

# 確率数理工学7

## 変数変換

$$x \sim f(x) \quad (\text{p.d.f.}) \quad (x \in \mathbb{R})$$

$$\begin{cases} y = \varphi(x) \\ x = \psi(y) \quad (\psi = \varphi^{-1}) \end{cases} \quad (y \in \mathbb{R}) \quad x \text{ と } y \text{ は 一対一.}$$

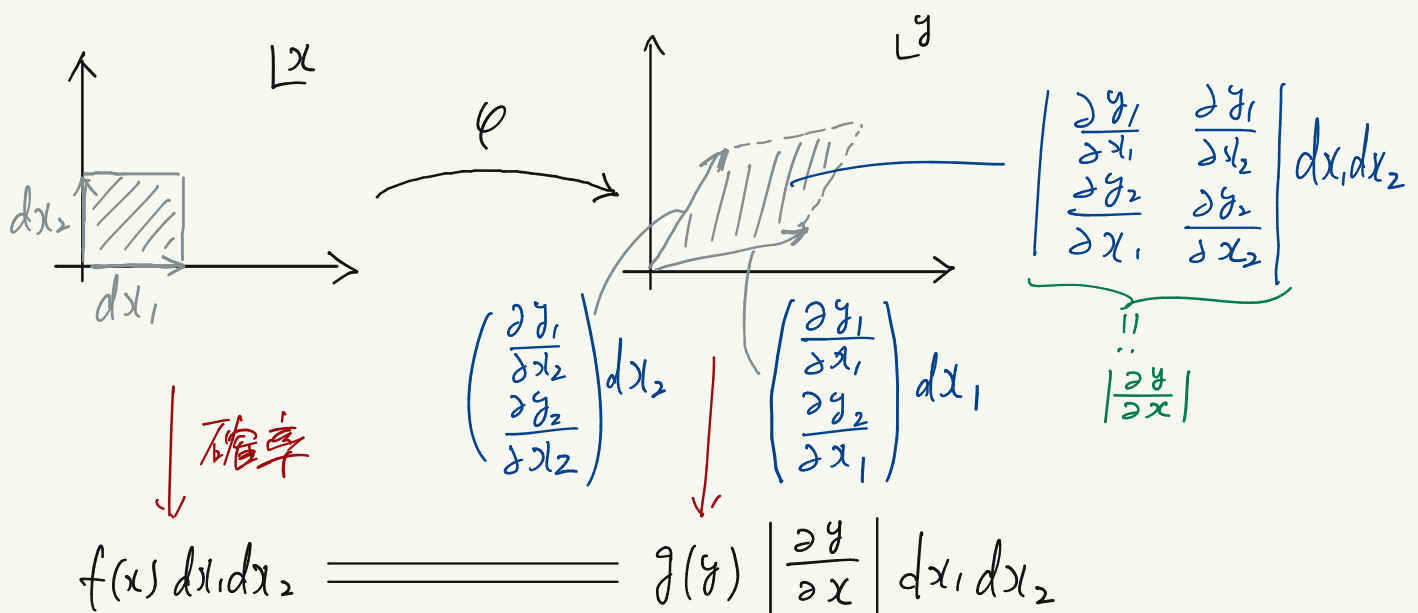
のとき、 $y$  の従う分布の p.d.f. は?

$$A: \quad \underbrace{g(y)}_{y \text{ の p.d.f.}} = \underbrace{f(\psi(y))}_{x \text{ の p.d.f.}} \underbrace{|\psi'(y)|}_{\text{変数変換によるスケールの調整}}$$

考え方:  $f(x) dx = f(x) \left| \frac{dx}{dy} \right| dy$

$$= f(x) |\psi'(y)| dy$$

$$= f(\psi(y)) |\psi'(y)| dy$$



$$\Rightarrow \underline{g(y) = \frac{1}{\left| \frac{\partial y}{\partial x} \right|} f(x) = \left| \frac{\partial x}{\partial y} \right| f(x)} \quad \leftarrow \text{逆変換}$$

$x, y \sim h(x, y)$  (同時分布の p.d.f.)

$$\begin{cases} u = \psi_1(x, y) \\ v = \psi_2(x, y) \end{cases}, \quad \begin{cases} x = \gamma_1(u, v) \\ y = \gamma_2(u, v) \end{cases} \quad \text{と書ける} \quad \text{と書ける} \quad \text{と書ける}$$

( $\psi, \gamma$  は 1対1対応に 微分可能)

$$\Rightarrow p(u, v) = h(\gamma_1(u, v), \gamma_2(u, v)) \left| \frac{\partial(x, y)}{\partial(u, v)} \right|$$

↑  
(u, v) の p.d.f.

← Jacobian

$$\text{f.t.L.} \quad \left| \frac{\partial(x, y)}{\partial(u, v)} \right| = \left| \det \begin{pmatrix} \frac{\partial \gamma_1}{\partial u} & \frac{\partial \gamma_1}{\partial v} \\ \frac{\partial \gamma_2}{\partial u} & \frac{\partial \gamma_2}{\partial v} \end{pmatrix} \right|$$

Ex.  $\begin{cases} X \sim N(0, 1) \\ Y \sim G(\frac{1}{2}, \frac{n}{2}) \end{cases}$  --- 自由度  $n$  の  $\chi^2$  分布  $\uparrow$  独立

$\rightarrow \frac{X}{\sqrt{Y/n}}$  の分布は? (検定統計量としてよく現れる)

例:  $X_1, \dots, X_n$  の平均は  $\mu$  否? ( $X_i \sim$  正規分布)

$$\frac{\frac{X_1 + \dots + X_n}{n} - \mu}{\sqrt{\frac{1}{n} \sum_{i=1}^n (X_i - \mu)^2}} \leftarrow N(0, 1) \cdot \sigma$$

$\leftarrow G(\frac{1}{2}, \frac{n}{2}) \cdot \sigma^2$   
in 帰無仮説

$$\begin{cases} u = \frac{x}{\sqrt{y/n}} \\ v = y \end{cases} \iff \begin{cases} x = u \sqrt{\frac{v}{n}} = \gamma_1(u, v) \\ y = v = \gamma_2(u, v) \end{cases}$$

$$h(x, y) = f(u)g(v) = \underbrace{c}_{\text{定数}} \cdot \exp\left(-\frac{x^2}{2}\right) \cdot y^{\frac{n}{2}-1} \exp\left(-\frac{y}{2}\right)$$

$$\left| \frac{\partial(x, y)}{\partial(u, v)} \right| = \left| \begin{pmatrix} \frac{\sqrt{v/n}}{2} & 0 \\ \frac{u}{2\sqrt{n}} & 1 \end{pmatrix} \right| = \sqrt{\frac{v}{n}} \quad \neq 1$$

$$\begin{aligned} p(u, v) &= h(\gamma_1(u, v), \gamma_2(u, v)) \sqrt{\frac{v}{n}} \\ &\cong \exp\left(-\frac{1}{2} u^2 \frac{v}{n}\right) v^{\frac{n}{2}-1} \exp\left(-\frac{v}{2}\right) \cdot v^{\frac{1}{2}} \\ &= v^{\frac{n+1}{2}-1} \exp\left(-\frac{v}{2} \left(1 + \frac{u^2}{n}\right)\right) \end{aligned}$$

よって  $p(u) = \int_0^{\infty} p(u, v) dv$   $\epsilon$  計算可能な  $u$ . ( $v = y \geq 0$  (a.s.))  
注意

$$p(u) = \int_0^{\infty} p(u,v) dv$$

$$\cong \left(1 + \frac{u^2}{n}\right)^{-\frac{n+1}{2}} \quad (\because \text{ガンマ分布の積分})$$

$$\int_{-\infty}^{\infty} p(u) du = 1 \quad \text{より}$$

$$p(u) = \frac{\Gamma(\frac{n+1}{2})}{\sqrt{n\pi} \Gamma(\frac{n}{2})} \left(1 + \frac{u^2}{n}\right)^{-\frac{n+1}{2}} : \text{自由度 } n \text{ の } t\text{-分布}$$

Ex.  $\begin{cases} X \sim N(0,1) \\ Y \sim N(0,1) \end{cases}$

$U = \frac{X}{Y}$  の分布は?  $\rightarrow p(u) = \frac{1}{\pi} \frac{1}{1+u^2} : \text{Cauchy分布}$

(上の例の  $n=1$  に対応)

畳み込み分布

絶対連続な場合

$$\left. \begin{matrix} X_1 \sim f_1(x_1) \\ X_2 \sim f_2(x_2) \end{matrix} \right\} \text{互に独立}$$

$Z = X_1 + X_2$  の分布は? 特にその p.d.f. は?

$$\begin{cases} z = x_1 + x_2 \\ \omega = x_2 \end{cases} \iff \begin{cases} x_1 = z - \omega \\ x_2 = \omega \end{cases}$$

$$\rightarrow \left| \frac{\partial(x_1, x_2)}{\partial(z, \omega)} \right| = \left| \begin{pmatrix} 1 & 0 \\ -1 & 1 \end{pmatrix} \right| = 1 \quad \text{より}$$

$$p(z, \omega) = f_1(x_1) f_2(x_2) \cdot 1 = f_1(z - \omega) f_2(\omega)$$

$$p(z) = \int_{-\infty}^{\infty} p(z, \omega) d\omega = \int_{-\infty}^{\infty} f_1(z - \omega) f_2(\omega) d\omega$$

畳み込み

Def (畳み込み分布)

密度  $f_1, f_2$  の畳み込み分布 (convolution) は密度関数

$$(f_1 * f_2)(z) \stackrel{\text{def}}{=} \int_{-\infty}^{\infty} f_1(z - \omega) f_2(\omega) d\omega$$

と与えられる。

## 離散の場合

$$p(z) = \sum_k f_1(z-k) f_2(k) = (f_1 * f_2)(z)$$

↑  
zのp.m.f.

## 一般の場合

分布関数:

$$F(z) = F_1 * F_2(z) = \int F_1(z-w) F_2(dw)$$

(証明) 特性関数  $\phi(t) = \int e^{itx} F(dz)$

↑  $p(x) = p_1(x-w)$  による積分

$$\begin{aligned} &= \iint e^{itx} dF_1(z-w) dF_2(w) \\ &= \int e^{itw} \underbrace{\int e^{it(z-w)} dF_1(z-w)}_{\phi_1(t)} dF_2(dw) \quad (F_{2.b.i.}) \\ &= \phi_1(t) \int e^{itw} dF_2(w) \\ &= \phi_1(t) \phi_2(t) \end{aligned}$$

よって特性関数と分布の一对一対応から従う。

同様に  $Z = X_1 + \dots + X_n$  の分布関数は

$$F = F_1 * \dots * F_n$$

と与えられる。

Ex. 二項分布の再生性

$$\begin{cases} X \sim B(n, \theta) \\ Y \sim B(m, \theta) \end{cases} \text{独立} \Rightarrow Z = X + Y \sim B(n+m, \theta)$$

Ex. Poisson分布

$$\begin{cases} X \sim P_0(\lambda_1) \\ Y \sim P_0(\lambda_2) \end{cases} \text{独立} \Rightarrow Z = X + Y \sim P_0(\lambda_1 + \lambda_2)$$

Ex. 正規分布

$$\begin{cases} X \sim N(\mu_1, \sigma_1^2) \\ Y \sim N(\mu_2, \sigma_2^2) \end{cases} \text{独立} \Rightarrow Z = X + Y \sim N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$$

Ex. ガウス分布

$$\begin{cases} X \sim G(\lambda, k_1) \\ Y \sim G(\lambda, k_2) \end{cases} \text{独立} \Rightarrow Z = X + Y \sim G(\lambda, k_1 + k_2)$$

# 確率評価に関する不等式

標本平均の期待値からのずれ、確率変数の大まか・収束を論じるための不等式を紹介

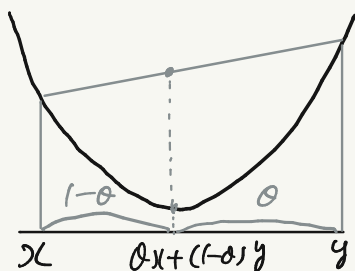
・期待値に関する不等式

Def (凸関数)

$f: \mathbb{R}^d \rightarrow \mathbb{R}$  の凸関数

def  $\forall x, y \in \mathbb{R}^d, 0 \leq \theta \leq 1$  に対し

$$f(\theta x + (1-\theta)y) \leq \theta f(x) + (1-\theta)f(y)$$



Ex.  $\exp(x), -\log(x), (x-a)^2, |x-a|, \log(1+\exp(-x))$  : 凸関数

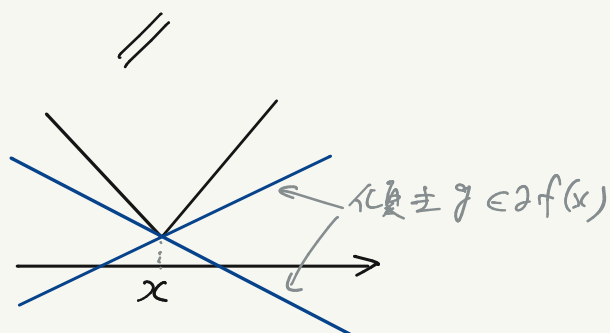
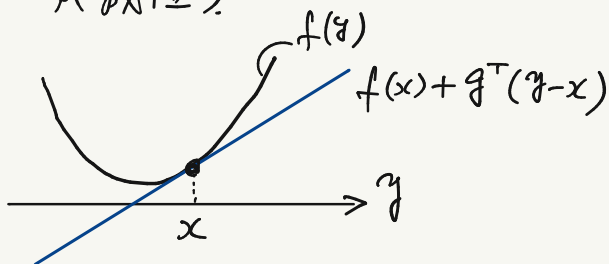
$\ln, \log(x)$  : 非凸関数

Prop  $f$  : 凸関数 ( $f: \mathbb{R}^d \rightarrow \mathbb{R}$ )

$\forall x \in \mathbb{R}^d$  に対し、ある  $g \in \mathbb{R}^d$  が存在し

$$f(y) \geq f(x) + g^T(y-x) \quad (\forall y \in \mathbb{R}^d)$$

⇔ 成り立つ



$$\partial f(x) = \{g \in \mathbb{R}^d \mid \forall y, f(y) \geq f(x) + g^T(y-x)\}$$

⇒ 劣微分 と言ふ ←  $\forall x \in \mathbb{R}^d, \partial f(x) \neq \emptyset$  (演習問題)

$\partial f(x)$  の元を 劣勾配 と言ふ

$f$  が微分可能ならば

$g = \nabla f(x)$  が  $\partial f(x)$  の唯一の元。

Thm (Jensenの不等式) ~~重~~ ~~要~~

$f: \mathbb{R}^d \rightarrow \mathbb{R}$   $\varepsilon$  凸関数 とある.

$X: \mathbb{R}^d$  値-r.v., 可積分,  $f(x)$  も可積分 から.

$$f(E[X]) \leq E[f(X)] //$$

(証明)  $g \in \partial f(E[X])$  とある.

すなわち  $\forall y \in \mathbb{R}^d$  に対し.

$$f(y) \geq f(E[X]) + g^T(y - E[X])$$

とある.  $y$  に  $E[X]$  期待値を取れば.

期待値の単調性より

$$E[f(X)] \geq f(E[X]) + \overbrace{g^T(E[X] - E[X])}^{=0} = f(E[X]) //$$

Cor (Youngの不等式)

$r \geq 1, s \geq 1$  かつ

$$\frac{1}{r} + \frac{1}{s} = 1$$

$a, b \in \mathbb{R}$  ならば.

$$|ab| \leq \frac{1}{r}|a|^r + \frac{1}{s}|b|^s$$

← 関数解析では  
よく出てくる.  
e.g. Sobolev 空間  
Besov 空間

←  $r=1$  のとき  $s=\infty$

Proof

$$-\log(|ab|) = -\log(|a|^{\frac{r}{s}} |b|^{\frac{s}{r}})$$

$$= -\frac{1}{r} \log(|a|^r) - \frac{1}{s} \log(|b|^s)$$

$$\geq -\log\left(\frac{1}{r}|a|^r + \frac{1}{s}|b|^s\right) \quad (\because -\log \text{ は凸}) //$$

$r=s=2$  のとき. 相加相乗平均の不等式 になる.

$$|ab| \leq \frac{1}{2}(|a|^2 + |b|^2)$$

(1) Hölder の不等式

$X, Y: r.v. \quad r \geq 1, s \geq 1, \frac{1}{r} + \frac{1}{s} = 1$  とおす.

$$E[|XY|] \leq E[|X|^r]^{\frac{1}{r}} E[|Y|^s]^{\frac{1}{s}}$$

$$(\quad = \|X\|_r \cdot \|Y\|_s)$$

(2) Minkowski の不等式

$r \geq 1$  とおす.

$$E[|X+Y|^r]^{\frac{1}{r}} \leq (E[|X|^r])^{\frac{1}{r}} + (E[|Y|^r])^{\frac{1}{r}}$$

$$(\quad \|X+Y\|_r \leq \|X\|_r + \|Y\|_r)$$

(1/3 ZEM)

(1) Young の不等式よりある  $\tilde{X}, \tilde{Y}$  に対し.

$$|\tilde{X}\tilde{Y}| \leq \frac{1}{r} |\tilde{X}|^r + \frac{1}{s} |\tilde{Y}|^s$$

とある. 両辺期待値を取れば.

$$E[|\tilde{X}\tilde{Y}|] \leq \frac{1}{r} E[|\tilde{X}|^r] + \frac{1}{s} E[|\tilde{Y}|^s].$$

$$\text{ここで, } \tilde{X} = \frac{X}{(E[|X|^r])^{\frac{1}{r}}}, \quad \tilde{Y} = \frac{Y}{(E[|Y|^s])^{\frac{1}{s}}} \quad \text{とおく.}$$

$$\frac{E[|XY|]}{E[|X|^r]^{\frac{1}{r}} E[|Y|^s]^{\frac{1}{s}}} \leq \frac{1}{r} \frac{E[|X|^r]}{E[|X|^r]} + \frac{1}{s} \frac{E[|Y|^s]}{E[|Y|^s]}$$

$$= \frac{1}{r} + \frac{1}{s} = 1 \quad //$$

$$(2) \quad E[|X+Y|^r] = E[|X+Y| \cdot |X+Y|^{r-1}]$$

$$\leq E[|X| \cdot |X+Y|^{r-1}] + E[|Y| \cdot |X+Y|^{r-1}]$$

$$\leq E[|X|^r]^{\frac{1}{r}} \cdot E[|X+Y|^{(r-1)s}]^{\frac{1}{s}}$$

$$+ E[|Y|^s]^{\frac{1}{s}} E[|X+Y|^{(r-1)r}]^{\frac{1}{r}} \quad (\text{Young 式})$$

$$(\text{ここで, } \frac{1}{s} + \frac{1}{r} = 1, \text{ かつ } s = \frac{r}{r-1} \text{ とおす})$$

$$= (E[|X|^r]^{\frac{1}{r}} + E[|Y|^s]^{\frac{1}{s}}) \cdot (E[|X+Y|^r])^{\frac{r-1}{r}}$$

$$\Rightarrow E[|X+Y|^r]^{\frac{1}{r}} \leq E[|X|^r]^{\frac{1}{r}} + E[|Y|^s]^{\frac{1}{s}} \quad //$$

Young の不等式より  $E[|XY|] \leq \frac{1}{r} \|X\|_r^r + \frac{1}{s} \|Y\|_s^s$  とある.

-  $\|X\|_p := E[|X|^p]^{1/p}$  ( $p \geq 1$ ) と書くと. Minkowski の不等式が)

= 似た三角不等式を満たす.  $\Rightarrow$   $L^p$ -ノルム と言う

- Hölder の不等式から  $\|\cdot\|_r$  と  $\|\cdot\|_s$  は互いに "双対"

$$\langle X, Y \rangle = E[XY] \text{ と書くと.}$$

$$\langle X, Y \rangle \leq \|X\|_r \|Y\|_s \text{ かつ } \|X\|_r = \sup_{\|Y\|_s \leq 1} \langle X, Y \rangle$$

$$\left( Y = \frac{\text{sign}(X) \cdot |X|^{r-1}}{\|X\|_r^{r-1}} \text{ とすれば } = \text{成り立つ} \right)$$

-  $\|X\|_p < \infty$  である r.v.  $X$  を  $L^p$ -可積分と言う

$L^p$ -可積分な r.v. 全体を  $L^p(P)$  と書く.

Eq 正確には

$L^p$  は高次元空間をなす.

$$(X=0 \Leftrightarrow \|X\|_p=0)$$

$L^p$  の元  $X, Y$  について  $\|X-Y\|_p = 0$  なる r.v. を同一視する.

$\|\cdot\|_p$  は  $L^p$  のノルムになる. したがって  $L^p$  は  $\|\cdot\|_p$  による完備空間であることも示せる.

つまり  $(X_n)_n$  が  $L^p$  内の  $\mathcal{F}$ -数列 (任意  $\epsilon > 0$ ,  $\exists N \in \mathbb{N}$  について  $\|X_n - X_m\|_p \leq \epsilon$  ( $n, m \geq N$ ))

ならば 必ず  $X \in L^p$  が存在して  $\|X_n - X\|_p \rightarrow 0$  である. (非自明) ← 次回の演習問題

に似た.  $L^p$  は Banach 空間になる.  $\Rightarrow$   $L^p$ -空間と言う. 出題予定

-  $L^2$ -空間は  $\langle X, Y \rangle$  を内積とした Hilbert 空間になる.

- 尤も  $\|X-Y\|_p = 0$  ならば  $X=Y$  (a.s.) である. (Markov の不等式が)

↓ 次回.

### Cor (Schwarz の不等式)

$$E[|XY|] \leq \|X\|_2 \cdot \|Y\|_2$$

$$\text{特に } \|X+Y\|_2 \leq \|X\|_2 + \|Y\|_2$$

Ex. Kullback-Leibler divergence (KL-ダイバージェンス, 相対エントロピー)

$P, Q$ : 2つの p.d.f.

$$D(P||Q) := \int p(x) \log\left(\frac{p(x)}{q(x)}\right) dx : \text{KL-divergence}$$

$\Rightarrow D(P||Q) \geq 0$  かつ  $D(P||Q) = 0 \Leftrightarrow P=Q$  である.

$$(\because \text{ Jensen が }) D(P||Q) = \int p(x) \left[ -\log\left(\frac{q(x)}{p(x)}\right) \right] dx \geq -\log\left(\int p(x) \frac{q(x)}{p(x)} dx\right) = -\log(1) = 0$$

$$= \text{ かつ } \frac{p(x)}{q(x)} = 1 \text{ (a.e.) ならば } P=Q \text{ (a.e.)}$$



参考: KL-div  $\varepsilon$  一般化して,  $f(1)=0$  なる凸関数を用いた

$$D_f(p||q) := \int p(x) f\left(\frac{q(x)}{p(x)}\right) dx$$

$\varepsilon$   $f$ -ダイバージェンスと書く。

例として:  $f(u) = -(u+1) \log\left(\frac{1+u}{2}\right) + u \log u$  を用いる。

$$D_f(p||q) = \frac{1}{2} \int p(x) \log\left(\frac{2p(x)}{p(x)+q(x)}\right) + q(x) \log\left(\frac{2q(x)}{p(x)+q(x)}\right) dx$$

$\varepsilon$  Jensen-Shannon エントロピーと書く。情報学習における GAN に  
用いられる。

# 演習問題

(1) 二項分布, ポアソン分布, 正規分布, カンマ分布の再生性を  
畳み込みを用いて示せ.

(2)  $X, Y$  が独立  $\Leftrightarrow$  連続分布であるとき  $P(X+Y=0)=0$  を示せ.

(3)  $X, Y$  が独立なとき. 可測な  $f, g$  に対し  $f(X), g(Y)$  が  
独立であることを示せ.

(4) 2つの独立同一分布に従う連続分布  
その和の p.d.f. を畳み込みを用いて計算せよ. (難)

(5)  $X, Y$  が独立にカンマ分布  $G(\lambda, \alpha), G(\lambda, \beta)$  に従うとき.

$$Z = \frac{X}{X+Y}$$

↑  
λは共通

の従う分布の p.d.f. を求めよ.

(6)  $f: \mathbb{R}^d \rightarrow \mathbb{R}$  を凸関数とする.

←  $f: \mathbb{R}^d \rightarrow \mathbb{R} \cup \{\infty\}$  とするとき注意

(i)  $\text{epi}(f) = \{(x, \omega) \in \mathbb{R}^{d+1} \mid f(x) \leq \omega\}$

が閉凸集合であることを示せ. なお  $f$  が連続関数にすぎない場合はよく.

(ii) 非空な閉凸集合  $H \in \mathbb{R}^{d+1}$  に対し, 以下が成り立つ

「 $H$  の境界上の点  $z_0 \in \partial H$  に対し, ある  $a \in \mathbb{R}^{d+1}$  ( $a \neq 0$ ) が存在して,

$$a^T z_0 \leq \inf_{z \in H} a^T z$$

が成り立つ」

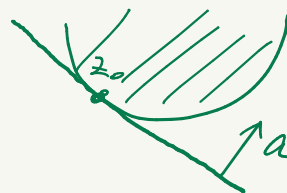
(支持超平面の定理, 分離定理)

これを (i) を用いて. 任意の  $x \in \mathbb{R}^d$  に対して.

ある正の実数  $\beta$  と  $g \in \mathbb{R}^{d+1}$  が存在して.

$$g^T x + \beta f(x) \leq g^T x' + \beta \omega \quad (\forall (x', \omega) \in \text{epi}(f))$$

が成り立つことを示し.  $\partial f(x)$  が非空であることを示せ.



(7)  $0 < \alpha < \beta < \infty$  に対し.

$$\|X\|_\alpha \leq \|X\|_\beta \quad (\alpha, \beta < 1 \text{ ではない})$$

を示せ.

(8)  $X \in L^2$  のとき,  $X \in L^1$  であることを示せ.

まず, 明らかに,  $\mu = E[X]$  に対し.

$$E[(X-\mu)^2] \leq E[(X-a)^2] \quad (a \in \mathbb{R})$$

を示せ.

(9)  $X_n \in L^1$  ( $n$ ) とおす. まず,  $\sup_n E[X_n] < \infty$  とす.

$X_n \nearrow X$  であることを,  $X \in L^1$  であることを, 特に

$$E[X_n] \nearrow E[X]$$

を示せ.

(10)  $X \geq 0, Y \geq 0$  かつ,  $p \geq 1$  に対し.

$$E[(X+Y)^p] \leq 2^{p-1} (E[X^p] + E[Y^p])$$

を示せ.