

鈴木讓「ベイジアンネットワーク入門」(培風館)

1. 確率論の基礎

1.4 Kullback-Leibler 情報量

鈴木讓

大阪大学

2010年5月6日

あらまし

- 1 期待値・分散
- 2 Kullback-Leibler 情報量
- 3 相互情報量

期待値・分散

 $(\Omega, \mathcal{F}, \mu)$ $X : \Omega \rightarrow \mathbb{R}$: 確率変数

$$\int_{x \in X(\Omega)} |x| \mu_X(dx) < \infty$$

期待値

$$E[X] := \int_{x \in X(\Omega)} x \mu_X(dx) = \int_{\omega \in \Omega} X(\omega) \mu(d\omega)$$

分散

$$V[X] := E[(X - E[X])^2] < \infty$$

$$V[X] = E[X^2 - 2XE[X] + E[X]^2] = E[X^2] - E[X]^2$$

期待値・分散 (続)

$$\text{(例)} f_X(x) := \frac{1}{\pi} \frac{1}{1+x^2}$$

$$\int_{-\infty}^{\infty} |x| f_X(x) dx = \lim_{t \rightarrow \infty} \frac{2}{\pi} \int_0^t \frac{x}{1+x^2} dx = \lim_{t \rightarrow \infty} \frac{1}{\pi} \log(1+t^2) = \infty$$

$$\text{(例)} X \sim N(\mu, \sigma^2)$$

$$\begin{aligned} \int_{-\infty}^{\infty} |x| f(x) dx &= \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{\infty} |x| \exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right] dx \\ &\leq |\mu| \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sigma} \exp^{-\frac{(x-\mu)^2}{2\sigma^2}} dx + \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sigma} |x-\mu| \exp^{-\frac{(x-\mu)^2}{2\sigma^2}} dx \\ &= |\mu| + \sqrt{\frac{2}{\pi}} \sigma < \infty \end{aligned}$$

期待値・分散 (続)

$$\begin{aligned}
 E[X] &= \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{\infty} x \exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right] dx \\
 &= \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{\infty} \sigma y \exp\left(-\frac{y^2}{2}\right) \sigma dy + \mu \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{\infty} \exp\left(-\frac{y^2}{2}\right) dy = \mu
 \end{aligned}$$

$$\begin{aligned}
 V[X] &= \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{\infty} (x-\mu)^2 \exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right] dx \\
 &= \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{\infty} \sigma^2 y^2 \exp\left(-\frac{y^2}{2}\right) \sigma dy = \sigma^2
 \end{aligned}$$

(例) $A \in \mathcal{F}$ の指示関数 I_A は確率変数になる (命題 1.4)。

$$E[I_A] = 1 \cdot \mu(A) + 0 \cdot \mu(\bar{A}) = \mu(A)$$

$$V[I_A] = E[I_A^2] - E[I_A]^2 = E[I_A] - E[I_A]^2 = \mu(A) - \mu(A)^2 = \mu(A)(1 - \mu(A))$$

期待値・分散 (続)

$X, Y : \Omega \rightarrow \mathbb{R}, a, b \in \mathbb{R}$

$$\begin{aligned} E[aX + bY] &= \int (ax + by)\mu_{XY}(dx, dy) \\ &= a \int x\mu_X(dx) + b \int y\mu_Y(dy) = aE[X] + bE[Y] \end{aligned}$$

X, Y : 独立のとき、

$$\begin{aligned} E[XY] &= \int xy\mu_{XY}(dx, dy) = \int xy\mu_X(dx)\mu_Y(dy) \\ &= \int x\mu_X(dx) \int y\mu_Y(dy) = E[X]E[Y]. \end{aligned}$$

$$E[(X - E[X])(Y - E[Y])] = E[X - E[X]]E[Y - E[Y]] = 0.$$

$$\begin{aligned} V(aX + bY) &= E[\{a(X - E[X]) + b(Y - E[Y])\}^2] \\ &= E[a^2(X - E[X])^2 + b^2(Y - E[Y])^2 + 2ab(X - E[X])(Y - E[Y])] \\ &= a^2V[X] + b^2V[Y] \end{aligned}$$

期待値・分散 (続)

(例) X_1, \dots, X_n : $E[X_i] = \mu$, $V[X_i] = \sigma^2$ なる独立な n 個の確率変数

$$E \left[\frac{1}{n} \sum_{i=1}^n X_i \right] = \frac{1}{n} \sum_{i=1}^n E[X_i] = \mu$$

$$V \left[\frac{1}{n} \sum_{i=1}^n X_i \right] = \frac{1}{n^2} \sum_{i=1}^n V[X_i] = \frac{\sigma^2}{n}$$

k 次モーメント

$E[|X|^k] < \infty$ のとき、 $E[X^k]$

$E[|X|^n] < \infty \implies E[|X|^k] < \infty$, $k = 1, \dots, n-1$ ($|x|^k \leq 1 + |x|^n$)

\mathcal{F} 上可測な関数の性質 \mathcal{F} 上可測な関数の性質

$g : \Omega \rightarrow \mathbb{R}$ が \mathcal{F} 上可測, $f : \mathbb{R} \rightarrow \mathbb{R}$ が連続

$\implies f \circ g : \Omega \rightarrow \mathbb{R}$ が \mathcal{F} 上可測

証明: 任意の閉区間は、可算個の閉区間の像である。実際、

$$\min f(x) = A, \max f(x) = B$$

$$\implies \exists a_1 \leq b_1 < a_2 \leq b_2 < \dots \in \mathbb{R} \text{ s.t.}$$

$$A = \min_{x \in [a_i, b_i]} f(x), B = \max_{x \in [a_i, b_i]} f(x), i = 1, 2, \dots$$

$$\{\omega \in \Omega \mid f(g(\omega)) \in [A, B]\} = \{\omega \in \Omega \mid g(\omega) \in \cup_{i=1}^{\infty} [a_i, b_i]\} \in \mathcal{F}$$

$$D \in \mathcal{B} \implies \{\omega \in \Omega \mid f(g(\omega)) \in D\} \in \mathcal{F}$$

Kullback-Leibler 情報量

 $(\Omega, \mathcal{F}, \mu)$ $\nu: \nu(\Omega) \leq 1$ を満足する測度, $\mu \ll \nu$ μ の ν に関する Kullback-Leibler 情報量

$$D(\mu||\nu) := \int_{\omega \in \Omega} \log \left(\frac{d\mu}{d\nu}(\omega) \right) \mu(d\omega)$$

$\frac{d\mu}{d\nu}$: \mathcal{F} 上可測より、 $\log\left(\frac{d\mu}{d\nu}\right)$: \mathcal{F} 上可測。 $z - 1 \geq \log z$, $z \in \mathbb{R}$ より、

$$\begin{aligned} D(\mu||\nu) &\geq \int_{\omega \in \Omega} (1 - 1/\frac{d\mu}{d\nu}(\omega)) \mu(d\omega) = 1 - \int_{\omega \in \Omega} \frac{d\nu}{d\mu}(\omega) \mu(d\omega) \\ &= 1 - \nu(\Omega) \geq 0 \end{aligned}$$

$$D(\mu_X||\nu_X) = \int_{x \in X(\Omega)} \log \left(\frac{d\mu_X}{d\nu_X}(x) \right) \mu_X(dx) \geq 0$$

Kullback-Leibler 情報量 (続)

(例) 確率空間 $(\Omega, \mathcal{F}, \mu)$ における離散確率変数 X について,
 $\mu_X(\{x\}) := \mu(\{\omega \in \Omega \mid X(\omega) = x\})$: X の確率測度,
 $\nu_X(\{x\}) := \nu(\{\omega \in \Omega \mid X(\omega) = x\})$: $\sum_{x \in X(\Omega)} \nu_X(\{x\}) \leq 1$ をなる測度

$$\text{support}(\mu_X) := \{x \in \mathbb{R} \mid \mu_X(\{x\}) > 0\}$$

$$\text{support}(\nu_X) := \{x \in \mathbb{R} \mid \nu_X(\{x\}) > 0\}$$

$$\mu_X \ll \nu_X \implies \text{support}(\mu_X) \subseteq \text{support}(\nu_X)$$

$$x \in \text{support}(\mu_X) \text{ に対して、} \frac{d\mu_X}{d\nu_X}(x) = \frac{\mu_X(\{x\})}{\nu_X(\{x\})}$$

$$D(\mu_X \parallel \nu_X) = \sum_{x \in \text{support}(\mu_X)} \mu_X(\{x\}) \log \frac{\mu_X(\{x\})}{\nu_X(\{x\})} \geq 0$$

Kullback-Leibler 情報量 (続)

λ_X : Lebesgue 測度

$\mu_X \ll \nu_X \ll \lambda_X$

$$f_X := \frac{d\mu_X}{d\lambda_X}, \quad g_X := \frac{d\nu_X}{d\lambda_X}$$

とおけば, 系 1.1 より, ν_X -測度 0 の点を除いて,

$$\frac{d\mu_X}{d\nu_X} = \frac{d\mu_X}{d\lambda_X} / \frac{d\nu_X}{d\lambda_X} = \frac{f_X}{g_X}$$

$$D(\mu_X || \nu_X) = \int_{x \in X(\Omega)} f_X(x) \log \frac{f_X(x)}{g_X(x)} \lambda_X(dx) \geq 0$$

相互情報量

 $(\Omega, \mathcal{F}, \mu)$ X, Y : 確率変数

- $\nu_{XY}(D, D') = \mu_X(D)\mu_Y(D')$ なる $\nu_{XY} : \mathcal{B}^2 \rightarrow [0, 1]$ は測度
- $\nu_{XY}(X(\Omega), Y(\Omega)) \leq 1$
- $\mu_{XY} \ll \nu_{XY}$

 X, Y の相互情報量

$$I(X, Y) := D(\mu_{XY} || \nu_{XY})$$

$$I(X, Y) \geq 0$$

$$I(X, Y) = I(Y, X)$$

相互情報量 (続)

$X(\Omega), Y(\Omega)$ が可算集合のとき

$$P_{XY}(x, y) := \mu_{XY}(\{x\}, \{y\})$$

$$P_X(x) := \mu_X(\{x\}), \quad P_Y(y) := \mu_Y(\{y\})$$

$$P_{X|Y}(x|y) := \frac{P_{XY}(x, y)}{P_Y(y)}$$

X, Y の相互情報量 $I(X, Y)$

$$I(X, Y) := \sum_{x \in X(\Omega)} \sum_{y \in Y(\Omega)} P_{XY}(x, y) \log \frac{P_{XY}(x, y)}{P_X(x)P_Y(y)} = I(Y, X)$$

エントロピー

X のエントロピー $H(X)$

$$H(X) := I(X, X) = \sum_{x \in X(\Omega)} -P_X(x) \log P_X(x) \geq 0$$

X, Y の同時エントロピー

$$H(XY) := \sum_{x \in X(\Omega)} \sum_{y \in Y(\Omega)} -P_{X,Y}(x, y) \log P_{X,Y}(x, y) \geq 0$$

Y のもとでの X の条件付エントロピー $H(X|Y)$

$$H(X|Y) := \sum_{x \in X(\Omega)} \sum_{y \in Y(\Omega)} -P_{XY}(x, y) \log P_{X|Y}(x|y) \geq 0$$

$$H(XY) = H(X) + H(Y|X), \quad I(X, Y) = H(X) - H(X|Y)$$