

Math remark fondation for analysis and probability

X

ElegantL^AT_EX Program

Update: February 9, 2025

Contents

1	Genenralisation of several inequalities	3
1.1	Jense,Hölder,Minkeoski's inequalities	3
1.2	Markov, Tchebychev, Cantelli's inequalities	8
2	Classic anaylsis theory and technic	12
2.1	Power series and analytic function	13

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

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

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 applying 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 probability space can be written 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 $t \mapsto \ln t$ 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 inequality 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 $f, g : (X, \mathcal{E}, \mu) \rightarrow \mathbb{C}$, we have

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

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 $t \mapsto \ln t$ is a concave function on $(0, +\infty)$, so we can estimate by Jense's inequality

$$\begin{aligned} \ln(uv) &= \ln(u + v) = \frac{1}{p} \ln u^p + \frac{1}{q} \ln v^q \\ &\leq \ln\left(\frac{1}{p} u^p + \frac{1}{q} v^q\right) \end{aligned}$$

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

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

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 (Cauchy-Schwartz)

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

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

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

$$| \langle u, v \rangle | \leq \|u\| \|v\|$$

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

$$\begin{aligned} p(t) &= t^2 z \bar{z} \langle u, u \rangle + tz \langle u, v \rangle + t \bar{z} \langle v, u \rangle + \langle v, v \rangle \\ &= t^2 z \bar{z} \langle u, u \rangle + tz \langle u, v \rangle + \overline{tz \langle u, v \rangle} + \langle v, v \rangle \\ &= t^2 \|u\|^2 + 2|\langle u, v \rangle| t + \|v\|^2 \end{aligned}$$

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 sufficient and necessary condition that

$$\Delta = 4|\langle u, v \rangle|^2 - 4\|u\|^2 \|v\|^2 \geq 0$$

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 (Minkowski)

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

$$\int_X |f + g|^p d\mu \leq \int_X |f|^p d\mu + \int_X |g|^p d\mu$$

The equality holds iff $|f| = c|g|$ almost everywhere (μ -a.e.) 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 (Ω, \mathcal{T}, P) to denote a probability space

Theorem 1.4 (Markov)

Suppose that $Y : \Omega \rightarrow \mathbb{R}^+$ 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

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

■

Theorem 1.5 (Tchebychev)

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

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

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

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

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 $t \geq \sigma$.

2 Classic analysis theory and technic

2.1 Power series and analytic function

This section contains 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 a special case of the series of the function, it gives by the following sequence:

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

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 sequence implies a real number M such that $a_n z_0^n \leq M$, so

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

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 can't converge to zero, so the power series diverge. ■

The lemma inspires us that the domain 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

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

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

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

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

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

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 $r \rightarrow R$. 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 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 (Taylor's series)

If $f(z)$ is the sum of the power series $\sum_{n \in \mathbb{N}} a_n z^n$ on the domain $D(0, R)$ 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 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 radius 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

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

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 $f : U \rightarrow \mathbb{K}$ 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 domain $D(z_0, r) \subset U$ 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 radius 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 Taylor's Formula, so the question in \mathbb{R} is equivalent to the question that for each smooth function whether its Taylor series converges to itself or not. In conclusion, the relation show as following 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 $f : U \rightarrow \mathbb{K}$, and take $a \in U$, the analytic function implies a positive $r > 0$ and a sequence (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 guarantee 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

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

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

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

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 Able 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 seriees 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 prove that

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

if $\sum_{n \in \mathbb{N}} a_n$ is a convergent 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, holomorphic function is just the differentiable functions in \mathbb{C} , why do we not just call them "a differentiable 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 diffrenetiable and with the form

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

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} :

analytic function \iff 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 serieis $\sum_{n \in \mathbb{N}} \frac{f^{(n)}(z_0)}{n!} (z - z_0)^n$ 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 we can get the result we hope. ■

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