

# Math remark fondation for analysis and probability

X

ElegantL<sup>A</sup>T<sub>E</sub>X Program

*Update: February 7, 2025*

# Contents

<b>1</b>	<b>Genenralisation of several inequalities</b>	<b>3</b>
1.1	Jense,Hölder,Minkeoski's inequalities . . . . .	3
1.2	Markov, Tchebychev, Cantelli's inequalities . . . . .	8
<b>2</b>	<b>Classic anaylsis theory and technic</b>	<b>12</b>
2.1	Power series . . . . .	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,  $(X, \mathcal{E}, \mu)$  is a probability space and  $f \in L^1(X)$  with  $\text{im} f \subset I$ , then  $\int_X f d\mu \in I$  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 ( $\mu$ -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  $(\Omega, \mathcal{T}, P)$  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

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

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

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 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  $\sum_{n \in \mathbb{N}} a_n z^n$  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 *power series* is given by

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