2020 年 8 月 5 日

量子情報数理特論
（第 13 回）量子 Stein の補題：順定理

電気通信大学 大学院情報理工学研究科

小川朋宏

12　量子Steinの補題：順定理

(最終レポート，その１)

講義中の関数$\psi(s)$と$\varphi(a)$のグラフを（ひとつのグラフ中に）を描け．その際，以下の特徴を明示せよ．

(1) 厳密に凸関数であること

(2) $\psi(0)$, $\psi(1)$の値

(3) $s=0$, $s=1$における接線の傾き

(4) $\varphi(a)$および$\varphi(a)-a$の値

提出方法：手書きで写真(jpeg, pdf)による提出を推奨（iPad等でもOK）

締切：2020年8月31日（本日から１週間以内を推奨）

12.1 量子 Neyman-Pearson 検定

Lemma 1 (前回 Lemma 2). 任意のエルミート作用素 A について,

$$\mathrm{Tr}\, A_+ = \mathrm{Tr}\, A\{A > 0\} = \max_{0 \le T \le I} \mathrm{Tr}\, AT \tag{1}$$

Lemma 2 (前回 Lemma 6). 任意のエルミート作用素 A, B について,

$$\min_{0 \le T \le I} \{\mathrm{Tr}\, A(I - T) + \mathrm{Tr}\, BT\} = \mathrm{Tr}\, A - \mathrm{Tr}(A - B)_+ \tag{2}$$

ここで, 最小化は $S := \{A - B > 0\}$ で達成される.

(別証明) $\mathrm{Tr}\, A(I - T) + \mathrm{Tr}\, BT = \mathrm{Tr}\, A - \mathrm{Tr}(A - B)T$ より,

$$(\text{左辺}) = \mathrm{Tr}\, A - \max_{0 \le T \le I} \mathrm{Tr}(A - B)T = (\text{右辺})$$

Lemma 3. 上記を $\rho, \sigma \in \mathcal{S}(\mathcal{H})$ の仮説検定問題に適用 ($A = \rho$, $B = e^a \sigma$ $(a \in \mathbb{R})$ とおく)

$$\min_{0 \le T \le I} \{\alpha(T) + e^a \beta(T)\} = \min_{0 \le T \le I} \{\mathrm{Tr}\, \rho\, (I - T) + e^a\, \mathrm{Tr}\, \sigma T\} = 1 - \mathrm{Tr}(\rho - e^a \sigma)_+ \tag{3}$$

最小化は $S(a) := \{\rho - e^a \sigma > 0\}$ (量子 Neyman-Pearson 検定) で達成される.

12.2 量子 Neyman-Pearson の補題

Lemma 4. 量子 Neyman-Pearson 検定 $S(a) := \{\rho - e^a\sigma > 0\}$ $(a \in \mathbb{R})$ と，
任意の検定 $0 \leq T \leq I$ について，

$$\alpha(T) \leq \alpha(S(a)) \implies \beta(T) \geq \beta(S(a))$$

（意味：第一種誤りで Neyman-Pearson 検定より性能が良いと，第二種誤りでは性能が悪くなる）

（証明） Lemma 3 より，任意の検定 $0 \leq T \leq I$ について，

$$\alpha(S(a)) + e^a\beta(S(a)) \leq \alpha(T) + e^a\beta(T) \tag{4}$$

これを移項することで，

$$0 \overset{\text{仮定}}{\leq} \alpha(S(a)) - \alpha(T) \leq e^a\{\beta(T) - \beta(S(a))\} \tag{5}$$

□

12.3 Neyman-Pearson検定の誤り確率の上界

Lemma 5 (復習，Audenaert *et al.* の不等式)**.** 任意の非負定値作用素A, Bと，$0 \leq \forall s \leq 1$について，

$$\min_{0 \leq T \leq I} \{\mathrm{Tr}\, A(I-T) + \mathrm{Tr}\, BT\} \leq \mathrm{Tr}\, A^s B^{1-s}. \tag{6}$$

$\rho, \sigma \in \mathcal{S}(\mathcal{H})$の仮説検定問題に適用，$A = \rho$, $B = e^a \sigma$ $(a \in \mathbb{R})$とおく．Lemma 3と合わせて，

$$\begin{aligned}\alpha(S(a)) + e^a \beta(S(a)) &= \min_{0 \leq T \leq I} \{\alpha(T) + e^a \beta(T)\} \\ &= \min_{0 \leq T \leq I} \{\mathrm{Tr}\, \rho\, (I-T) + e^a\, \mathrm{Tr}\, \sigma T\} = 1 - \mathrm{Tr}(\rho - e^a \sigma)_+ \leq e^{a(1-s)}\, \mathrm{Tr}\, \rho^s \sigma^{1-s}\end{aligned}$$

よって，

$$\alpha(S(a)) \leq \alpha(S(a)) + e^a \beta(S(a)) \leq e^{a(1-s)}\, \mathrm{Tr}\, \rho^s \sigma^{1-s} \tag{7}$$

$$e^a \beta(S(a)) \leq \alpha(S(a)) + e^a \beta(S(a)) \leq e^{a(1-s)}\, \mathrm{Tr}\, \rho^s \sigma^{1-s} \tag{8}$$

Lemma 6. 任意の$a \in \mathbb{R}$と$0 \leq s \leq 1$について，

$$\alpha(S(a)) \leq e^{a(1-s)}\, \mathrm{Tr}\, \rho^s \sigma^{1-s}, \quad \beta(S(a)) \leq e^{-as}\, \mathrm{Tr}\, \rho^s \sigma^{1-s} \tag{9}$$

12.4　大偏差理論 (Large Deviation Theory)

- $\rho^{\otimes n}$, $\sigma^{\otimes n}$ の仮説検定について，$\rho \leftarrow \rho_n = \rho^{\otimes n}$, $\sigma \leftarrow \sigma_n = \sigma^{\otimes n}$, $a \leftarrow na$ の置き換えをする.
- $S_n(a) := \{\rho_n - e^{na}\sigma_n > 0\}$ を量子Neyman-Pearson検定として，(9)より，

$$\alpha_n(S_n(a)) \leq e^{na(1-s)} \operatorname{Tr} \rho_n^s \sigma_n^{1-s}, \quad \beta_n(S_n(a)) \leq e^{-nas} \operatorname{Tr} \rho_n^s \sigma_n^{1-s} \tag{10}$$

- $\operatorname{Tr} \rho_n^s \sigma_n^{1-s} = \operatorname{Tr} \left(\rho^{\otimes n}\right)^s \left(\sigma^{\otimes n}\right)^{1-s} = \operatorname{Tr} \left(\rho^s\right)^{\otimes n} \left(\sigma^{1-s}\right)^{\otimes n} = \operatorname{Tr} \left(\rho^s \sigma^{1-s}\right)^{\otimes n} = \left(\operatorname{Tr} \rho^s \sigma^{1-s}\right)^n$ に注意.

Definition 1. $\psi(s) := \log \operatorname{Tr} \rho^s \sigma^{1-s}, \quad \varphi(a) := \max_{0 \leq s \leq 1} \{as - \psi(s)\}$

(10)より，

$$\frac{1}{n} \log \alpha_n(S_n(a)) \leq -a(s-1) + \frac{1}{n} \log \operatorname{Tr} \rho_n^s \sigma_n^{1-s} = -\{as - \psi(s)\} + a \tag{11}$$

$$\frac{1}{n} \log \beta_n(S_n(a)) \leq -as + \frac{1}{n} \log \operatorname{Tr} \rho_n^s \sigma_n^{1-s} = -\{as - \psi(s)\} \tag{12}$$

(11), (12)の上界を$0 \leq s \leq 1$について最適化すると，

Lemma 7. 任意の$a \in \mathbb{R}$について，

$$\frac{1}{n} \log \alpha_n(S_n(a)) \leq -\{\varphi(a) - a\} \tag{13}$$

$$\frac{1}{n} \log \beta_n(S_n(a)) \leq -\varphi(a) \tag{14}$$

12.5　関数 $\psi(s)$ の性質

$$\psi(s) = \log \operatorname{Tr} \rho^s \sigma^{1-s} \tag{15}$$

- スペクトル分解を考える．

$$\rho = \sum_i a_i E_i, \quad \sigma = \sum_j b_j F_j$$

- $P(i,j) := a_i \operatorname{Tr} E_i F_j, \quad Q(i,j) := b_j \operatorname{Tr} E_i F_j$ とおくと，$P(i,j)$, $Q(i,j)$ は確率分布である．

$$\sum_i \sum_j P(i,j) = \operatorname{Tr} \left(\sum_i a_i E_i \right) \left(\sum_j F_j \right) = \operatorname{Tr} \rho = 1 \quad \text{（$Q(i,j)$ についても同様）}$$

- $P(i,j)$, $Q(i,j)$ で表すと，

$$\operatorname{Tr} \rho^s \sigma^{1-s} = \sum_i \sum_j a_i^s b_j^{1-s} \operatorname{Tr} E_i F_j = \sum_i \sum_j (a_i \operatorname{Tr} E_i F_j)^s (b_i \operatorname{Tr} E_i F_j)^{1-s} = \sum_i \sum_j P(i,j)^s Q(i,j)^{1-s}$$

- $x = (i,j)$ と変数の書き換えを行えば，

$$\psi(s) = \log \sum_{x \in \mathcal{X}} P(x)^s Q(x)^{1-s}$$

は $P(x)$ と $Q(x)$ の（古典）f-ダイバージェンス（の一種）とみなせる．

Definition 2. 確率分布 $P(x)$, $Q(x)$ について，

$$D_f(P||Q) = \sum_{x\in\mathcal{X}} P(x) f\left(\frac{Q(x)}{P(x)}\right) \tag{16}$$

Example 1. $f(t) = -\log t$ のときは，相対エントロピー $D(P||Q)$ である．

$$D_f(P||Q) = -\sum_{x\in\mathcal{X}} P(x) \log\left(\frac{Q(x)}{P(x)}\right) = \sum_{x\in\mathcal{X}} P(x) \log\frac{P(x)}{Q(x)} \tag{17}$$

Example 2. $f(t) = t\log t$ のときは，相対エントロピー $D(Q||P)$ である．

$$D_f(P||Q) = \sum_{x\in\mathcal{X}} P(x) \left(\frac{Q(x)}{P(x)}\right) \log\left(\frac{Q(x)}{P(x)}\right) = \sum_{x\in\mathcal{X}} Q(x) \log\frac{Q(x)}{P(x)} \tag{18}$$

Example 3. $f(t) = t^{1-s}$ のとき，

$$D_f(P||Q) = \sum_{x\in\mathcal{X}} P(x) \left(\frac{Q(x)}{P(x)}\right)^{1-s} = \sum_{x\in\mathcal{X}} P(x)^s Q(x)^{1-s} \tag{19}$$

12.7　f-ダイバージェンスの単調性

条件付き確率 $W(y|x)$ $(x \in \mathcal{X}, y \in \mathcal{Y})$ から定まる古典通信路：

$$P \in \mathcal{P}(\mathcal{X}) \longmapsto PW \in \mathcal{P}(\mathcal{Y}) \quad \left(PW(y) := \sum_{x \in \mathcal{X}} P(x)W(y|x) \right)$$

Theorem 1 (単調性). $f(t)$ が凸関数であるとき，

$$D_f(P||Q) \geq D_f(PW||QW) \tag{20}$$

$f(t)$ が厳密に凸関数であるとき，等号成立の必要十分条件は

$$\exists V(x|y)\,(\text{条件付き確率}), \quad PWV = P, \quad QWV = Q \tag{21}$$

（証明）　入力のシンボル x が，$P(x)$ または $Q(x)$ で発生する場合，入出力 (x,y) の同時分布は，

$$P_{XY}(x,y) := P(x)W(y|x), \quad Q_{XY}(x,y) := Q(x)W(y|x) \tag{22}$$

で与えられる．

$$P_Y(y) = \sum_{x \in \mathcal{X}} P_{XY}(x,y) = \sum_{x \in \mathcal{X}} P(x)W(y|x) = PW(y) \tag{23}$$

$$Q_Y(y) = \sum_{x \in \mathcal{X}} Q_{XY}(x,y) = \sum_{x \in \mathcal{X}} Q(x)W(y|x) = QW(y) \tag{24}$$

同時分布について，yを条件とする形で無理やり書き換える．

$$P_{XY}(x,y) = P_Y(y)V_P(x|y), \quad Q_{XY}(x,y) = Q_Y(y)V_Q(x|y) \tag{25}$$

入力分布P, Qに依存した条件付き分布になることに注意．

$$\begin{aligned} D_f(P_{XY}\|Q_{XY}) &= \sum_{x\in\mathcal{X}}\sum_{y\in\mathcal{Y}} P_{XY}(x,y) f\left(\frac{Q_{XY}(x,y)}{P_{XY}(x,y)}\right) = \sum_{x\in\mathcal{X}}\sum_{y\in\mathcal{Y}} P(x)W(y|x) f\left(\frac{Q(x)W(y|x)}{P(x)W(y|x)}\right) \\ &= \sum_{x\in\mathcal{X}}\sum_{y\in\mathcal{Y}} P(x)W(y|x) f\left(\frac{Q(x)}{P(x)}\right) = \sum_{x\in\mathcal{X}} P(x) f\left(\frac{Q(x)}{P(x)}\right) = D_f(P\|Q) \end{aligned} \tag{26}$$

$$\begin{aligned} D_f(P_{XY}\|Q_{XY}) &= \sum_{x\in\mathcal{X}}\sum_{y\in\mathcal{Y}} P_Y(y)V_P(x|y) f\left(\frac{Q_Y(y)V_Q(x|y)}{P_Y(y)V_P(x|y)}\right) \\ &= \sum_{y\in\mathcal{Y}} P_Y(y) \sum_{x\in\mathcal{X}} V_P(x|y) f\left(\frac{Q_Y(y)V_Q(x|y)}{P_Y(y)V_P(x|y)}\right) \\ &\geq \sum_{y\in\mathcal{Y}} P_Y(y) f\left(\sum_{x\in\mathcal{X}} V_P(x|y)\frac{Q_Y(y)V_Q(x|y)}{P_Y(y)V_P(x|y)}\right) \end{aligned} \tag{27}$$

$$= \sum_{y\in\mathcal{Y}} P_Y(y) f\left(\frac{Q_Y(y)}{P_Y(y)}\right) = D_f(PW\|QW) \tag{28}$$

- 等号成立条件について，(21) $\Rightarrow$ (20) の等号成立を示す．(21) を仮定すると，既に証明した (20) を二回使うことで，

$$D_f(P||Q) \geq D_f(PW||QW) \geq D_f(PWV||QWV) = D_f(P||Q) \tag{29}$$

両辺が一致して，(20) の等号が成立することが示された．

- $f(t)$ が厳密に凸関数であるとき，(20) の等号成立 $\Rightarrow$ (21) を示す．$f(t)$ が厳密に凸関数である場合，(20) の等号成立するのは，各 $y \in \mathcal{Y}$ について (27) における関数の引数がすべて等しい場合に限られる．すなわち，

$$\forall y \in \mathcal{Y}, \exists C_y, \forall x \in \mathcal{X}, \frac{Q_Y(y)V_Q(x|y)}{P_Y(y)V_P(x|y)} = \frac{Q_Y(y)}{P_Y(y)} = C_y \tag{30}$$

これより，

$$\forall y \in \mathcal{Y}, \forall x \in \mathcal{X}, \frac{V_Q(x|y)}{V_P(x|y)} = 1 \tag{31}$$

$V(x|y) := V_Q(x|y) = V_P(x|y)$ とおけば，(21) より，

$$P_{XY}(x,y) = P(y)V(x|y), \quad Q_{XY}(x,y) = Q(y)V(x|y) \tag{32}$$

この周辺分布をとれば，

$$P(x) = \sum_{y\in\mathcal{Y}} P_{XY}(x,y) = \sum_{y\in\mathcal{Y}} P(y)V(x|y), \quad Q(x) = \sum_{y\in\mathcal{Y}} Q_{XY}(x,y) = \sum_{y\in\mathcal{Y}} Q(y)V(x|y) \tag{33}$$

$\square$

Lemma 8 (正値性)**.** $f(t)$ が凸関数であるとき，

$$D_f(P||Q) \geq D_f(P_0||P_0) = f(1) \tag{34}$$

$f(t)$ が厳密に凸な関数であるとき，等号成立の必要十分条件は $P = Q$ である.

- $f(t) = -t^{1-s}$ は $0 \leq s \leq 1$ で凸関数であるから，

$$D_f(P||Q) = -\sum_{x\in\mathcal{X}} P(x)\left(\frac{Q(x)}{P(x)}\right)^{1-s} = -\sum_{x\in\mathcal{X}} P(x)^s Q(x)^{1-s} \geq f(1) = -1 \tag{35}$$

すなわち，

$$\psi(s) = \log \sum_{x\in\mathcal{X}} P(x)^s Q(x)^{1-s} \leq 0 \quad (0 \leq s \leq 1) \tag{36}$$

- $f(t) = t^{1-s}$ は $s \leq 0$ または $s \geq 1$ で凸関数であるから，

$$D_f(P||Q) = \sum_{x\in\mathcal{X}} P(x)\left(\frac{Q(x)}{P(x)}\right)^{1-s} = \sum_{x\in\mathcal{X}} P(x)^s Q(x)^{1-s} \geq f(1) = 1 \tag{37}$$

すなわち，

$$\psi(s) = \log \sum_{x\in\mathcal{X}} P(x)^s Q(x)^{1-s} \geq 0 \quad (s \leq 0, s \geq 1) \tag{38}$$

12.9 関数 $\psi(s)$ の性質

Lemma 9. $\psi(s) = \log \sum_{x \in \mathcal{X}} P(x)^s Q(x)^{1-s}$ について，以下の性質が成り立つ．

(i) $\psi(0) = 0$, $\psi(1) = 0$

(ii) $\psi(s) \leq 0$ $(0 \leq s \leq 1)$, $\psi(s) \geq 0$ $(s \leq 0, s \geq 1)$.

(iii) $\psi'(0) = -D(Q||P)$, $\psi'(1) = D(P||Q)$.

(iv) $P \neq Q$ ならば，$\psi''(s) > 0$ である（$\psi(s)$ は厳密に凸な関数）

（証明）(i) は自明，(ii) は既に示した．
定義より，$e^{\psi(s)} = \sum_{x \in \mathcal{X}} P(x)^s Q(x)^{1-s}$ であり，

$$P_s(x) := e^{-\psi(s)} P(x)^s Q(x)^{1-s} = \frac{P(x)^s Q(x)^{1-s}}{\sum_{x \in \mathcal{X}} P(x)^s Q(x)^{1-s}} \tag{39}$$

とおくと，$P_s(x)$ は確率分布で

$$P_0(x) = Q(x), \quad P_1(x) = P(x) \tag{40}$$

微分を計算する．

$$
\begin{aligned}
\psi'(s) &= \frac{\sum_{x\in\mathcal{X}} P(x)^s Q(x)^{1-s}\left\{\log P(x) - \log Q(x)\right\}}{\sum_{x\in\mathcal{X}} P(x)^s Q(x)^{1-s}} \\
&= e^{-\psi(s)} \sum_{x\in\mathcal{X}} P(x)^s Q(x)^{1-s}\left\{\log P(x) - \log Q(x)\right\} \qquad (41) \\
&= \sum_{x\in\mathcal{X}} P_s(x)\left\{\log P(x) - \log Q(x)\right\} \\
&= \mathrm{E}\Big[\log P(X) - \log Q(X) \;\Big|\; X \sim P_s\Big], \qquad (42)
\end{aligned}
$$

ここで，$s=0$, $s=1$とすると(iii)が示される．(41)より，

$$
\begin{aligned}
\psi''(s) &= -\psi'(s) e^{-\psi(s)} \sum_{x\in\mathcal{X}} P(x)^s Q(x)^{1-s}\left\{\log P(x) - \log Q(x)\right\} \\
&\qquad + e^{-\psi(s)} \sum_{x\in\mathcal{X}} P(x)^s Q(x)^{1-s}\left\{\log P(x) - \log Q(x)\right\}^2 \\
&= \sum_{x\in\mathcal{X}} P_s(x)\left\{\log P(x) - \log Q(x)\right\}^2 - \psi'(s) \sum_{x\in\mathcal{X}} P_s(x)\left\{\log P(x) - \log Q(x)\right\} \\
&= \mathrm{E}\Big[\left\{\log P(X) - \log Q(X)\right\}^2\Big] - \Big\{\mathrm{E}\left[\log P(X) - \log Q(X)\right]\Big\}^2 \\
&= V\Big[\log P(X) - \log Q(X) \;\Big|\; X \sim P_s\Big] > 0 \qquad (43)
\end{aligned}
$$

最後の符号号は分散が正であることを用いた．よって(iv)が示された．

12.10　$\psi(s)$ と $\varphi(a)$ のグラフ