$ is continous, so $f(K)$ will be compact in $\mathbb{R}$, so $|f|$ must attain a maximum in K, with the strong statement we know that the maximum can not be in $int(K)$, so in boundary. $\blacksquare$

2.3   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*

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

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

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

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

we conclude the properties as folloing

Proposition 2.11 *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 2.12 (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 $\displaystyle\sum_{n\in\mathbb{N}} P(A_i) < +\infty$, then the probability that infinitely many of them occur is zero.
(b) If $\displaystyle\sum_{n\in\mathbb{N}} P(A_i) = +\infty$ 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) $\displaystyle\sum_{n \in \mathbb{N}} P(A_i) < +\infty \implies P(\limsup A_i) = 0$.

(b) $\displaystyle\sum_{n \in \mathbb{N}} P(A_i) = +\infty \wedge \text{independence} \implies P(\limsup A_i) = 1$.

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

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

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

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

∎

2.4 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 experence, 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 2.13 (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

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

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 2.14 (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*

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

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 2.14.1 *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 2.15 (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:

$$\begin{cases} x = \dfrac{z + \bar{z}}{2} \\ y = \dfrac{z - \bar{z}}{2i} \end{cases} \implies \begin{cases} \dfrac{\partial}{\partial z} = \dfrac{\partial}{\partial x}\dfrac{\partial x}{\partial z} + \dfrac{\partial}{\partial y}\dfrac{\partial y}{\partial z} = \dfrac{1}{2}\left(\dfrac{\partial}{\partial x} - i\dfrac{\partial}{\partial y}\right) \\ \dfrac{\partial}{\partial \bar{z}} = \dfrac{\partial}{\partial x}\dfrac{\partial x}{\partial \bar{z}} + \dfrac{\partial}{\partial y}\dfrac{\partial y}{\partial \bar{z}} = \dfrac{1}{2}\left(\dfrac{\partial}{\partial x} + i\dfrac{\partial}{\partial y}\right) \end{cases}$$

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

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

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

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

if the limit exists as z tends to zero, them the term having $\dfrac{\bar{z}}{z}$ must vanish, that means $\dfrac{\partial f}{\partial \bar{z}} = 0$, 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

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

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:**

$$\begin{cases} \dfrac{\partial u}{\partial x} = \dfrac{\partial v}{\partial y} \\ \dfrac{\partial u}{\partial y} = -\dfrac{\partial v}{\partial x} \end{cases}$$

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 $f = adz + bd\bar{z} \in V$ and $u \in \mathbb{C}^2$

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

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

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) $\dfrac{\partial f}{\partial \bar{z}} = 0$ 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 $\dfrac{\partial f}{\partial \bar{z}}(z_0)$.

(3) Just calculate

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

∎

Corollary 3.2.1 *For a holomorphic function $f = u + iv$, we have some details:*
(a) If we see f as a real-value function $F(x, y)$, then for any defined $z = x + iy$

$$det J_F(x, y) = |f'(z)|$$

(b) $f'(z) = u_x + iv_x = v_y - iu_y$

Proof. The result can be caculated by using C-R equation. ∎

4 Functional Anaylsis

4.1 Hilbert Space

It is boring to talk about the finite-dimensional normed space, that is because each such space can be seen the same with Euclidean space $\mathbb{R}^n$ with a certain norm (both in terms of algebraic and topological structures), and functional space is usually a infinite-dmiensional space, so our linear algebra will be not enough (for someone only know martix), that will be the main content in Hilbert space: genenralize the topic about inner product in linear algebra, refine the result with some idea of topology.

Here are some two simple inspiration for the next content. The first is a simple result: **Any finite-dimension subspace of a normed vector space is closed.** The statement will be not correct if the dimension is infinite, hence we can speculate that some good properties we hope will still hold if the subspace is closed. The second is about duality and dual space, that is a beautiful and strong structure. In finite-dimensional case, a linear map can be seen as a matrix M and we often talk about its transpose M^T, it appears frequently in a number of topics involving orthogonal groups, but when the dimension is infinite, it is not easy to see them as a matrix and not easy to consider its transpose, so we need a equivalent concept: adjoint operator, it refers to inner product.

Moreover, the good properties about orthogonality s in finite-dimensional vector space is as following: if $\{e_1, .., e_k\}$ is a orthonormal basis for a vector space E with a inner product, then for any $x \in E$ we have a decomposition:

$$x = \langle e_1, x \rangle e_1 + ... + \langle e_k, x \rangle e_k$$

and its norm will be

$$\|x\|^2 = |\langle e_1, x \rangle|^2 + ... + |\langle e_k, x \rangle|^2$$

That will be very useful for anaylsis. For example, in fourier anaylsis we usually consider a orthonormal basis $\{e^{inx}\}_{i\in\mathbb{Z}}$, but in the time of Fourier mathmaticien just know how to expand function as a fourier series:

$$f \approx \sum_{n\in\mathbb{Z}} \hat{f}(n)e^{inx}$$

it is diffcult to study the convergence of the series. That is a very classic question, new tool and techinc come from algebra gives a very bright methodology. Here i really suggest reader to read some materials: The chapter 6 of [1] it covers complete definition of inner product in finite-dimension space, some techinc will be the same in Hilbert space; The chapter 3, section F of [1], it gives a rough introduction about dual space in the finite-dimension, it is a good

motivation and beginning for someone who wants to know duality; First 3 chapters of [2], it gives a clear motivation about Fourier analysis, and show the diffculties about the topic without advanced tool.

4.1.1 Prerequiste about inner product

Definition 4.1 *An inner product on a $\mathbb{K}$-vector space V is a function:*

$$\langle \cdot, \cdot \rangle : V \times V \to \mathbb{K}$$
$$(u, v) \mapsto \langle u, v \rangle$$

which satisfies following properties:

Positivity: $\langle v, v \rangle \geq 0$ *for all $v \in V$.*

Definiteness: $\langle v, v \rangle = 0$ *if and only if $v = 0$.*

Additivity in first slot: $\langle u + v, w \rangle = \langle u, w \rangle + \langle v, w \rangle$ *for all $u, v, w \in V$.*

Homogeneity in first slot: $\langle \lambda u, v \rangle = \lambda \langle u, v \rangle$ *for all $\lambda \in \mathbb{F}$ and all $u, v \in V$.*

Conjugate symmetry: $\langle u, v \rangle = \overline{\langle v, u \rangle}$ *for all $u, v \in V$.*

 Remark When $\mathbb{K} = \mathbb{R}$, the inner product is actually a symmetric bilinear form which is positive definite; When $\mathbb{K} = \mathbb{C}$, the inner product is actually a Hermitien form which is positive definite.

Inner product has a closed relation with norm, each inner product can define a standard norm as following:

$$\|x\| = \sqrt{\langle x, x \rangle}$$

It is easy to verify that it is a norm (a nice exercise). But for the uniqueness, we will state below:

Proposition 4.1 *Inner product defines a norm uniquely up to homeomorphism. More precisely, if $\langle \cdot, \cdot \rangle$ is a inner product of a vector space V, then there exists a natural map $f : \mathbb{R}^+ \to \mathbb{R}^+$ such that $f(\langle u, u \rangle)$ defines a norm on V.*

Proof. Firstly, for any $x = \langle u, u \rangle$, and any $a \in \mathbb{K}$, we have $f(\langle au, au \rangle) = f(|a|^2 x) = |a| f(x)$. Notice that f must be continous since $\langle u, u \rangle$ defines a norm, and each norm can be seen as a continous map. Hence we just need to find the possible continous function f such that $f(a^2 x) = af(x)$ for any $a, x > 0$, it is not diffcult to get that the unique possible is the map with the form $x \mapsto C\sqrt{x}$. Notice that $\|x\| = \sqrt{\langle x, x \rangle}$ is a norm as we state above, so for any other $\|x\|_c = c\sqrt{\langle x, x \rangle}$ with positive number c, it is still a norm with simple verification. and notice that

$1/M\|x\|_c \le \|x\| \le M\|x\|_c$ for $M \ge \max(c, 1/c)$, so the topology they define are equivalent. ■

According to the standard norm above, we can conclude several properties about inner product:

Proposition 4.2 *Suppose that $\langle \cdot, \cdot \rangle$ is a product of a vector space V, then*
*(a) **Cauchy-Schwartz inequality:** $\langle x, y \rangle \le \|x\| \cdot \|y\|, \quad \forall x, y \in V$*

The equality holds if and only if $x = \lambda y$ with $\lambda \in \mathbb{K} - \{0\}$.
*(b) **Pythagore's formula:** for any $x, y \in V$,*

$$\|x + y\|^2 = \|x\|^2 + 2 \cdot Re(\langle x, y \rangle) + \|y\|^2$$

*(c) **Parallelogram Indentity:** for any $x, y \in V$,*

$$\|x + y\|^2 + \|x - y\|^2 = 2\|x\|^2 + 2\|y\|^2$$

Proof. good exercise. ■

Parallelogram indentity needs an attention here, it is important to determine whether a norm can invite an inner product, the proof is a little complicated.

Theorem 4.3 *A normed vector space can have inner product structure if and only if the norm satisfies **Parallelogram identity**.*

more precisely, for an normed vector space $(V, \| \cdot \|)$ satisfying P.I. we can define inner product as following:
-If V is over $\mathbb{R}$:
$$\langle x, y \rangle := \frac{\|x + y\|^2 - \|x - y\|^2}{4}$$

-If V is over $\mathbb{C}$:

$$\langle x, y \rangle := \frac{\|x + y\|^2 - \|x - y\|^2}{4} - i \cdot \frac{\|x + iy\|^2 - \|x - iy\|^2}{4}$$

Proof. We just need to prove the necessity. It refers to some algebraic caculation so omit some parts. If V is over $\mathbb{R}$, we prove the homogeneity. We fix $u \in V$ such that $\phi_u : x \mapsto \langle x, u \rangle$. then we need to prove $\phi_u(kx) = k\phi_u(x)$ for any $k \in \mathbb{R}$, it is easy to observe that ϕ_u is a continuous map since $\| \cdot \|$ is continous. Moreover, we can prove the additivity: $\phi_u(x+y) = \phi_u(x) + \phi_u(y)$, then by the two propeties (coninitous and additivity), ϕ_u is homogenous automatically.

If V is over $\mathbb{C}$, similarly... ■

Orthogonality is a core concept in inner product, it involves the algebraic treatment of the geometric concept of perpendicularity. here are some vocabulary:

Definition 4.2 $\langle \cdot, \cdot \rangle$ *is the inner product of a vector space.*

   *-**orthogonal**: we call two vector $x, y \in V$ is orthogonal if*

$$\langle x, y \rangle = 0$$

or equiavlently, it satisfying $\|x + y\|^2 = \|x\|^2 + \|y\|^2$.

   *-**orthonormal list**: a set of vector $\{e_i\}_{i \in I}$ is a orthonormal list if for any $m, n \in I$*

$$\langle e_m, e_n \rangle = \delta_{m,n}$$

*In particular if the list spans the vector space, we call it **orthonormal base**.*

   *-**orthogonal complement**: for a non-empty subset $U \subset V$, the orthogonal complement of U is defined by a vector space(**needs a verfication!**) as following:*

$$U^\perp := \{u \in V \,|\, \langle u, v \rangle = 0 \quad \forall v \in V\}$$

There exists a classic method to generate a orthonormal list according to a given linearly independent list or a base, it is called **Gram-Schmidt procedure**.

Proposition 4.4

Finally we discuss the orthogonal projection, it refers to the decompostition of the space, which will be fundamental if we have base function.

Theorem 4.5 (Existence and Uniqueness of projection)
Suppose that F is a closed and convex subset of a vector space V with a inner product $\langle \cdot, \cdot \rangle$, that for any $x \in V$ there exists a unique vector x_F such that

$$\|x - x_F\| = d(x, F) := \inf_{y \in F} \|x - y\|$$

Proof. equivalently, we just need to prove the set $S = \{\|x - y\| : y \in F\}$ is closed, then clearly a such vector x_F will exist by **Supremum Property** of the real number.

such vector is unique, we just suppose that there exists another vector $x' \in F$ such that $\|x - x'\| = d(x, F)$, then by parallegoram equality, we will have

$$\|2x - x_F - x'\|^2 + \|x' - x_F\|^2 = 2\|x - x_F\|^2 + 2\|x - x'\|^2 = 4d^2(x, F)$$

then rewrite it we have

$$\|x' - x_F\| = 4d^2(x, F) - 4\|x - \frac{x_F + x'}{2}\|^2 \leq 0$$

So immediately $x_F = x'$ ∎

Notice that the uniqueness depends on the inner product, for the generla normed space we cant conclude the uniqueness although the existence still holds. With the theorem it makes sense to give orthogonal projection a the formal definition(**Notice that a subspace is indeed convex**):

Definition 4.3 *Give a closed subspace $F \subset V$, the (orthogonal) projection of some $x \in V$ on F is a vector $P_F(x)$ satisfies*

$$\|x - P_F(x)\| = \min_{y \in F} \|x - y\|$$

*or equivalently(**needs a proof**)*

$$\langle x - P_F(x),\, y \rangle = 0 \quad \forall y \in F$$

Accoding to these properties we can conclude the fundamental decomposition accodring to the orthogonality:

Theorem 4.6 *F is a closed of vector space V, then*

$$F \oplus F^{\perp} = V$$

Proof. It is immediately if we write any $x \in V$ as the form:

$$P_F(x) + (x - P_F(x)) = x$$

∎

So it makes sense to consider projection map

$$P_F : V \to F, x \mapsto P_F(x)$$

Here are its properties:

Proposition 4.7 *F is a closed subspace of vector space V.*
(a) P_F is a linear map in $\mathcal{L}(V)$.
(b) P_F is continous.
(c) P_F is surjective with $P_F(x) = x$ for any $x \in F$.
(f) $\mathrm{Im} P_F \oplus \ker P_F = V$
(d) $P_F \circ P_F = P_F$
(e) P_F is a self-adjoint opreator, i.e. for any $x, y \in V$

$$\langle P_F(x),\, y \rangle = \langle x,\, P_F(y) \rangle$$

notice that if $\dim V < +\infty$ and $M = [P_F]_{\mathcal{B}}$, the condition is same with $\overline{M^T} = M$.

With the baisc knowledge we can now formally define the hilbert space:

Definition 4.4 $(H, \langle \cdot, \cdot \rangle)$ *is a **pre-hilbert space** if H is a inner product space over $\mathbb{R}$ or $\mathbb{C}$; if the a pre-hilbertspace is complete under the standard norm $\|x\| = \sqrt{\langle x, x \rangle}$, then it is a **hilbert space**.*

Example 4.1 *Consider the case when $H = \mathbb{C}$, complex number forms a complet space by the modulus $\|z\|^2 = z\overline{z}$, so we can deine in it $\langle x, y \rangle = x\overline{y}$. And we can generalize it for $H = \mathbb{C}^n$, it gives **general Euclidean space** with the inner product:*

$$\langle x, y \rangle = x_1\overline{y_1} + ... + x_n\overline{y_n}$$

It is the classic case of hilbert space.

Example 4.2 *L^p space is a hilbert space only if $p = 2$, since its norm must satisfy the parallegoram indentity, so we take two disjoint and finite measurable set A and B, then consider $f = 1_A$ and $g = 1_B$, then by indentity:*

$$|\mu(A)|^{2/p} + |\mu(B)|^{2/p} = |\mu(A) + \mu(B)|^{2/p}$$

If p is finite, the equality holds only if $p = 2$. For $p = +\infty$, same identity we have the contradiction:

$$2\|f\|_\infty^2 + 2\|g\|_\infty^2 = 4 \neq 2 = \|f + g\|_\infty^2 + \|f - g\|_\infty^2$$

In the case $p = 2$ we can define inner product on $L^2(X, \mathcal{E}, \mu)$ space by

$$\langle f, g \rangle = \int_X = f\overline{g}d\mu$$

4.1.2   Hilbert base

5 Probability

5.1 Random variable

Random variable is the core of Probability, just like the place of function in Analysis, or generally probability is just a special version of measure theory, here are the correspondence:

$$\text{Probability Space} \Longleftrightarrow \text{Measure space}$$
$$\text{random variable} \Longleftrightarrow \text{measurable function}$$

A probability space is usually denoted by the form

$$(\Omega, \mathcal{T}, P)$$

It refers to the probability experiment observed. Here Ω is the sample space, it conatains all possible outcomes; $\mathcal{T}$ is a sigma-algebra which contains all event we expected; Finally, probability P is the measure satisfying $P(\Omega) = 1$. (What that means? it means it is a function which assigns a **"number between 0 and 1"** to each event in $\mathcal{T}$, and it keeps linear in the sense of disjoint, this type of correspondence between expected events and a number with order structure, "1" here is not important because any finite measure can be rewritten to be a probability).

In a philosophical sense, the definition of the sample space is objective, whereas the definition of the event space is subjective, accompanied by a constraint in the form of a logical axiom. For example, we choose a ball from n balls, here the sample space is surely $\Omega = \{x_1, ..., x_n\}$ by labeling balls, but the choose of event space $\mathcal{T}$ can be different: (1) we can take $\mathcal{T} = \mathcal{P}(\Omega)$, then in this stuation we can choose all single ball respectively; (2) we can take

$$\mathcal{T} = \{\{x_1\}, \{x_2, ..., x_n\}, \Omega, \emptyset\},$$

actually it is same with Bernoulli test,in fact the stuation means that we can take either the ball labelled x_1 or the other rest balls. (1) and (2) are totally different, it depends on what exactly experiment we want to observe.

We start the definition of random variable by observing above two examples, we want to do some mathematics about this two cases, clearly this two cases are discrete and simple, in mathematics it is easy to identify $S_k = \{1, 2, ..., k\}$ and we take $\mathcal{E} = \mathcal{P}(S)$, here we call the measurable space $(S_k, \mathcal{E}_k)$ as a **k-standard space**, in above example (1) can be seen the case here $k = n$, example (2) can be seen the case here $k = 2$, the correspondence is based on the measurable function:

-case (1)

$$X_1 : \Omega \to S_n, \quad x_i \mapsto i$$

-case (2)

$$X_2 : \Omega \to S_2, \quad x_i \mapsto \begin{cases} 1 & i = 1 \\ 2 & i \neq 1 \end{cases}$$

in case (1) we can define the probability $P(\{x_i\}) = p_i$, then measurable function X_1 invites a probability for standard space by measure by image: $P(X_1 = i) = P \circ X_1^{-1}(\{i\}) = p_i$. Similarly for case (2), so random variable is just the measurable function, it conatins too much infrmation:exact details of the event space $\mathcal{T}$; the corresponding standard space, a certain well-defied probability experiment (like Bernoulli test, normal distribution); the probability of the event.

A most important things here is that a random variable is similar to "a quoteint map", in probability we usually prefer to care about the result of the random variable rather than the sample space, or in other words we care about the event space. For example, iif we throw the dice twice, we're probably more interested in the sum of the two dice points than in the exact number of points.

«««< HEAD =======

5.1.1 Basic Property

Definition 5.1 *Let $(\Omega, \mathcal{T}, P)$ be a probability space, then*

- Random variable $X : \Omega \to \mathbb{R}$ is a measurable function. The expectation of the random variable is the integral defined as following (Surely it is at least integrable)

$$\mathbb{E}(X) = \int_\Omega X dP$$

and the variance of the random variable is defined as following (it is at least square integrable)

$$\mathbb{V}(X) = \mathbb{E}[(X - \mathbb{E}(X))^2] = \mathbb{E}(X^2) - \mathbb{E}(X)^2$$

the covriance of two random variable X, Y is defined as following

$$\mathrm{Cov}(X, Y) = \mathbb{E}[(X - \mathbb{E}(X))(Y - \mathbb{E}(Y))]$$

- The loi of the random variable $X : \Omega \to \mathbb{R}$ is the probability measure defined on the standard Borel space $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ as following

$$P_X(A) = P(X \in A) = P \circ X^{-1}(A)$$

For any borel set $A \subset \mathbb{R}$

The properties of of the expectation and variance is as following

Proposition 5.1 *Let X, Y be two (square) integrable random variable and $c \in \mathbb{R}$, then*

-Expectation:

(1) Linear: $\mathbb{E}(X + cY) = \mathbb{E}(X) + c\mathbb{E}(Y)$

(2) Monotone: If $X(w) \leq Y(w)$ for almost every $w \in \Omega$, then

$$\mathbb{E}(X) \leq \mathbb{E}(Y)$$

(3) Convergence: $(X_n)_{n \geq 0}$ is the sequence of random variable converging to the random variable X, and there exists a integrable random variable Y such that $|X_n| \leq Y$ for any n, then

$$\mathbb{E}(X_n) \xrightarrow[n \to \infty]{} \mathbb{E}(X)$$

(4) For any event A, we have $P(A) = E(1_A)$

-Variance:

(1) $\mathbb{V}(cX) = c^2 \cdot \mathbb{V}(X)$

(2) $\mathbb{V}(X + Y) = \mathbb{V}(X) + \mathbb{V}(Y) + 2 \cdot \mathrm{Cov}(X, Y)$

Above properties is just the translation from integral theory, there are nothing new and the most improtant things is to understand the sense behind the definition: (1) the expectation shows the real average of a random variable in the long term. (2) the variance reflects the the degree of dispersion of the value of the random variable. (3) the covriance reflects the degree of the correlation of two random variable.

»»»> 86b7e621dbb65ab673d629589d16e23a2743cac9

5.2 Independence

Definition 5.2 *Given a probability space $(\Omega, \mathcal{T}, P)$*

- *events $A_1, A_2, ... \in \mathcal{T}$ are called independent if for any $i \neq j$,*

$$P(A_i \cap A_j) = P(A_i)P(A_j)$$

- *random variables $X_1, X_2, ...$ are called independent if for any $i \neq j$,*

$$P(X_i \in A; Xj \in B) = P(X_i \in A)P(X_j \in B) \quad \forall A, B \in \mathcal{B}(\mathbb{R})$$

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 5.2 *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 5.2.1 *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

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

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

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

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 after caculation. ∎

5.3 Convergence

The convergence of the sequence of random variable is realted to functional sequence, we can classfiy different types of convergence, they are very delicated.

Definition 5.3 *Given a sequence of random variable $(Y_n)_{n \in \mathbb{N}}$ of $(\Omega, \mathcal{T}, P)$, and $p \geq 1$:*

- *$(Y_n)_{n \in \mathbb{N}}$ converges to Y **by norme** L^p if*

$$
\|Y_n - Y\|_p \xrightarrow[n \to \infty]{} 0
$$

- *$(Y_n)_{n \in \mathbb{N}}$ converges to Y **by measure** (probability) if*

$$
\forall c > 0, \quad P(|Y_n - Y| > c) \xrightarrow[n \to \infty]{} 0
$$

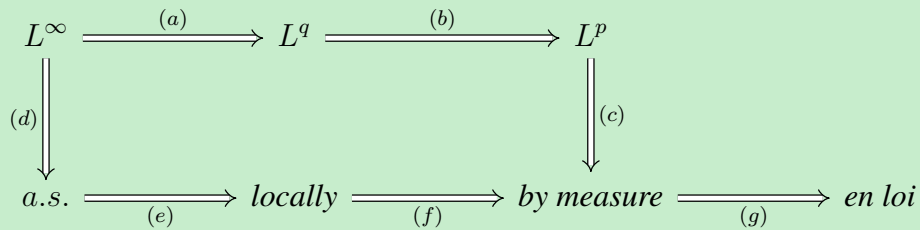- *$(Y_n)_{n \in \mathbb{N}}$ converges to Y **almost surely** if*

$$
\text{for almost every } w \in \Omega, \quad Y_n(w) \xrightarrow[n \to \infty]{} Y(w)
$$

- *$(Y_n)_{n \in \mathbb{N}}$ converges to Y **en loi** if for any continous bounded function $f : \mathbb{R} \to \mathbb{R}$:*

$$
E[f(Y_n)] \xrightarrow[n \to \infty]{} E[f(Y)]
$$

- *$(Y_n)_{n \in \mathbb{N}}$ **locally converges** to Y if there exists a subsequence $Y_{\varphi(n)}$ almost surely convering to Y.*

Proposition 5.3 (Implication of the convergence)

where $1 \leq p \leq q \leq \infty$

Proof. (a) ∎

5.3.1 Stlutsky's theorem

Theorem 5.4 *Given two sequences of random variable X_n and Y_n on $\mathbb{R}^d$ and $\mathbb{R}^m$ respectively. If X_n converge en loi to X and Y_n converge by probability to a constant c, then couple (X_n, Y_n) converge en loi to the couple (X, c)*

Proof. ∎

Just choice suitable bounded and continous function f then immediately we have below strong result, it is more useful:

Corollary 5.4.1 *In the same case we can conclude that*
(1) The r.v. sequence $X_n + Y_n$ converges en loi to $c + X$.
(2) The r.v. sequence $X_n Y_n$ converges en loi to cX.

5.3.2 characteristic function

We notice that the convergence of the r.v. sequence is essentially the convergence of the sequence of functions. A classic techinc is from fourier analysis and integral transform, which allows equivalently determine the convergence of the functions after transform.

Definition 5.4 (characteristic function)
For a random variable $X : \Omega \to \mathbb{R}$, its characteristic function is the Fourier transform of its loi (probability measure by image):

$$\varphi_X(t) := \mathcal{F}(P_X)(t) = E[e^{itX}]$$

Here are some basic properties of characteristic function:
(1) $\varphi : \mathbb{R} \to \mathbb{C}$ is a continous and bounded function:

$$|\varphi_X(t)| \leq E[|e^{itX}|] \leq 1$$

(2) $\varphi_X(0) = 1$.
(3) If $X \in L^p(\Omega)$, then $\varphi_X \in C^p(\mathbb{R})$ with a good derivative at $t = 0$:

$$\varphi'_X(0) = iE(X), \quad \varphi''_X(0) = -E(X^2)$$

The proof is based in the integral by parameter. Notice that if X is a discrete random, then its characteristic function will still be continous and even differentiable, and the differentiation allow us to do carefully analysis.

Proposition 5.5 *Two random variables X and Y are identically distributed if and only if they have the smae characteristic function: $\varphi_X = \varphi_Y$.*

Remark The sufficence is claerly since $P_X = P_Y$ means they will define the same measure on a real measurable space. We will not prove the necessity directly since it will be a immediate result from Levy's convergence theorem and that will be the core of the section.

We rewrite the definition of the convergence en loi by transfert theorem:

$$E[f(X_n)] \xrightarrow[\to +\infty]{} E[f(X)]$$

which will be equivalent to

$$\int_\Omega f \, dP_{X_n} \xrightarrow[\to +\infty]{} \int_\Omega f \, dP_X$$

That means the convergence depends on the measure invited by the sequence of random variable X_n, specially if we take $f(x) = e^{itx}$ we will obtain the convergence:

$$\varphi_{X_n}(t) \xrightarrow[\to +\infty]{} \varphi_X(t), \quad \forall t \in \mathbb{R}$$

or using the language of the Fourier transform:

$$\hat{P}_X(t) \xrightarrow[\to +\infty]{} \hat{P}_X(t), \quad \forall t \in \mathbb{R}$$

Notice there the form of function f does not satisfy the definition of the convergence en loi, so deeply it refers to prove the equivalence of two method of convergence about measure:

Lemma 5.6 *Suppose that μ and $(\mu)_{n \in \mathbb{N}}$ are probability measure on a real measurable space $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$, then following statements are equiavlent:*

(a) $\displaystyle\int_\Omega f \, d\mu_n \xrightarrow[n \to \infty]{} \int_\Omega f \, d\mu$ *for any continuous and bounded function $f : \mathbb{R} \to \mathbb{R}$,*

(b) $\displaystyle\int_\Omega f \, d\mu_n \xrightarrow[n \to \infty]{} \int_\Omega f \, d\mu$ *for any $f \in C^\infty(\mathbb{R})$ with compact support,*

(c) $\hat{\mu}_n \xrightarrow[n \to \infty]{} \mu$ *for any $t \in \mathbb{R}$.*

Proof. ∎

Then we can conclude an method to determiner the convergence of random variable, which is different Stlusky's theorem, and usually more useful to handle the case if the random variable is exactly given:

Theorem 5.7 (Levy)

$$X_n \xrightarrow{en\ loi} X_n \iff \varphi_{X_n} \xrightarrow{\forall t \in \mathbb{R}} \varphi_X$$

5.4   Law of large number

Definition 5.5 *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 identically distrbuted random variable.

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

Proposition 5.8 *If X_1 and X_2 are two independent and 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 independent and identically distributed.*

Proof. The condition for independence can be that f is just a borel function. We just need to verify that for any Borel set B, there exists a borel set A such that

$$(f \circ X)^{-1}(B) = X^{-1} \circ f^{-1}(B) = X^{-1}(A)$$

then it is clear. For id. we notic 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. $\blacksquare$

We have same result for the independence of two random variable, Now we can conclude the important theorem in probability theory:

Theorem 5.9 (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 $\dfrac{S_n}{n}$ converges to zero almost surely.

Corollary 5.9.1 *For any borel set A, we have a almost surely convergence*

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

Proof. Notice that $\#\{i \le n | X_i(w) \in A\} = \sum_{i=0}^{n} 1_{\{X_i \in A\}}$, so we take $Y_k = 1_{X_k \in A}$ as the new sequence of random variable, then using the theorem we get that the it will converge to $E[Y_0] = P(X_0 \in A)$. ∎

Theorem 5.10 (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 $\dfrac{S_n}{n}$ converges to $E[X_0]$ 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. ∎

   Notice that the condition of integrable random variable is important, that means that $E(|X|) < +\infty$, it makes sense to talk about the law of grand number if the experence is finite. Here we give a theortical application of law of large number. If X_1, X_2, X_3 are independent, then...

5.5   Central limit theorem

   As its name, CLT shows that the sum of large numbers of independent random variable nearly subjects to the normal disrtibution (Guass distribution). We will prove this theorem via the characteristic function.

Lemma 5.11 *Suppose that $Z \sim \mathcal{N}(0,1)$ is a random variable subjecting to the standard normal distribution, then its characteristic function will be*

$$\varphi_Z(t) = e^{-t^2/2}$$

Gnenerally if $Y \sim \mathcal{N}(\mu, \sigma^2)$ is an other random vraible sujecting to a normal distrbution, then its characteristic function will be

$$\varphi_Y(t) = e^{it\mu - \frac{\sigma^2}{2}t^2}$$

Proof. wait.... ∎

   With the help of the lemma we can prove CLT. Notice that the convergence of the characteristic function will reflect the convergence of the distribution, accoring to theorem of Levy.

Theorem 5.12 *In a probability space $(\Omega, \mathcal{T}, P)$, let $(X_n)_{n \in \mathbb{N}}$ be a sequence of random vraiable which staifies:*

(1) X_n is independent and identaically distributed.

(2) $X_n \in L^2(\Omega)$ squeare integrable.

(3) $V(X_n) \neq 0$

Then For the sum of random variable $S_n = \sum_{k=1}^{n} X_k$, we have convergence en loi:

$$\frac{\sqrt{n}}{\sigma(X_0)} \left(\frac{S_n}{n} - \mathbb{E}(X_0) \right) \xrightarrow[n \to \infty]{en\ loi} X \sim \mathcal{N}(0,1)$$

In particular, for any interval $[a, b] \subset \mathbb{R}$ we have the convergence about the probability:

$$P\left(a \leq \frac{\sqrt{n}}{\sigma(X_0)} \left(\frac{S_n}{n} - \mathbb{E}(X_0) \right) \leq b \right) \xrightarrow[n \to \infty]{} \frac{1}{\sqrt{2\pi}} \int_a^b e^{-x^2/2} \, dx.$$

Proof. ∎

Usually we take a sequence with zero experence, then we have an elegant formula:

$$\frac{S_n}{\sqrt{n}} \xrightarrow[n \to \infty]{en\ loi} X \sim \mathcal{N}(0, V(X_0))$$

5.6 Examples of random vraiable

5.7 discrete case

A random variable $X : \Omega \to \mathbb{R}$ is discrete if the image of X is discrete, in that case we have

$$P(A) = \sum_{a \in A} P(\{a\})$$

the probability depends on each single point. Following we firstly define some standard distribution:

Example 5.1 (Uniform distribution)

We put $S_n = \{1, ..., n\}$ and $(S_n, \mathcal{P}(S_n))$ is a measurable space equipped with the probability

$$P(A) = \frac{\#A}{n}$$

Then the probability space is a standard uniform distribution, more generally, the definition can be extended to $(E, \mathcal{P}(E), P)$ with E a finite subset of $\mathbb{R}$, and P is a probability defined as following

$$P(A) = \frac{\#A}{\#E}$$

It is the simplest case, we call $X \sim \text{Unif}(E)$ a random variable X subjects to a unifrom disrtibution on E if X is discrete, and $(X(\Omega), \mathcal{P}(X(\Omega)), P_X)$ equals to the standard uniform distribution $(E, \mathcal{P}(E), P)$ defined above.

Example 5.2 (Bernoulli Test)

We take $B = \{0, 1\}$ and $(B, \mathcal{P}(B))$ a measurable space equipped with the testing probability

$$P(\{0\}) = 1 - p, \quad P(\{1\}) = p$$

for some $p \in [0, 1]$, then we call the $(B, \mathcal{P}(B), P)$ a standard bernoulli p-testing.

6 Classic Mechanics

It is too complicated to give a complete description of the mathematics structure of mechanic, it is one of the oldest topic in physical and mathematics.

6.1 Conservation of mechanical energy

Work naturally is a line integral of a vector field, here the vector is the force acting on the object, formally

$$W_f = \int_{t_0}^{t_1} \mathbf{f} \cdot \mathbf{v} dt = \int_C \mathbf{f} d\mathbf{r}$$

here $\mathbf{f}$ is the force.

there exists a direct relationship between work and kinetic energy, to apply Newton's second law, we can find

$$\mathbf{f} \cdot \mathbf{v} = m\ddot{\mathbf{r}} \cdot \dot{\mathbf{r}} = \frac{1}{2}m\frac{d}{dt}(\dot{\mathbf{r}} \cdot \dot{\mathbf{r}}) = \frac{1}{2}m\frac{d}{dt}(\mathbf{v}^2) = \frac{d}{dt}\left(\frac{1}{2}m\mathbf{v}^2\right).$$

Therefore we can conclude the **work-energy theorem**: for a single practicle, the work done by the net force equals the change of kinetic energ,

$$W_f = \int_{t_0}^{t_1} \mathbf{f} \cdot \mathbf{v} dt = T(t_1) - T(t_0)$$

Consider the model of free falling, the object fall from a certain height, the gravitational potential energy decrease as the kinectic energy incrases, it is easy to verify the sum of energy is the same at any time; Conisder another model, when a moving car suddenly stalls, its kinetic energy gradually decreases to zero as the speed decreases, the energy loss in this process is caused by the friction from the ground. Comparing this two example, we can find that the conservation of the energy depends the system we observe and the force acting on the object: in the two system, if we contain the road into the system, the kinectic energy in facts is transformed into the heat energy, which can not denoted by potential energy, so in the example there exist no potential energy, but the energy is conserved in the sense of the larger system.

Formally, we define conservative force by defining the conservative vector field

Definition 6.1 *Let $F : U \subset \mathbb{R}^3 \to \mathbb{R}^3$ be a vector field, it is said to be conservative if there exists a C^1 scalar function $\varphi : U \to \mathbb{R}^3$ such that*

$$-\nabla\varphi = F$$

References

[1] Sheldon Axler. *Linear algebra done right*. Springer Nature, 2024.

[2] Elias M Stein and Rami Shakarchi. *Fourier analysis: an introduction*. Vol. 1. Princeton University Press, 2011.