確率数理工学7

③ 変数変換

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

$$\begin{cases} y = \varphi(x) & (y \in \mathbb{R}) \\ x = \psi(y) & (\psi = \varphi^{-1}) \end{cases}$$

xとyは一対一.

のとき. yの従う分布のp.d.f.は?

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

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

$$\begin{pmatrix} \frac{\partial y_1}{\partial x_2} \\ \frac{\partial y_2}{\partial x_2} \end{pmatrix} dx_2 \qquad \begin{pmatrix} \frac{\partial y_1}{\partial x_1} \\ \frac{\partial y_2}{\partial x_1} \end{pmatrix} dx_1$$

$$\underbrace{\begin{vmatrix} \frac{\partial y_1}{\partial x_1} & \frac{\partial y_1}{\partial x_2} \\ \frac{\partial y_2}{\partial x_1} & \frac{\partial y_2}{\partial x_2} \end{vmatrix}}_{= \left|\frac{\partial y}{\partial x}\right|} dx_1 dx_2$$

確率

$$f(x)\,dx_1 dx_2 = g(y) \left|\frac{\partial y}{\partial x}\right| dx_1 dx_2$$

$$\Longrightarrow \quad g(y) = \frac{1}{\left|\frac{\partial y}{\partial x}\right|} f(x) = \left|\frac{\partial x}{\partial y}\right| f(x)$$ ← チェックせよ.

$x, y \sim h(x,y)$ （同時分布の p.d.f.）

$$\begin{cases} u = \varphi_1(x,y) \\ v = \varphi_2(x,y) \end{cases}, \quad \begin{cases} x = \psi_1(u,v) \\ y = \psi_2(u,v) \end{cases}$$ と書けるとき

（φ, ψ は 1対1対応かつ微分可能）

$$\Longrightarrow \ p(u,v) = h(\psi_1(u,v), \psi_2(u,v)) \left| \frac{\partial(x,y)}{\partial(u,v)} \right|$$

(u,v) の p.d.f.

Jacobian

ただし、$$\left| \frac{\partial(x,y)}{\partial(u,v)} \right| = \left| \det \begin{pmatrix} \frac{\partial \psi_1}{\partial u} & \frac{\partial \psi_1}{\partial v} \\ \frac{\partial \psi_2}{\partial u} & \frac{\partial \psi_2}{\partial v} \end{pmatrix} \right|$$

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

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

例：$X_1, \ldots, X_n$ の平均は μ か？（$X_i \sim$ 正規分布）

$$\frac{\frac{X_1 + \cdots + X_n}{n} - \mu}{\sqrt{\frac{1}{n}\sum_{i=1}^{n}(X_i - \hat{\mu})^2}}$$

$\leftarrow N(0,1)\cdot\sigma$

$\leftarrow G(\frac{1}{2}, \frac{n-1}{2})\cdot\sigma^2$

in 帰無仮説

$$\begin{cases} u = \dfrac{x}{\sqrt{y/n}} \\ v = y \end{cases} \Longleftrightarrow \begin{cases} x = u\sqrt{\dfrac{v}{n}} = \psi_1(u,v) \\ y = v = \psi_2(u,v) \end{cases}$$

$$h(x,y) = f(x)g(y) = C \cdot \exp\left(-\frac{x^2}{2}\right) \cdot y^{\frac{n}{2}-1} \exp\left(-\frac{y}{2}\right)$$

C：定数

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

$$p(u,v) = h(\psi_1(u,v), \psi_2(u,v)) \sqrt{\frac{v}{n}}$$

$$\cong \exp\left(-\frac{1}{2}u^2 \frac{v}{n}\right) v^{\frac{n}{2}-1} \exp\left(-\frac{v}{2}\right) \cdot v^{\frac{1}{2}}$$

$$= v^{\frac{n+1}{2}-1} \exp\left(-\frac{v}{2}\left(1 + \frac{u^2}{n}\right)\right)$$

あとは $p(u) = \int_0^\infty p(u,v)\,dv$ を計算すれば良い。（$v = y \geq 0$ (a.s.) に注意）

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

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

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

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

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

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

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

$Z = X_1 + X_2$ の分布は？ 特にその p.d.f. は？

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

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

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

$$p(z) = \int_{-\infty}^{\infty} p(z,w)\,dw = \underbrace{\int_{-\infty}^{\infty} f_1(z-w) f_2(w)\,dw}_{\text{畳み込み}}$$

Def（畳み込み分布）

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

$$(f_1 * f_2)(z) \overset{\text{def}}{=} \int_{-\infty}^{\infty} f_1(z-w) f_2(w)\,dw$$

で与えられる。 //

離散の場合

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

$\uparrow$ 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 $\phi(t) = \int e^{itz} F(dz)$

$$= \iint e^{itz} dF_1(z-w) dF_2(dw)$$

$$= \int e^{itw} \underbrace{\int e^{it(z-w)} dF_1(z-w)}_{= \phi_1(t)} dF_2(dw) \quad \text{(Fubini)}$$

$$= \phi_1(t) \int e^{itw} dF_2(w)$$

$$= \phi_1(t) \phi_2(t)$$

あとは特性関数と分布の一対一対応から従う。 //

同様にして $Z = X_1 + \cdots + X_n$ の分布関数は

$$F = F_1 * \cdots * F_n$$

として与えられる。

Ex. 二項分布の再生性

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

Ex. Poisson分布

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

Ex. 正規分布

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

Ex. ガンマ分布

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

確率評価に関する不等式

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

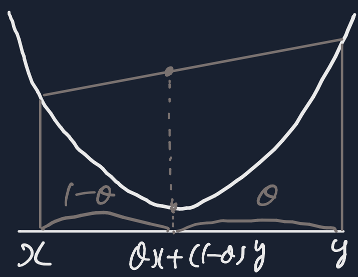
・期待値に関する不等式

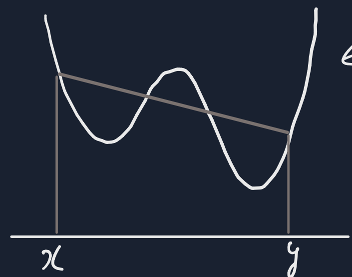
Def（凸関数）

$f : \mathbb{R}^d \to \mathbb{R}$ が凸関数

$\overset{\text{def}}{\iff}$ $^\forall x, y \in \mathbb{R}^d$, $0 \le {}^\forall\theta \le 1$ に対し、

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





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

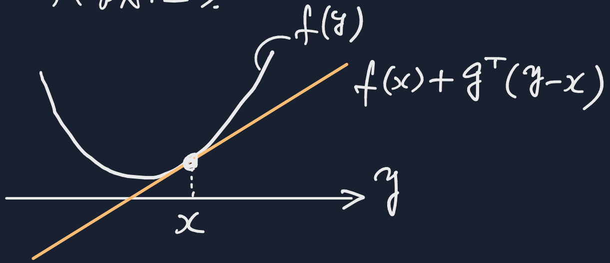
$\sqrt{x}$, $\log(x)$: 非凸関数

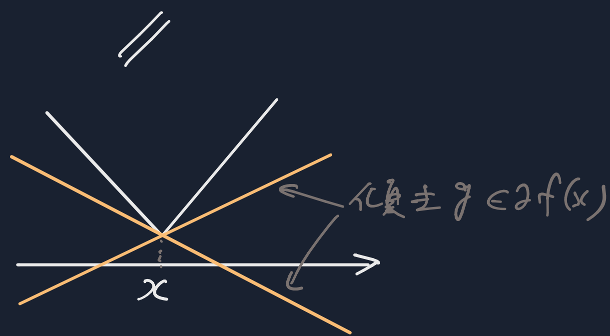
Prop f : 凸関数 $(f : \mathbb{R}^d \to \mathbb{R})$

$^\forall x \in \mathbb{R}^d$ において、ある $g \in \mathbb{R}^d$ が存在して、

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

が成り立つ





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

⟹ 劣微分 と言う ← ※ $^\forall x \in \mathbb{R}^d$ で $\partial f(x) \ne \phi$ （演習問題）

$\partial f(x)$ の元を劣勾配と言う

f が微分可能なら

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

Thm (Jensenの不等式) ㊙ 重要

$f: \mathbb{R}^d \to \mathbb{R}$ を凸関数とする.

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

$$f(E[X]) \leqq E[f(X)]$$

//

(証明) $g \in \partial f(E[X])$ とする.

すると, $\forall y \in \mathbb{R}^d$ に対し,

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

である. y について期待値を取れば,

期待値の単調性より

$$E[f(X)] \geqq f(E[X]) + g^T\underbrace{(E[X] - E[X])}_{=0} = f(E[X])$$

//

Cor (Youngの不等式)

$r \geqq 1$, $s \geqq 1$ が

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

をみたすなら,

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

関数解析でもよく出てきます.
e.g. Sobolev 空間
Besov 空間

Proof

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

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

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

//

$r = s = 2$ のとき, 相加相乗平均の不等式になる.

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

Thm

(1) Hölder の不等式

X, Y : r.v. $r \geqq 1, s \geqq 1, \frac{1}{r}+\frac{1}{s}=1$ とする。

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

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

(2) Minkowski の不等式

$r \geqq 1$ のとき。

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

$$(\|X+Y\|_r \leqq \|X\|_r + \|Y\|_r)$$

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

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

である。両辺期待値を取れば

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

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

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

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

//

(2)

$$\begin{aligned}
E[|X+Y|^r] &= E[|X+Y| \cdot |X+Y|^{r-1}] \\
&\leqq E[|X| \cdot |X+Y|^{r-1}] + E[|Y| \cdot |X+Y|^{r-1}] \\
&\leqq E[|X|^r]^{\frac{1}{r}} \cdot E[|X+Y|^{(r-1)s}]^{\frac{1}{s}} \\
&\quad + E[|Y|^r]^{\frac{1}{r}} E[|X+Y|^{(r-1)s}]^{\frac{1}{s}} \quad (\text{Young より})
\end{aligned}$$

(ただし、$\frac{1}{s}+\frac{1}{r}=1$, つまり $s=\frac{r}{r-1}$ とする)

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

$$\Longrightarrow E[|X+Y|^r]^{\frac{1}{r}} \leqq E[|X|^r]^{\frac{1}{r}} + E[|Y|^r]^{\frac{1}{r}}$$

//

Young の不等式より $E[|XY|] \leqq \frac{1}{r}\|X\|_r^r + \frac{1}{s}\|Y\|_s^s$ でもある。

- $\|X\|_p := E[|X|^p]^{\frac{1}{p}}$ $(p \geqq 1)$ と書くと、Minkowski の不等式より
  これは 3角不等式を満たす。$\Longrightarrow$ <u>L^p-ノルムと言う</u>

- Hölder の不等式から $\|\cdot\|_r$ と $\|\cdot\|_s$ は 互いに "双対"
  $\langle X, Y \rangle = E[XY]$ と書くと、
  $\langle X, Y \rangle \leqq \|X\|_r \|Y\|_s$ かつ $\|X\|_r = \sup_{\|Y\|_s \leqq 1} \langle X, Y \rangle$

  $\left(Y = \frac{\mathrm{sign}(X) \cdot |X|^{r-1}}{\|X\|_r^{r-1}} \right.$ とすれば $=$ をみたす$\left.\right)$

- $\|X\|_p < \infty$ 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。
  L^p の元 X, Y で $\|X - Y\|_p = 0$ なる r.v. を同一視することで、
  $\|\cdot\|_p$ は L^p のノルムになる。さらに、L^p は $\|\cdot\|_p$ について完備であることも示せる。
  つまり $(X_n)_n$ が L^p 内のコーシー列 $(\forall \varepsilon > 0, \exists N \in \mathbb{N}$ で、$\|X_n - X_m\|_p \leqq \varepsilon$ $(\forall n, m \geqq N))$
  なら、ある $X \in L^p$ が存在して、$\|X_n - X\|_p \longrightarrow 0$ である。(非自明) ← 次回の演習問題で出題予定
  これより、L^p は Banach 空間になる。$\Longrightarrow$ L^p-空間と言う

  より正確には $L^p/\sim$ なる商空間を考える。$(X \sim Y \Leftrightarrow \|X - Y\|_p = 0)$

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

- もっとも $\|X - Y\|_p = 0$ なら $X = Y$ (a.s.) である。(Markov の不等式より)
  ↳ 次回。

<u>Cor</u> (Schwartz の不等式)

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

特に $\|X + Y\|_2 \leqq \|X\|_2 + \|Y\|_2$

<u>Ex.</u> 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.

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

$\Longrightarrow$ $D(p \| q) \geqq 0$ かつ、$D(p \| q) = 0 \Longleftrightarrow p = q$ である。

$(\because)$ Jensen より $D(p \| q) = \int p(x) \left[-\log\left(\frac{q(x)}{p(x)}\right) \right] dx \geqq -\log\left(\int \cancel{p(x)} \frac{q(x)}{\cancel{p(x)}} dx \right) = -\log(1) = 0$

$=$ は $\frac{p(x)}{q(x)} = 1$ (a.s.) のとき、つまり $p = q$ (a.s.) //

参考: KL-divを一般化して、$f(1)=0$ なる凸関数を用いた

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

を f-ダイバージェンスと言う.

例えば: $f(u) = -(u+1)\log\left(\frac{1+u}{2}\right) + u\log u$ を用いると.

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

を Jensen-Shannon エントロピーと言い、深層学習における GAN に用いられている.