不等式の話

三角不等式が重要．三角不等式たちがどこから来たか．

(1) 絶対値 $|x+y| \leqq |x|+|y|$．

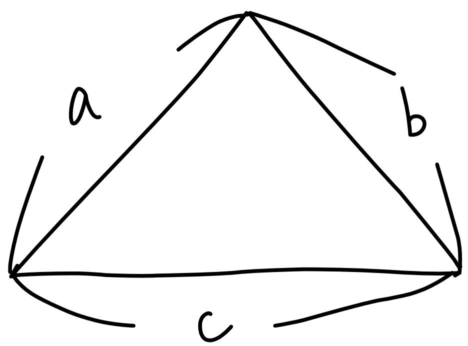
これは $|x-y| \geqq \max\{|x|-|y|,\ |y|-|x|\}$ の形でも使われる．

$|x| = |x-y+y| \leqq |x-y|+|y| \quad \therefore \quad |x-y| \geqq |x|-|y|$

$|x-y| = |y-x| \geqq |y|-|x|$．

$|x-y| = |x-z+z-y| \leqq |x-z|+|z-y|$ の形でもよく使われる．

(2) 平面上での直線距離 (Euclid距離)



$c \leqq a+b$

(3) $\mathbb{R}^2$ 上の2つの点 $(0,0)$ と (x,y) の距離 $\|(x,y)\|_1$ を

$$\|(x,y)\|_1 = |x|+|y|$$

と定めることもできる（ℓ^1 ノルム，ℓ^1 距離）．このとき，

$$\begin{aligned}\|(x,y)+(x',y')\|_1 &= \|(x+x',y+y')\|_1 = |x+x'|+|y+y'| \\ &\leqq |x|+|x'|+|y|+|y'| = \|(x,y)\|_1 + \|(x',y')\|_1.\end{aligned}$$

三角不等式

(4) 一般に $p \geqq 1$ に対して，$\mathbb{R}^2$ 上の ℓ^p ノルム $\|\cdot\|_p$ を

$$\|(x,y)\|_p = \left(|x|^p + |y|^p\right)^{\frac{1}{p}}$$

と定めることができる．これの三角不等式は Minkowski の不等式の特別な場合になっている．

これを証明したい．

Jensenの不等式

Iは$\mathbb{R}$内の区間であるとし, $\mathcal{F}$はI上の実数値函数からなるベクトル空間で1とxを含むものとする. 写像 $E[\cdot]:\mathcal{F}\to\mathbb{R},\ f(x)\mapsto E[f(x)]$ は $E[x]\in I$ および以上の条件をみたしていると仮定する:

(E1) 線形性: $E[\alpha f(x)+\beta g(x)]=\alpha E[f(x)]+\beta E[g(x)]$ $(f,g\in\mathcal{F},\ \alpha,\beta\in\mathbb{R})$

(E2) 単調性: $f,g\in\mathcal{F}$がI上で$f\leqq g$をみたす $\Longrightarrow E[f(x)]\leqq E[g(x)]$.

(E3) 規格化: $E[1]=1$.

このとき, $E[\cdot]:\mathcal{F}\to\mathbb{R}$を期待値汎函数と呼ぶ.

定理 $f\in\mathcal{F}$が$\overset{\text{下}}{\text{上}}$に凸な函数ならば $E[f(x)]\overset{\geqq}{\leqq}f(E[x])$. □

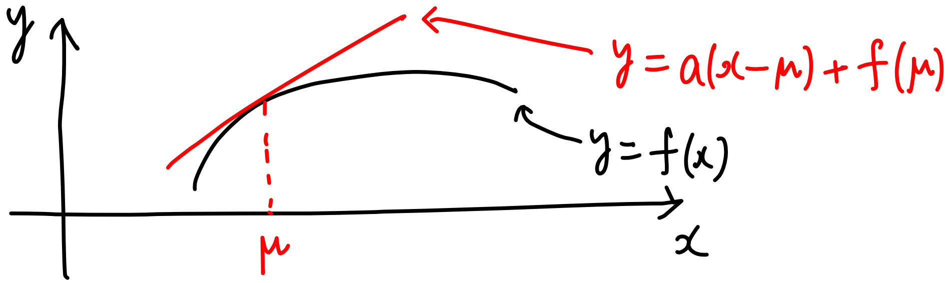
例 $I=\mathbb{R}_{>0},\ x_1,\dots,x_n\in I,\ E[f(x)]=\dfrac{1}{n}\sum_{i=1}^{n}f(x_i)$ のとき, $f(x)=\log x$ に Jensenの不等式を適用すると, $\log(x_1\cdots x_n)^{1/n}=\dfrac{1}{n}\sum_{i=1}^{n}\log x_i\leqq\log\left(\dfrac{1}{n}\sum_{i=1}^{n}x_i\right)$

ゆえに, $(x_1\cdots x_n)^{1/n}\leqq\dfrac{1}{n}\sum_{i=1}^{n}x_i$ (相加相乗平均の不等式!) □

$\boxed{\text{定理}}$ $f\in\mathcal{F}$ が$\overset{\text{下}}{\text{上}}$に凸な函数ならば $E[f(x)] \overset{\geqq}{\leqq} f(E[x])$. □

$\boxed{\text{証明}}$ この講義では f が C^2 級で I 上で $f'' \leqq 0$ となる場合にのみ証明しておけば十分なので，そのように仮定する.

このとき，$\mu := E[x] \in I$ における $y=f(x)$ の接線 $y=a(x-\mu)+f(\mu)$ が存在して，I 上で $f(x) \leqq a(x-\mu)+f(\mu)$ となる.



ゆえに (E2), (E1), (E3) を順番に使えば

$$\begin{aligned}
E[f(x)] &\leqq E[a(x-\mu)+f(\mu)] && \text{by 単調性 (E2)}\\
&= a(E[x]-\mu E[1]) + f(\mu)E[1] && \text{by 線形性 (E1)}\\
&= a(E[x]-\mu) + f(\mu) && \text{by 規格化 (E3)}\\
&= f(E[x]) && \text{by } \mu = E[x].
\end{aligned}$$

$\boxed{\text{q.e.d.}}$

$\boxed{\text{Youngの不等式}}$ $p, q > 1$, $\frac{1}{p} + \frac{1}{q} = 1$ とする。このとき、

$$a, b \geqq 0 \Longrightarrow ab \leqq \frac{a^p}{p} + \frac{b^q}{q}.$$

$\boxed{\text{証明}}$ $ab = 0$ の場合は自明なので $a, b > 0$ と仮定してよい。

$E[f(x)] = \frac{1}{p} f(a^p) + \frac{1}{q} f(b^q)$ と $f(x) = \log x$ に Jensenの不等式を使うと、

$$\log(ab) = \frac{1}{p} \log a^p + \frac{1}{q} \log b^q = E[\log x] \leqq \log E[x] = \log\left(\frac{a^p}{p} + \frac{b^q}{q}\right).$$

$$\therefore \quad ab \leqq \frac{a^p}{p} + \frac{b^q}{q}.$$

$\boxed{\text{q.e.d.}}$

$\boxed{\text{Hölderの不等式}}$ $p, q > 1$, $\frac{1}{p} + \frac{1}{q} = 1$ とする.

$$a_1, \dots, a_n, b_1, \dots, b_n \geqq 0 \Rightarrow \sum_{i=1}^n a_i b_i \leqq \left(\sum_{i=1}^n a_i^p\right)^{1/p} \left(\sum_{i=1}^n b_i^q\right)^{1/q}.$$

$p=q=2$ の場合の Cauchy-Schwarz の不等式の一般化

$\boxed{\text{証明}}$ $a_1 = \cdots = a_n = 0$ または $b_1 = \cdots = b_n = 0$ のときは自明なので,

そうではないと仮定し,

$$A_i = \frac{a_i}{\left(\sum_{i=1}^n a_i^p\right)^{1/p}}, \quad B_i = \frac{b_i}{\left(\sum_{i=1}^n b_i^q\right)^{1/q}}$$

$$A_i B_i \leqq \frac{A_i^p}{p} + \frac{B_i^q}{q}$$

とおいて, $\sum_{i=1}^n A_i B_i \leqq 1$ を示せばよい. このとき, Youngの不等式より,

$$\sum_{i=1}^n A_i B_i \leqq \sum_{i=1}^n \left(\frac{A_i^p}{p} + \frac{B_i^q}{q}\right) = \frac{1}{p} \underbrace{\sum_{i=1}^n A_i^p}_{=1} + \frac{1}{q} \underbrace{\sum_{i=1}^n B_i^q}_{=1} = \frac{1}{p} + \frac{1}{q} = 1.$$

$\boxed{\text{q.e.d.}}$

$\boxed{\text{Minkowskiの不等式}}$ $p>1$, $x_i, y_i \in \mathbb{R}$ のとき,

三角不等式の一種

$$(**) \quad \left(\sum_{i=1}^n |x_i+y_i|^p\right)^{1/p} \leqq \left(\sum_{i=1}^n |x_i|^p\right)^{1/p} + \left(\sum_{i=1}^n |y_i|^p\right)^{1/p},$$

$\boxed{\text{証明}}$ $a_i=|x_i|$, $b_i=|y_i|$ とおくと, $((**)\text{の左辺}) \leqq \left(\sum_{i=1}^n (a_i+b_i)^p\right)^{1/p}$ なので

$$(**)' \quad \left(\sum_{i=1}^n (a_i+b_i)^p\right)^{1/p} \leqq \left(\sum_{i=1}^n a_i^p\right)^{1/p} + \left(\sum_{i=1}^n b_i^p\right)^{1/p}$$

を示せばよい.

$q>1$ を $\frac{1}{p}+\frac{1}{q}=1$ で定めると, $p+q=pq$, $p=(p-1)q$ なので

$$\sum_{i=1}^n (a_i+b_i)^p \overset{\text{ポイント}}{=} \sum_{i=1}^n a_i(a_i+b_i)^{p-1} + \sum_{i=1}^n b_i(a_i+b_i)^{p-1}$$

Hölderの不等式

$$\leqq \left(\sum_{i=1}^n a_i^p\right)^{1/p}\left(\sum_{i=1}^n (a_i+b_i)^{(p-1)q}\right)^{1/q} + \left(\sum_{i=1}^n b_i^p\right)^{1/p}\left(\sum_{i=1}^n (a_i+b_i)^{(p-1)q}\right)^{1/q}$$

$$= \left(\sum_{i=1}^n a_i^p\right)^{1/p}\left(\sum_{i=1}^n (a_i+b_i)^p\right)^{1-\frac{1}{p}} + \left(\sum_{i=1}^n b_i^p\right)^{1/p}\left(\sum_{i=1}^n (a_i+b_i)^p\right)^{1-\frac{1}{p}}$$

なので両辺を $\left(\sum_{i=1}^n (a_i+b_i)^p\right)^{1-\frac{1}{p}}$ でわると $(**)'$ を得る.

$\boxed{\text{q.e.d.}}$

$\boxed{\text{Gibbsの情報不等式}}$ $p_i, q_i > 0$, $\sum_{i=1}^n p_i = \sum_{i=1}^n q_i = 1 \Rightarrow \sum_{i=1}^n q_i \log\frac{q_i}{p_i} \geqq 0$.

$\boxed{\text{証明}}$ $E[f(x)] = \sum_{i=1}^n f\left(\frac{q_i}{p_i}\right)p_i$ と $f(x) = x\log x$ (下に凸) にJensenの不等式を使う. $f'(x) = \log x + 1$, $f''(x) = \frac{1}{x} > 0$ $(x>0)$ なので $f(x)$ は下に凸である. ゆえに,

$$\sum_{i=1}^n q_i \log\frac{q_i}{p_i} = \sum_{i=1}^n p_i \frac{q_i}{p_i}\log\frac{q_i}{p_i} = E[f(x)] \geqq f(E[x]) = f\left(\sum_{i=1}^n \frac{q_i}{p_i}p_i\right) = f(1) = 0. \quad \square$$

$\boxed{\text{注意}}$ $D(q\|p) = \sum_{i=1}^n q_i \log\frac{q_i}{p_i}$ は確率分布 p と q のちがいの大きさを測る量としてよく使われ, Kullback-Leibler 情報量と呼ばれる. $D(q\|p)$ は p と q について対称ではないが, Sanovの定理という「KL情報量は p による q のシミュレーションの誤差の大きさと解釈できること」を意味する良い性質を持つ. 後で説明する「距離」の性質をみたさないが距離のように使われるものとして非常に重要である. $\square$