Chapter5 Measures of Information and Uncertainty

竹田航太

2021 年 5 月 16 日

目次

1 Variance, Information and Entropy

1.1 Variance

Definition 1.1. $(V, \|\cdot\|)$*: ノルム空間,* μ を V 上の確率測度で平均 $m \in V$ を持つとする．μ に従う確率変数 X の分散（*variance*）を次で定める．

$$\mathrm{Var}(\mu) = \int_V \|x-m\|^2 d\mu(x) = \mathbb{E}_{X\sim\mu}[\|X-m\|^2]$$

また，共分散作用素（*covariance operator*）

$$\mathrm{Cov}(\mu) = \mathbb{E}_{X\sim\mu}[(X-m)\otimes(X-m)]$$

Remark 1.2. $n \in \mathbb{N}$ に対して，$V = \mathbb{R}^n$ のとき $\mathrm{Cov}(\mu)$ は $n \times n$ の対称半正定値行列になる．よって，$\mathrm{Cov}(\mu)$ は n 個の非負の固有値 $\lambda_1 \geq \cdots \geq \lambda_n \geq 0$ をもち，対応する固有ベクトルを $v_1, v_2, \cdots, v_n$ とすると，v_1 の方向で確率変数 X の不確実性が強い．

不確実性を物理量に導入することによる量子力学における古典的な結果が次の不確実性原理である．

まず，量子力学における数学的設定を与える．適切な（$L^2(\mathbb{R}^n;\mathbb{C})$ など）Hilbert 空間 H の単位要素 ψ を用いて，確率分布は $p=|\psi|^2$ で書かれる．物理量は H 上のエルミート作用素として定式化される．

Definition 1.3. 物理量 $A: H \to H$ に対して，分布 $p=|\psi|^2$ に対する平均 $\langle A\rangle$ と分散 σ_A は次で与えられる．

$$\begin{aligned}\langle A\rangle &= \langle \psi, A\psi\rangle\\ \sigma_A^2 &= \langle (A-\langle A\rangle)^2\rangle\end{aligned}$$

また，交換子 *(commutator)* も次のように定義する．

$$\begin{aligned}[A,B] &= AB-BA\\ \{A,B\} &= AB+BA\end{aligned}$$

Theorem 1.4 (Uncertainty Principle)**.** *$A,B: H\to H$; エルミート作用素，$\psi\in H, \|\psi\|_H = 1$ とする．*

このとき，A,B の不確実性の積は次のように下から評価される．

$$\sigma_A^2\sigma_B^2 \geq \left|\frac{\langle\{A,B\}\rangle - 2\langle A\rangle\langle B\rangle}{2}\right|^2 + \left|\frac{\langle [A,B]\rangle}{2}\right|^2$$

Proof. Cauchy-Schwarz の不等式を使う． □

1.2 Information and Entropy

有限の状態空間 V に値をとる確率変数 $X:\Omega\to V$ を考える．X の分布は μ とする．

Definition 1.5. 確率変数 X の値が x であるという事象のもつ情報量 $I(x)=I_X(x)$ を次で定める．

$$I(x) \coloneqq -\log(P(X=x)) = -\log\mu(x)$$

また，分布 μ の持つ *(Shannon)* エントロピー $H(\mu)$ は情報量の平均であり，次で定められる．

$$H(\mu) \coloneqq -\sum_{x\in V}\mu(x)\log(\mu(x))$$

Remark 1.6. エントロピーは $H(X)$ や $S(X)$ と書かれることもある．共立出版の『確率論的エントロピー』では条件つきエントロピーや相対エントロピーなどの関連概念を含めた性質や公理系によるエントロピーの特徴づけがなされている．

Definition 1.7. $V = \mathbb{R}^n$ という連続的な空間の場合にも同様に情報量，エントロピーを定義する．X の分布 μ のルベーグ測度に対する密度関数を ρ として，

$$I(x) = -\log \rho(x)$$
$$H(\mu) = -\int_{\mathbb{R}^n} \rho(x) \log \rho(x) dx$$

Proposition 1.8. μ, ν を V 上の確率測度とする．このとき積測度 $\mu \otimes \nu$ は次を満たす．

$$H(\mu \otimes \nu) = H(\mu) + H(\nu)$$

Proof. 定義に従って計算する．$(\mu \otimes \nu)(A \times B) = \mu(A)\nu(B)$ を使う． □

2 Information Gain, Distance and Divergence

TODO:

2.1 KL divergence

エントロピーの定義では暗に一様な測度を利用している（有限空間における数え上げ測度，$\mathbb{R}^n$ におけるルベーグ測度）．しかし，無限次元 Banach 空間では一様な測度は存在しない．

Definition 2.1 (Kullback–Leibler divergence)**.** μ, ν を可測空間 $(\mathcal{X}, \mathcal{F})$ 上の σ 有限測度とする．μ から ν への *Kullback–Leibler divergence*（*KL-divergence*）を次で定める．

$$D_{KL}(\mu||\nu) := \int_{\mathcal{X}} \frac{d\mu}{d\nu} \log \frac{d\mu}{d\nu} d\nu = \int_{\mathcal{X}} \log \frac{d\mu}{d\nu} d\mu$$

ただし，$\mu \ll \nu$ のときのみ上式で定め，そうでない場合は $+\infty$ とする．$\frac{d\mu}{d\nu}$ は μ の ν に対する *Radon-Nicodým* 導関数

Remark 2.2. *KL divergence* は相対エントロピーともいう．$\mathcal{X}$ が有限集合か $\mathbb{R}^n$ で ν が一様測度の場合には *KL-divergence* はエントロピーに一致する．

Remark 2.3. *Gelfand-Kolmogorov-Yaglom* の定理により *Kl divergence* が *wll-defined* であることが保証される．

Remark 2.4. μ, ν を可測空間 $(\mathcal{X}, \mathcal{F})$ 上の σ 有限測度として以下が成り立つ．

(1) $D_{KL}(\mu||\nu) \geq 0$ かつ $= 0 \Leftrightarrow \mu = \nu$
(2) 一般には $D_{KL}(\mu||\nu) \neq D_{KL}(\nu||\mu)$
(3) 上記のことから *KL divergence* は距離にはならないが位相を定めることはできる．
$\mathcal{M} = \mathcal{M}_+(\mathcal{X})$ *or* $\mathcal{M}_1(\mathcal{X})$ 上の開球を $U(\mu, \epsilon) = \{\nu; D_{KL}(\mu||\nu) < \epsilon\}$ で与える．

Theorem 2.5 (Pinsker の不等式)**.** $\forall \mu, \nu \in \mathcal{M}_1(\mathcal{X})$ に対して,

$$d_{TV}(\mu, \nu) \leq \sqrt{2D_{KL}(\mu||\nu)}$$

が成り立つ．ただし，$d_{TV}(\mu, \nu) = \sup_{A \in \mathcal{F}} |\mu(A) - \nu(A)|$

Proof. 符号付き速度に対する Hahn 分解を使う． □

Remark 2.6. 逆方向の不等号は成り立たない．つまり，$\forall \epsilon > 0, \exists \mu, \nu \in \mathcal{M}_1(\mathcal{X})$ *s.t.* $d_{TV}(\mu, \nu) < \epsilon, D_{KL}(\mu, \nu) = \infty$ が成り立つ．

2.2 Divergences and Other Distance

より一般の距離のような関数を定める．

Definition 2.7 (f-divergence)**.** μ, ν を可測空間 $(\mathcal{X}, \mathcal{F})$ 上の σ 有限測度とする．$f : [0, \infty) \to \mathbb{R} \cup \{\infty\}$ が凸函数で，$f(1) = 0$ を満たすとする．

このとき，*f-divergence* を次で定める．

$$D_f(\mu||\nu) = \int_{\mathcal{X}} f\left(\frac{d\mu}{d\nu}\right) d\nu \tag{2.1}$$

ただし，$\mu \ll \nu$ のときのみ上式で定め，それ以外の場合は $+\infty$ とする．

また，$\mu, \nu \ll \rho$ となる参照測度 ρ に対して,

$$D_f(\mu||\nu) = \int_{\mathcal{X}} f\left(\frac{d\mu}{d\rho}/\frac{d\nu}{d\rho}\right) \frac{d\nu}{d\rho} d\rho \tag{2.2}$$

と書くこともできる．

Lemma 2.8. *(2.2)* は *well-defined* である．つまり，参照測度 ρ の取り方に依存しない．

Remark 2.9. *f-divergence* によっていくつかの測度間の擬似距離関数を表すことができる．

(1) *total variation*d_{TV} は $f(t) = |t - 1|$ に対する *f-divergence.*

(2) *KL-divergence* は $f(t) = t \log t$ に対する *f-divergence.*

(3) *Hellinger distance*d_H は $f(t) = |\sqrt{t} - 1|^2$ に対する *f-divergence.*
$d_H(\mu, \nu)^2 = \int_{\mathcal{X}} \left|\sqrt{\frac{d\mu}{d\rho}} - \sqrt{\frac{d\nu}{d\rho}}\right|^2 d\rho = \int_{\mathcal{X}} \left|\sqrt{\frac{d\mu}{d\nu}} - 1\right|^2 d\nu$

Theorem 2.10 (Kraft)**.**

$$d_H(\mu, \nu)^2 \leq d_{TV}(\mu, \nu) \leq 2d_H(\mu, \nu)$$

Proposition 2.11. $(V, \|\cdot\|)$ を *Banach* 空間とし，$f : \mathcal{X} \to V$ は $\mu, \nu \in \mathcal{M}_1(\mathcal{X})$ に関して *2* 次モーメントが有限とする．このとき以下の不等式が成り立つ．

$$\|\mathbb{E}_\mu[f] - \mathbb{E}_\nu[f]\| \leq 2\sqrt{\mathbb{E}_\mu[\|f\|^2] + \mathbb{E}_\nu[\|f\|^2]}d_H(\mu, \nu)$$

2.3 Topology of Weak Convergence of Probability Measures

Definition 2.12 (Lévy-Prokhorov distance)**.** (Ω, d) を距離空間とし，μ, ν を Ω 上の確率測度とする．*Lévy-Prokhorov* 距離 d_{LP} を次で定める．

$$d_{LP}(\mu, \nu) = \inf\{\epsilon > 0; \mu(A) \leq \nu(A^\epsilon) + \epsilon, \nu(A) \leq \mu(A^\epsilon) + \epsilon, \forall A \in \mathcal{F}\}$$

ただし，$A^\epsilon = \cup_{a \in A} B_\epsilon(a)$ は A の ϵ 近傍.

Theorem 2.13 (Portmanteau theorem for weak convergence)**.** Ω を可分な位相空間とする．$(\mu_n)_{n\in\mathbb{N}} \subset \mathcal{M}_1(\Omega), \mu \in \mathcal{M}_1(\Omega)$ に対して，以下は同値.

(a) $\limsup_{n\to\infty} \mu_n(F) \leq \mu(F)\ \forall F \overset{closed}{\subset} \Omega$
(b) $\liminf_{n\to\infty} \mu_n(U) \leq \mu(U)\ \forall U \overset{open}{\subset} \Omega$
(c) $\lim_{n\to\infty} \mu_n(A) = \mu(A)$ *with* $\mu(\partial A)(= \mu(\bar{A}) - \mu(\mathring{A})) = 0$
(d) $\lim_{n\to\infty} \int_\Omega f d\mu_n = \int_\Omega f d\mu\ \forall f$: 有界，連続
(e) $\lim_{n\to\infty} \int_\Omega f d\mu_n = \int_\Omega f d\mu\ \forall f$: 有界，*Lipschitz* 連続
(f) $\limsup_{n\to\infty} \int_\Omega f d\mu_n \leq \int_\Omega f d\mu\ \forall f$: 上半連続，上に有界
(g) $\liminf_{n\to\infty} \int_\Omega f d\mu_n \geq \int_\Omega f d\mu\ \forall f$: 下半連続，下に有界
(h) Ω が d で距離づけされているとき，$\lim_{n\to\infty} d_{LP}(\mu_n, \mu) = 0$

Proof. まず．(a) ⇔ (b) を示した後に，((a) + (b)) ⇒ (c) ⇒ (d) ⇒ (e) ⇒ ((a) + (b)) を示し，((a) + (b)) ⇔ (h)，(d) ⇔ ((f)+(g)), (f) ⇔ (g) を示す． □

参考文献

[1] Timothy John Sullivan. *Introduction to uncertainty quantification*, volume 63. Springer, 2015.