付録　証明・説明と解答例

定理 2.5(➔p.12) の証明

(1) の証明: 1 次元データ $x = (x_1, x_2, \ldots, x_n)$ を小さい順に並び替えて

$$a_1 \leqq a_2 \leqq \cdots \leqq a_{n-1} \leqq a_n$$

となったとする。

(i) $n = 1$ のとき，$L(a) = |a_1 - a|$ を最小化する a が $a = a_1$ で与えられることは明らか。

(ii) $n = 2$ のとき

$$\begin{aligned} L(a) &= |a_1 - a| + |a_2 - a| \\ &= \begin{cases} (a_1 - a) + (a_2 - a) = (a_2 - a_1) + 2(a_1 - a) & \text{if } a < a_1, \\ (a - a_1) + (a_2 - a) = a_2 - a_1 & \text{if } a_1 \leqq a \leqq a_2, \\ (a - a_1) + (a - a_2) = (a_2 - a_1) + 2(a - a_2) & \text{if } a_2 < a \end{cases} \end{aligned}$$

であるから，$a_1 \leqq a \leqq a_2$ のときに $L(a)$ は最小値をとる。特に $a = \frac{a_1 + a_2}{2}$ のときにも最小値をとる。

(iii) $n \geqq 3$ のとき: $a < a_1$ のときは $L(a) > L(a_1)$ であり，$a_n < a$ のときは $L(a) > L(a_n)$ であるから，$L(a)$ が $a \in (-\infty, a_1) \cup (a_n, \infty)$ のときに最小値をとることはない。そこで以下 $a_1 \leqq a \leqq a_n$ とすると，

$$\begin{aligned} L(a) = \sum_{i=1}^{n} |a_i - a| &= \underbrace{|a_1 - a|}_{\substack{\| \\ a - a_1}} + \sum_{i=2}^{n-1} |a_i - a| + \underbrace{|a_n - a|}_{\substack{\| \\ a_n - a}} \\ &= \underbrace{(a_n - a_1)}_{\substack{a\text{ によらない} \\ \text{数になってる}}} + \sum_{i=2}^{n-1} |a_i - a| \end{aligned}$$

となるから，$\sum_{i=2}^{n-1} |a_i - a|$ を最小化する a を見つければよい。この操作を繰り返していくと，n の偶奇に応じて，結果として上の (i) または (ii) の場合に帰着され，a が中央値の場合において $L(a)$ が最小値をとることがわかる。

(2) の証明: 1 次元データ $x = (x_1, x_2, \ldots, x_n)$ に対して $L(a)$ は a に関する二次関数となる。これを平

方完成すると

$$
\begin{aligned}
L(a) &= \sum_{i=1}^{n}(x_i - a)^2 = \sum_{i=1}^{n}(a^2 - 2x_i a + (x_i)^2) \\
&= n\left\{a^2 - 2\left(\frac{1}{n}\sum_{i=1}^{n} x_i\right)a + \frac{1}{n}\sum_{i=1}^{n}(x_i)^2\right\} \\
&= n\left(a - \frac{1}{n}\sum_{i=1}^{n} x_i\right)^2 + n\left\{\frac{1}{n}\sum_{i=1}^{n}(x_i)^2 - \left(\frac{1}{n}\sum_{i=1}^{n} x_i\right)^2\right\} \\
&= n\left(a - \frac{1}{n}\sum_{i=1}^{n} x_i\right)^2 + nv
\end{aligned}
$$

となり，ゆえに $L(a)$ は $a = \frac{1}{n}\sum_{i=1}^{n} x_i$ において最小値をとる。

公式 4.18(➔p. 118) の「説明」

4.2 節 (➔p. 116) における問題設定を確認しよう。
ある母集団について $(2+1)$-種類の「変量」X_1, X_2, Y を考えるとき，標本調査をして母集団から N 個の標本 $\omega_1, \omega_2, \ldots, \omega_N$ を無作為に選び，これらに関して変量 X_1, X_2, Y の値を調べて左下図のように表にまとめ，対応するデータ行列を X とする。

標本点 \ 変量	説明変数		目的変数
	X_1	X_2	Y
ω_1	x_{11}	x_{12}	y_1
ω_2	x_{21}	x_{22}	y_2
$\vdots$	$\vdots$	$\vdots$	$\vdots$
ω_N	x_{N1}	x_{N2}	y_N

$$
X = \begin{pmatrix} x_{11} & x_{12} \\ x_{21} & x_{22} \\ \vdots & \vdots \\ x_{N1} & x_{N2} \end{pmatrix} = \begin{pmatrix} \boldsymbol{x}_1 & \boldsymbol{x}_2 \end{pmatrix}
$$

変量 Y の予測値 $\widehat{Y} = b_1 X_1 + b_2 X_2 + c$ をとったとき，$\widehat{Y} = (X_1, X_2)\begin{pmatrix} b_1 \\ b_2 \end{pmatrix} + c$ と書き直した形に照らし合わせて次のように表しておく。

$$
\hat{\boldsymbol{y}} = \begin{pmatrix} \hat{y}_1 \\ \hat{y}_2 \\ \vdots \\ \hat{y}_N \end{pmatrix} = \underbrace{(\boldsymbol{x}_1, \boldsymbol{x}_2)}_{=X}\underbrace{\begin{pmatrix} b_1 \\ b_2 \end{pmatrix}}_{=\boldsymbol{b}} + \underbrace{\begin{pmatrix} c \\ c \\ \vdots \\ c \end{pmatrix}}_{=\boldsymbol{c}} = X\boldsymbol{b} + \boldsymbol{c},
$$

$$
\hat{\boldsymbol{\varepsilon}} = \begin{pmatrix} \hat{\varepsilon}_1 \\ \hat{\varepsilon}_2 \\ \vdots \\ \hat{\varepsilon}_N \end{pmatrix} = \begin{pmatrix} y_1 - \hat{y}_1 \\ y_2 - \hat{y}_2 \\ \vdots \\ y_N - \hat{y}_N \end{pmatrix} = \boldsymbol{y} - \hat{\boldsymbol{y}}
$$

変量 X_1, X_2 に関する 1 次元データを縦に並べたベクトル $\boldsymbol{x}_1, \boldsymbol{x}_2$ が多重共線性をもたないとき，関数

$(b_1, b_2, c) \mapsto \displaystyle\sum_{i=1}^{N} \left\{y_i - \left(b_1 x_{i1} + b_2 x_{i2} + c\right)\right\}^2$ $(= \|\boldsymbol{y} - (X\boldsymbol{b} + \boldsymbol{c})\|^2 = \|\boldsymbol{y} - \widehat{\boldsymbol{y}}\|^2 = \|\hat{\boldsymbol{\varepsilon}}\|^2)$ の最小点 $(\hat{b}_1, \hat{b}_2, \hat{c})$ を求めたいのであった。
そこで中心化行列 P と平均化行列 Q (定義 4.6 ➔ p. 112) を用いて $\|\hat{\boldsymbol{\varepsilon}}\|^2$ を少し計算してみると，

$$
\begin{aligned}
\|\hat{\boldsymbol{\varepsilon}}\|^2 = \|\boldsymbol{y} - (X\boldsymbol{b} + \boldsymbol{c})\|^2 &\overset{\substack{\text{命題 4.12–(1)} \\ (➔\ \text{p. 112})}}{=} \|(P+Q)\big(\boldsymbol{y} - (X\boldsymbol{b} + \boldsymbol{c})\big)\|^2 \\
&\overset{\substack{\text{命題 4.5–(2)} \\ (➔\ \text{p. 108})}}{=} \|\underbrace{P\big(\boldsymbol{y} - (X\boldsymbol{b} + \boldsymbol{c})\big)}_{\substack{\| \\ P\boldsymbol{y} - PX\boldsymbol{b} - (\underbrace{P\boldsymbol{c}}_{=\boldsymbol{0}})}}\|^2 + \|\underbrace{Q\big(\boldsymbol{y} - (X\boldsymbol{b} + \boldsymbol{c})\big)}_{\substack{\| \\ -(\underbrace{Q\boldsymbol{c}}_{=\boldsymbol{c}} - Q\boldsymbol{y} + QX\boldsymbol{b})}}\|^2 \\
&= \underbrace{\|P\boldsymbol{y} - PX\boldsymbol{b}\|^2}_{\substack{\text{この値の大小をコントロール} \\ \text{するのは } b_1, b_2 \text{ の値であり，} \\ c \text{ の値にはよらない。}}} + \underbrace{\|\boldsymbol{c} - Q\boldsymbol{y} + QX\boldsymbol{b}\|^2}_{\substack{\text{この値の大小をコントロール} \\ \text{するのは } b_1, b_2 \text{ と } c \text{ の値であり，} \\ \boldsymbol{c} = Q\boldsymbol{y} - QX\boldsymbol{b} \\ \text{のときに最小値 } 0 \text{ をとる。}}}
\end{aligned}
$$

となる。ゆえに，$\|\hat{\boldsymbol{\varepsilon}}\|^2$ を最小化するには以下の手順を踏めばよい。

(1) まず $\|P\boldsymbol{y} - PX\boldsymbol{b}\|^2$ の値を最小化する $(b_1, b_2) = (\hat{b}_1, \hat{b}_2)$ を求め，

(2) その後，式 $\boldsymbol{c} = Q\boldsymbol{y} - QX\boldsymbol{b}$ によって $c = \hat{c}$ を求める。

この手順を実行する前に，少し準備をしておく。ベクトル $\boldsymbol{b}$ が平面 $\mathbb{R}^2$ を動くとき，$PX\boldsymbol{b}$ は実は，空間 $\mathbb{R}^N$ 内の $\boldsymbol{0}$ を含むようなある (一般には多次元の) 平面全体に渡って動くことがわかる。この平面を $\mathrm{Im}(PX)$ という記号で表す。
まず (1) を実行するために，以下の補題を用意する。

補題 1 $\mathbb{R}^N$ 内の二つのベクトル $\mathbf{v}$，$\mathbf{w}$ を

$$\mathbf{v} = PX(X^\top PX)^{-1}X^\top P\boldsymbol{y}, \quad \mathbf{w} = P\boldsymbol{y} - PX(X^\top PX)^{-1}X^\top P\boldsymbol{y}$$

により定めると，次が成り立つ。

(i) $P\boldsymbol{y} = \mathbf{v} + \mathbf{w}$

(ii) 任意の $\boldsymbol{b} \in \mathbb{R}^2$ に対して，$\langle \mathbf{w}, PX\boldsymbol{b} \rangle = 0$ が成り立つ。

← ベクトル $\mathbf{w}$ は平面 $\mathrm{Im}(PX)$ 内のどのベクトルとも「直交」する，つまり $\mathbf{w}$ は平面 $\mathrm{Im}(PX)$ に直交しているということ。

証明. (i) は明らかである。(ii) $\boldsymbol{b} \in \mathbb{R}^2$ を任意とすると，次のように確かめられる。

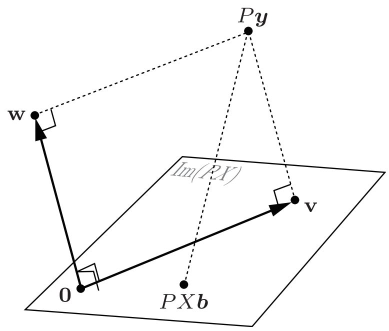
$$
\begin{aligned}
\langle \mathbf{w}, PX\boldsymbol{b} \rangle &= \left\langle P\boldsymbol{y} - PX(X^\top PX)^{-1}X^\top P\boldsymbol{y}, PX\boldsymbol{b} \right\rangle \\
&= \underbrace{\langle P\boldsymbol{y}, PX\boldsymbol{b} \rangle}_{\substack{\| \\ \langle \boldsymbol{y}, PX\boldsymbol{b} \rangle}} - \underbrace{\left\langle PX(X^\top PX)^{-1}X^\top P\boldsymbol{y}, PX\boldsymbol{b} \right\rangle}_{\substack{\| \\ \langle \underbrace{\underbrace{X^\top P^\top PX}_{\substack{\| \\ X^\top PX}}(X^\top PX)^{-1}}_{\substack{\| \\ E_2}} X^\top P\boldsymbol{y}, \boldsymbol{b} \rangle}} \\
&= \langle \boldsymbol{y}, PX\boldsymbol{b} \rangle - \langle X^\top P\boldsymbol{y}, \boldsymbol{b} \rangle = \langle \boldsymbol{y}, PX\boldsymbol{b} \rangle - \langle \boldsymbol{y}, PX\boldsymbol{b} \rangle = 0
\end{aligned}
$$

□

いま，$\boldsymbol{b}$ が平面 $\mathbb{R}^2$ 内を動き回るとき，

$$\|P\boldsymbol{y} - PX\boldsymbol{b}\|^2 = (P\boldsymbol{y}\ \textbf{と平面}\ \mathrm{Im}(PX)\ \textbf{内の点}\ PX\boldsymbol{b}\ \textbf{との距離})^2$$

の値を最小化する $\boldsymbol{b} = \hat{\boldsymbol{b}}$ は，



$$\begin{aligned}
&PX\hat{\boldsymbol{b}} \\
&= (\textbf{点}\ P\boldsymbol{y}\ \textbf{の平面}\ \mathrm{Im}(PX)\ \textbf{への直交射影}) \\
&\overset{\text{補題 1}}{=} \mathbf{v} \\
&= PX(X^\top PX)^{-1}X^\top P\boldsymbol{y}
\end{aligned}$$

をみたさなければならない。この最左辺と最右辺に $(X^\top PX)^{-1}X^\top$ を左から掛けると

$$\underbrace{(X^\top PX)^{-1}X^\top PX}_{=E_2}\hat{\boldsymbol{b}} = \underbrace{(X^\top PX)^{-1}X^\top PX}_{=E_2}(X^\top PX)^{-1}X^\top P\boldsymbol{y},$$

つまり

$$\hat{\boldsymbol{b}} = (\underbrace{X^\top PX}_{=NS_{X,X}})^{-1}X^\top P\boldsymbol{y} = \frac{1}{N}(S_{X,X})^{-1}X^\top P\boldsymbol{y} \tag{A.1}$$

となり $\hat{\boldsymbol{b}}$ が求まった。中心化行列 P が $P = P^2$ をみたすことに注意して，これをもう少し変形すると，

$$\begin{aligned}
\hat{\boldsymbol{b}} &= \frac{1}{N}(S_{X,X})^{-1}X^\top P\boldsymbol{y} \\
&= \frac{1}{N}(S_{X,X})^{-1} \times \underbrace{(X^\top P)}_{\begin{pmatrix} x_{11}-\overline{x}_1 & x_{21}-\overline{x}_1 & \cdots & x_{N1}-\overline{x}_1 \\ x_{12}-\overline{x}_2 & x_{22}-\overline{x}_2 & \cdots & x_{N2}-\overline{x}_2 \end{pmatrix}} \times \underbrace{(P\boldsymbol{y})}_{\begin{pmatrix} y_1-\overline{y} \\ y_2-\overline{y} \\ \vdots \\ y_N-\overline{y} \end{pmatrix}} \\
&= \frac{1}{N}(S_{X,X})^{-1}\begin{pmatrix} \sum_{i=1}^{N}(x_{i1}-\overline{x}_1)(y_i-\overline{y}) \\ \sum_{i=1}^{N}(x_{i2}-\overline{x}_2)(y_i-\overline{y}) \end{pmatrix} = (S_{X,X})^{-1}\begin{pmatrix} s(\boldsymbol{x}_1, \boldsymbol{y}) \\ s(\boldsymbol{x}_2, \boldsymbol{y}) \end{pmatrix}
\end{aligned}$$

これで公式 4.18 (➔ p. 118) における $\hat{\boldsymbol{b}} = \begin{pmatrix} \hat{b}_1 \\ \hat{b}_2 \end{pmatrix}$ が求まった。いま，手順 (2) を実行すると，

$$\begin{pmatrix} \hat{c} \\ \hat{c} \\ \vdots \\ \hat{c} \end{pmatrix} = Q\boldsymbol{y} - QX\hat{\boldsymbol{b}} = \overbrace{\begin{pmatrix} \overline{y} \\ \overline{y} \\ \vdots \\ \overline{y} \end{pmatrix}}^{Q\boldsymbol{y}} - \overbrace{\begin{pmatrix} \overline{x}_1 & \overline{x}_2 \\ \overline{x}_1 & \overline{x}_2 \\ \vdots & \vdots \\ \overline{x}_1 & \overline{x}_2 \end{pmatrix}}^{QX}\overbrace{\begin{pmatrix} \hat{b}_1 \\ \hat{b}_2 \end{pmatrix}}^{\hat{\boldsymbol{b}}},$$

つまり $\hat{c} = \overline{y} - (\overline{x}_1\hat{b}_1 + \overline{x}_2\hat{b}_2) = \overline{y} - \left\langle \begin{pmatrix} \overline{x}_1 \\ \overline{x}_2 \end{pmatrix}, \begin{pmatrix} \hat{b}_1 \\ \hat{b}_2 \end{pmatrix} \right\rangle$ となり，公式 4.18 (➔ p. 118) が示された。

命題 4.20(➔ p. 119), 命題 4.21(➔ p. 120) の証明

補題 2 重回帰係数 $\hat{\boldsymbol{b}}$ と $\hat{c}$ に対して，(i) $P\hat{\boldsymbol{y}}$ と $P\hat{\boldsymbol{\varepsilon}}$ は直交する。　(ii) $\hat{\boldsymbol{\varepsilon}} = P\hat{\boldsymbol{\varepsilon}}$

証明. (i) 補題 1 (➔ p. 3) より，二つのベクトル

$$\mathbf{v} = PX(X^\top PX)^{-1}X^\top P\boldsymbol{y}, \quad \mathbf{w} = P\boldsymbol{y} - PX(X^\top PX)^{-1}X^\top P\boldsymbol{y}$$

は直交するのであった。式 (A.1) (➔ p. 4) に注意すると，

$$\mathbf{v} = PX\underbrace{(X^\top PX)^{-1}X^\top P\boldsymbol{y}}_{=\hat{\boldsymbol{b}}} = PX\hat{\boldsymbol{b}} \overset{\substack{P\hat{\boldsymbol{c}}=\boldsymbol{0}\\ \text{が成り立つ}\\ \text{のだった。}}}{=} P(\underbrace{X\hat{\boldsymbol{b}}+\hat{\boldsymbol{c}}}_{=\hat{\boldsymbol{y}}}) = P\hat{\boldsymbol{y}},$$

$$\begin{aligned}\mathbf{w} &= P\boldsymbol{y} - PX\underbrace{(X^\top PX)^{-1}X^\top P\boldsymbol{y}}_{=\hat{\boldsymbol{b}}} = P\boldsymbol{y} - PX\hat{\boldsymbol{b}} \\ &= P(\boldsymbol{y} - X\hat{\boldsymbol{b}}) \overset{\substack{P\hat{\boldsymbol{c}}=\boldsymbol{0}\\ \text{が成り立つ}\\ \text{のだった。}}}{=} P\big(\underbrace{\boldsymbol{y} - (\underbrace{X\hat{\boldsymbol{b}}+\hat{\boldsymbol{c}}}_{=\hat{\boldsymbol{y}}})}_{=\hat{\boldsymbol{\varepsilon}}}\big) = P\hat{\boldsymbol{\varepsilon}}\end{aligned}$$

となり，ゆえに $P\hat{\boldsymbol{y}}$ と $P\hat{\boldsymbol{\varepsilon}}$ は直交する。
(ii) $\hat{\boldsymbol{\varepsilon}} = \boldsymbol{y} - \hat{\boldsymbol{y}} = \boldsymbol{y} - (X\hat{\boldsymbol{b}} + \hat{\boldsymbol{c}}) = \boldsymbol{y} - \big(X\hat{\boldsymbol{b}} + (Q\boldsymbol{y} - QX\hat{\boldsymbol{b}})\big) = (E_N - Q)\boldsymbol{y} - (E_N - Q)X\hat{\boldsymbol{b}} = P\boldsymbol{y} - PX\hat{\boldsymbol{b}} = P(\boldsymbol{y} - X\hat{\boldsymbol{b}}) = P(\boldsymbol{y} - (X\hat{\boldsymbol{b}} + \hat{\boldsymbol{c}})) = P(\boldsymbol{y} - \hat{\boldsymbol{y}}) = P\hat{\boldsymbol{\varepsilon}}$ □

命題 4.20 の証明. 命題 4.5–(2) (➔ p. 108) より $\|\hat{\varepsilon}\|^2 = \|P\hat{\varepsilon}\|^2 + \|Q\hat{\varepsilon}\|^2$ である一方で，補題 2–(ii) より $P\hat{\boldsymbol{\varepsilon}} = \hat{\boldsymbol{\varepsilon}}$ となる。ゆえに $\|Q\hat{\varepsilon}\|^2 = 0$，つまり $\overline{\hat{\varepsilon}} = \frac{1}{N}\sum_{i=1}^{N}\hat{\varepsilon}_i = 0$ でなければならない。 □

命題 4.21 の証明. $\boldsymbol{y} = \hat{\boldsymbol{y}} + \hat{\boldsymbol{\varepsilon}}$ であることを思い出すと，次の式変形が確かめられる。

$$\begin{aligned} s(\boldsymbol{y},\boldsymbol{y}) &\overset{\substack{\text{命題 4.12–(3)}\\ (\text{➔ p. 112})}}{=} \frac{1}{N}\langle \boldsymbol{y}, P\boldsymbol{y}\rangle = \frac{1}{N}\langle \hat{\boldsymbol{y}}+\hat{\boldsymbol{\varepsilon}}, P(\hat{\boldsymbol{y}}+\hat{\boldsymbol{\varepsilon}})\rangle \\ &= \frac{1}{N}\langle \hat{\boldsymbol{y}}, P\hat{\boldsymbol{y}}\rangle + \frac{2}{N}\underbrace{\langle \hat{\boldsymbol{y}}, P\hat{\boldsymbol{\varepsilon}}\rangle}_{\substack{\| \\ \langle P\hat{\boldsymbol{y}}, P\hat{\boldsymbol{\varepsilon}}\rangle \\ \| \\ 0}} + \frac{1}{N}\langle \hat{\boldsymbol{\varepsilon}}, P\hat{\boldsymbol{\varepsilon}}\rangle \overset{\substack{\text{命題 4.12–(3)}\\ (\text{➔ p. 112})}}{=} s(\hat{\boldsymbol{y}},\hat{\boldsymbol{y}}) + s(\hat{\boldsymbol{\varepsilon}},\hat{\boldsymbol{\varepsilon}}) \end{aligned}$$

□

命題 4.23(➔ p. 120) の証明

まず，$s(\boldsymbol{y}, \hat{\boldsymbol{y}})$ を次のように計算しておく。

$$\begin{aligned} s(\boldsymbol{y}, \hat{\boldsymbol{y}}) &\overset{\substack{\text{命題 4.12–(3)}\\(\text{➔ p. 112})}}{=} \frac{1}{N}\langle \boldsymbol{y}, P\hat{\boldsymbol{y}}\rangle = \frac{1}{N}\langle \hat{\boldsymbol{y}} + \hat{\boldsymbol{\varepsilon}}, P\hat{\boldsymbol{y}}\rangle = \frac{1}{N}\langle \hat{\boldsymbol{y}}, P\hat{\boldsymbol{y}}\rangle + \frac{1}{N}\langle \hat{\boldsymbol{\varepsilon}}, P\hat{\boldsymbol{y}}\rangle \\ &\overset{\substack{\text{命題 4.5–(1)}\\(\text{➔ p. 108})}}{=} \frac{1}{N}\langle \hat{\boldsymbol{y}}, P\hat{\boldsymbol{y}}\rangle + \frac{1}{N}\underbrace{\langle P\hat{\boldsymbol{\varepsilon}}, P\hat{\boldsymbol{y}}\rangle}_{=0} = \frac{1}{N}\langle \hat{\boldsymbol{y}}, P\hat{\boldsymbol{y}}\rangle \overset{\substack{\text{命題 4.5–(1)}\\(\text{➔ p. 108})}}{=} \frac{1}{N}\|P\hat{\boldsymbol{y}}\|^2 \end{aligned} \tag{A.2}$$

そこで R^2 の定義に基づくと，R は次のように計算できる。

$$\begin{aligned} R &= \frac{\sqrt{s(\hat{\boldsymbol{y}}, \hat{\boldsymbol{y}})}}{\sqrt{s(\boldsymbol{y}, \boldsymbol{y})}} \overset{\substack{\text{命題 4.12–(3)}\\(\text{➔ p. 112})}}{=} \frac{\sqrt{\frac{1}{N}\langle \hat{\boldsymbol{y}}, P\hat{\boldsymbol{y}}\rangle^2}}{\sqrt{\frac{1}{N}\langle \boldsymbol{y}, P\boldsymbol{y}\rangle^2}} \\ &\overset{\substack{\text{命題 4.5–(1)}\\(\text{➔ p. 108})}}{=} \frac{\sqrt{\|P\hat{\boldsymbol{y}}\|^2}}{\sqrt{\|P\boldsymbol{y}\|^2}} = \frac{\|P\hat{\boldsymbol{y}}\|}{\|P\boldsymbol{y}\|} = \frac{\|P\hat{\boldsymbol{y}}\|^2}{\|P\boldsymbol{y}\|\|P\hat{\boldsymbol{y}}\|} \overset{\text{(A.2)}}{=} \frac{s(\hat{\boldsymbol{y}}, \boldsymbol{y})}{\sqrt{s(\boldsymbol{y}, \boldsymbol{y})}\sqrt{s(\hat{\boldsymbol{y}}, \hat{\boldsymbol{y}})}} \end{aligned}$$

定理 4.30(➔ p. 130) の証明

ある母集団から N 個の標本点 $\omega_1, \omega_2, \ldots, \omega_N$ を取り，それぞれについて 2 種類の変量 X_1, X_2 が取った値を調べたときに

標本点 \ 変量	X_1	X_2
ω_1	x_{11}	x_{12}
ω_2	x_{21}	x_{22}
$\vdots$	$\vdots$	$\vdots$
ω_N	x_{N1}	x_{N2}

$$X = \begin{pmatrix} x_{11} & x_{12} \\ x_{21} & x_{22} \\ \vdots & \vdots \\ x_{N1} & x_{N2} \end{pmatrix} = \begin{pmatrix} \boldsymbol{x}_1 & \boldsymbol{x}_2 \end{pmatrix}$$

というデータを得たとする。このとき関数

$$f : \{\mathbf{w} \in \mathbb{R}^2 : \|\mathbf{w}\|^2 = 1\} \ni \mathbf{w} \mapsto f(\mathbf{w}) = \underbrace{(X\mathbf{w} \text{ の「情報量」})}_{= s(X\mathbf{w}, X\mathbf{w})} \in \mathbb{R}$$

が最大値をとる点 (ベクトル) $\mathbf{w}_1 = \begin{pmatrix} w_{11} \\ w_{21} \end{pmatrix}$ を見つけたいのであった。

Lagrange の未定乗数法 (定理 3.10 ➔ p. 100) を用いるために，関数 $f, g : \mathbb{R}^2 \to \mathbb{R}$ を $\mathbf{w} = \begin{pmatrix} w_1 \\ w_2 \end{pmatrix} \in \mathbb{R}^2$ に対して次により定める。

$$\begin{aligned} f(\mathbf{w}) &= (X\mathbf{w} \text{ の「情報量」}) = s(X\mathbf{w}, X\mathbf{w}) = \langle \mathbf{w}, S_{X,X}\mathbf{w}\rangle \\ &= s(\boldsymbol{x}_1, \boldsymbol{x}_1)(w_1)^2 + 2s(\boldsymbol{x}_1, \boldsymbol{x}_2)w_1 w_2 + s(\boldsymbol{x}_2, \boldsymbol{x}_2)(w_2)^2, \\ g(\mathbf{w}) &= \|\mathbf{w}\|^2 - 1 = (w_1)^2 + (w_2)^2 - 1 \end{aligned}$$

このとき $\|\mathbf{w}\|^2 = 1 \Leftrightarrow g(\mathbf{w}) = 0$ であるから $\{\mathbf{w} \in \mathbb{R}^2 : \|\mathbf{w}\|^2 = 1\} = \{\mathbf{w} \in \mathbb{R}^2 : g(\mathbf{w}) = 0\}$ である。そこで領域 $\{\mathbf{w} \in \mathbb{R}^2 : g(\mathbf{w}) = 0\}$ 上での関数 f の極値点の候補をすべてを求めよう。このためには，

Lagrange の未定乗数法 (➔定理 3.10, p. 100) より，新たな変数 $\lambda \in \mathbb{R}$ を導入して $(2+1)$ 個の式からなる次の連立方程式を $(\mathbf{w}^*, \lambda)$ に関して解けばよい。

$$\begin{cases} \dfrac{\partial f}{\partial w_1}(\mathbf{w}^*) - \lambda \dfrac{\partial g}{\partial w_1}(\mathbf{w}^*) = 0, \\ \dfrac{\partial f}{\partial w_2}(\mathbf{w}^*) - \lambda \dfrac{\partial g}{\partial w_2}(\mathbf{w}^*) = 0, \\ g(\mathbf{w}^*) = 0 \end{cases} \tag{A.3}$$

このために，(A.3) に現れる偏微分を計算しておくと，次を得る。

$$\begin{aligned} \frac{\partial f}{\partial w_1}(\mathbf{w}) &= \frac{\partial}{\partial w_1}\Big(s(\boldsymbol{x}_1, \boldsymbol{x}_1)(w_1)^2 + 2s(\boldsymbol{x}_1, \boldsymbol{x}_2) w_1 w_2 + s(\boldsymbol{x}_2, \boldsymbol{x}_2)(w_2)^2 \Big) \\ &= 2s(\boldsymbol{x}_1, \boldsymbol{x}_1) w_1 + 2s(\boldsymbol{x}_1, \boldsymbol{x}_2) w_2 = 2(S_{X,X}\mathbf{w})_1, \\ \frac{\partial f}{\partial w_2}(\mathbf{w}) &= 2s(\boldsymbol{x}_1, \boldsymbol{x}_2) w_1 + 2s(\boldsymbol{x}_2, \boldsymbol{x}_2) w_2 = 2(S_{X,X}\mathbf{w})_2, \\ \frac{\partial g}{\partial w_1}(\mathbf{w}) &= \frac{\partial}{\partial w_1}\Big((w_1)^2 + (w_2)^2 - 1 \Big) = 2w_1, \quad \frac{\partial g}{\partial w_2}(\mathbf{w}) = 2w_2 \end{aligned}$$

これらを (A.3) に代入すれば，

$$\begin{cases} 2(S_{X,X}\mathbf{w}^*)_1 - \lambda 2(\mathbf{w}^*)_1 = 0, \\ 2(S_{X,X}\mathbf{w}^*)_2 - \lambda 2(\mathbf{w}^*)_2 = 0, \\ \|\mathbf{w}^*\|^2 = 1 \end{cases}$$

であり，以上から，関数 $f(\mathbf{w})$ が $\mathbf{w} = \mathbf{w}^*$ において最大値をとるならば，

$$S_{X,X}\mathbf{w}^* = \lambda \mathbf{w}^*, \tag{A.4}$$

$$\|\mathbf{w}^*\|^2 = 1 \tag{A.5}$$

が成り立つことがわかったのである。式 (A.5) は $\mathbf{w}^*$ が単位ベクトルであることを表しており，式 (A.4) は λ は共分散行列 $S_{X,X}$ の固有値であり，$\mathbf{w}^*$ はその固有ベクトルであることを意味している。また，その最大値は $f(\mathbf{w}^*) = \langle \mathbf{w}^*, S_{X,X}\mathbf{w}^* \rangle = \langle \mathbf{w}^*, \lambda \mathbf{w}^* \rangle = \lambda \|\mathbf{w}^*\|^2 = \lambda$ となる。以上のことを踏まえて，証明に入る。

定理 4.30 の証明. (1)⇒(2): ベクトル $\mathbf{w}_1 \in \mathbb{R}^2$ は $\|\mathbf{w}_1\|^2 = 1$ をみたすとする。また，共分散行列 $S_{X,X}$ の最大固有値を λ_1 とし，ベクトル $\boldsymbol{u} \in \mathbb{R}^2$ は固有値 λ_1 に関する $S_{X,X}$ の固有ベクトルであり，$\|\boldsymbol{u}\|^2 = 1$ をみたすとする。
関数 $f(\mathbf{w})$ が $\mathbf{w} = \mathbf{w}_1$ において最大値をとるとすると，上で見たようにある実数 λ について $S_{X,X}\mathbf{w}_1 = \lambda \mathbf{w}_1$ が成り立つのだった。このとき，

$$\begin{aligned} \lambda_1 = \lambda_1 \|\boldsymbol{u}\|^2 = \langle \boldsymbol{u}, \lambda_1 \boldsymbol{u} \rangle &= \langle \boldsymbol{u}, S_{X,X}\boldsymbol{u} \rangle \\ &\leqq \max_{\substack{\mathbf{w} \in \mathbb{R}^2: \\ \|\mathbf{w}\|^2 = 1}} \underbrace{\langle \mathbf{w}, S_{X,X}\mathbf{w} \rangle}_{= f(\mathbf{w})} \\ &= f(\mathbf{w}_1) = \langle \mathbf{w}_1, S_{X,X}\mathbf{w}_1 \rangle = \underbrace{\langle \mathbf{w}_1, \lambda \mathbf{w}_1 \rangle}_{= \lambda \|\mathbf{w}_1\|^2} = \lambda \end{aligned} \tag{A.6}$$

が成り立つ。つまり $S_{X,X}$ の固有値 λ が $\lambda \geqq \lambda_1$ をみたさなければならないが，一方で λ_1 は $S_{X,X}$ の固有値の中で最も大きいから $\lambda \leqq \lambda_1$，よって $\lambda = \lambda_1$ が成り立つ。ゆえに $S_{X,X}\mathbf{w}_1 = \lambda \mathbf{w}_1 = \lambda_1 \mathbf{w}_1$，つまり $\mathbf{w}_1$ は λ_1 に関する固有ベクトルである。
(2)⇒(1): 共分散行列 $S_{X,X}$ の最大固有値を λ_1 とし，$\mathbf{w}_1$ をその固有ベクトルで $\|\mathbf{w}_1\|^2 = 1$ をみたす

ものとする。上で見たことから，

$$\begin{aligned}\max_{\substack{\mathbf{w}\in\mathbb{R}^2:\\ \|\mathbf{w}\|^2=1}} f(\mathbf{w}) &= (S_{X,X}\text{ の固有値のうちのいずれか})\\ &\leqq \lambda_1 = \langle \mathbf{w}_1, S_{X,X}\mathbf{w}_1\rangle = f(\mathbf{w}_1) \leqq \max_{\substack{\mathbf{w}\in\mathbb{R}^2:\\ \|\mathbf{w}\|^2=1}} f(\mathbf{w})\end{aligned}$$

であり，ゆえに上の不等式はすべて等号で結ばれる。特に $\max_{\substack{\mathbf{w}\in\mathbb{R}^2:\\ \|\mathbf{w}\|^2=1}} f(\mathbf{w}) = f(\mathbf{w}_1)$ となり，これで関数 $f(\mathbf{w})$ が $\mathbf{w}=\mathbf{w}_1$ において最大値をとることが示された。

□

命題 4.32(➲ p. 132)，命題 4.33(➲ p. 133) の証明

データ行列

$$X = \begin{pmatrix} x_{11} & x_{12} \\ x_{21} & x_{22} \\ \vdots & \vdots \\ x_{N1} & x_{N2} \end{pmatrix} = \begin{pmatrix} \boldsymbol{x}_1 & \boldsymbol{x}_2 \end{pmatrix}$$

の共分散行列 (**定義 4.5–(4)** ➲ **p. 110**) を $S_{X,X} = \begin{pmatrix} s(\boldsymbol{x}_1,\boldsymbol{x}_1) & s(\boldsymbol{x}_1,\boldsymbol{x}_2) \\ s(\boldsymbol{x}_2,\boldsymbol{x}_1) & s(\boldsymbol{x}_2,\boldsymbol{x}_2) \end{pmatrix}$ で表していたのだった。$S_{X,X}$ の固有値の中でも最大のものを λ_1 とし，ベクトル $\mathbf{w}_1 \in \mathbb{R}^2$ は固有値 λ_1 に関する $S_{X,X}$ の固有ベクトルであり，かつ $\|\mathbf{w}_1\|^2 = 1$ をみたすとする。2 次単位行列を $E_2 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$ で表していたことを思い出そう (**定義 B.1–(ii)** ➲ **p. 1**)。

補題 3

(i) ($S_{X(E_2-\mathbf{w}_1\otimes\mathbf{w}_1),X(E_2-\mathbf{w}_1\otimes\mathbf{w}_1)}$ と $S_{X,X}$ の関係式) 次が成り立つ。

$$\begin{aligned}S_{X(E_2-\mathbf{w}_1\otimes\mathbf{w}_1),X(E_2-\mathbf{w}_1\otimes\mathbf{w}_1)} &= S_{X,X} - S_{X,X}(\mathbf{w}_1\otimes\mathbf{w}_1)\\ &\quad - (\mathbf{w}_1\otimes\mathbf{w}_1)S_{X,X} + (\mathbf{w}_1\otimes\mathbf{w}_1)S_{X,X}(\mathbf{w}_1\otimes\mathbf{w}_1)\end{aligned}$$

(ii) 任意の $\mathbf{v}\in\mathbb{R}^2$ に対して，次が成り立つ。

$$S_{X(E_2-\mathbf{w}_1\otimes\mathbf{w}_1),X(E_2-\mathbf{w}_1\otimes\mathbf{w}_1)}\mathbf{v} = S_{X,X}\mathbf{v} - \lambda_1\langle \mathbf{v},\mathbf{w}_1\rangle\mathbf{w}_1$$

証明. (i) 命題 4.13–(2) (➲ **p. 113**) を用いると，次の式変形が確かめられる。

$$\begin{aligned}&S_{X(E_2-\mathbf{w}_1\otimes\mathbf{w}_1),X(E_2-\mathbf{w}_1\otimes\mathbf{w}_1)}\\ &= (E_2-\mathbf{w}_1\otimes\mathbf{w}_1)^\top S_{X,X}(E_2-\mathbf{w}_1\otimes\mathbf{w}_1)\\ &= (S_{X,X}-(\mathbf{w}_1\otimes\mathbf{w}_1)S_{X,X})(E_2-\mathbf{w}_1\otimes\mathbf{w}_1)\\ &= S_{X,X} - S_{X,X}(\mathbf{w}_1\otimes\mathbf{w}_1) - (\mathbf{w}_1\otimes\mathbf{w}_1)S_{X,X} + (\mathbf{w}_1\otimes\mathbf{w}_1)S_{X,X}(\mathbf{w}_1\otimes\mathbf{w}_1)\end{aligned}$$

(ii) $\mathbf{v}$ を任意とすると，(i) を用いて

$$\begin{aligned}&S_{X(E_2-\mathbf{w}_1\otimes\mathbf{w}_1),X(E_2-\mathbf{w}_1\otimes\mathbf{w}_1)}\mathbf{v}\\ &= S_{X,X}\mathbf{v} - S_{X,X}(\mathbf{w}_1\otimes\mathbf{w}_1)\mathbf{v} - (\mathbf{w}_1\otimes\mathbf{w}_1)S_{X,X}\mathbf{v} + (\mathbf{w}_1\otimes\mathbf{w}_1)S_{X,X}(\mathbf{w}_1\otimes\mathbf{w}_1)\mathbf{v}\end{aligned}$$

となる。このうち,

$$(\text{右辺第 2 項}) = -S_{X,X}(\mathbf{w}_1 \otimes \mathbf{w}_1)\mathbf{v} \overset{\substack{\text{命題 4.7–(a)}\\ \text{(➡ p. 109)}}}{=} -\langle \mathbf{w}_1, \mathbf{v}\rangle \underbrace{S_{X,X}\mathbf{w}_1}_{=\lambda_1 \mathbf{w}_1} = -\lambda_1 \langle \mathbf{w}_1, \mathbf{v}\rangle \mathbf{w}_1,$$

$$\begin{aligned}(\text{右辺第 3 項}) &= -(\mathbf{w}_1 \otimes \mathbf{w}_1)S_{X,X}\mathbf{v} \overset{\substack{\text{命題 4.7–(a)}\\ \text{(➡ p. 109)}}}{=} -\langle \mathbf{w}_1, S_{X,X}\mathbf{v}\rangle \mathbf{w}_1 \\ &\overset{\substack{\text{公式 B.51–(1)(➡ p. 26)}\\ \text{と } S_{X,X} \text{ の実対称性}}}{=} -\langle \underbrace{S_{X,X}\mathbf{w}_1}_{=\lambda_1 \mathbf{w}_1}, \mathbf{v}\rangle \mathbf{w}_1 = -\lambda_1 \langle \mathbf{w}_1, \mathbf{v}\rangle \mathbf{w}_1,\end{aligned}$$

$$\begin{aligned}(\text{右辺第 4 項}) &= (\mathbf{w}_1 \otimes \mathbf{w}_1)S_{X,X}(\mathbf{w}_1 \otimes \mathbf{w}_1)\mathbf{v} \\ &= \langle \mathbf{w}_1, \mathbf{v}\rangle (\mathbf{w}_1 \otimes \mathbf{w}_1)\underbrace{S_{X,X}\mathbf{w}_1}_{=\lambda_1 \mathbf{w}_1} \overset{\substack{\text{命題 4.7–(c)}\\ \text{(➡ p. 109)}}}{=} \lambda_1 \langle \mathbf{w}_1, \mathbf{v}\rangle \mathbf{w}_1\end{aligned}$$

となる。以上から，次を得る。

$$\begin{aligned}S_{X(E_2-\mathbf{w}_1\otimes\mathbf{w}_1),X(E_2-\mathbf{w}_1\otimes\mathbf{w}_1)}\mathbf{v} &= S_{X,X}\mathbf{v} - \lambda_1\langle \mathbf{w}_1, \mathbf{v}\rangle \mathbf{w}_1 - \lambda_1\langle \mathbf{w}_1, \mathbf{v}\rangle \mathbf{w}_1 + \lambda_1\langle \mathbf{w}_1, \mathbf{v}\rangle \mathbf{w}_1 \\ &= S_{X,X}\mathbf{v} - \lambda_1\langle \mathbf{w}_1, \mathbf{v}\rangle \mathbf{w}_1\end{aligned}$$

□

命題 4.32 の証明. 補題 3–(i) を用いて計算すると

$$\begin{aligned}&\mathrm{tr}(S_{X(E_2-\mathbf{w}_1\otimes\mathbf{w}_1),X(E_2-\mathbf{w}_1\otimes\mathbf{w}_1)}) \\ &= \mathrm{tr}(S_{X,X}) - \mathrm{tr}\big(S_{X,X}(\mathbf{w}_1 \otimes \mathbf{w}_1)\big) \\ &\qquad - \mathrm{tr}\big((\mathbf{w}_1 \otimes \mathbf{w}_1)S_{X,X}\big) + \mathrm{tr}\big((\mathbf{w}_1 \otimes \mathbf{w}_1)S_{X,X}(\mathbf{w}_1 \otimes \mathbf{w}_1)\big)\end{aligned}$$

となる。このうち

$$\begin{aligned}(\text{右辺第 3 項}) &= -\mathrm{tr}\big((\mathbf{w}_1 \otimes \mathbf{w}_1)S_{X,X}\big) \overset{\substack{\text{練習問題 4.2}\\ \text{(➡ p. 107)}}}{=} -\mathrm{tr}\big(S_{X,X}(\mathbf{w}_1 \otimes \mathbf{w}_1)\big), \\ (\text{右辺第 4 項}) &= \mathrm{tr}\big((\mathbf{w}_1 \otimes \mathbf{w}_1)S_{X,X}(\mathbf{w}_1 \otimes \mathbf{w}_1)\big) \\ &\overset{\substack{\text{練習問題 4.2}\\ \text{(➡ p. 107)}}}{=} \mathrm{tr}\big(S_{X,X}(\mathbf{w}_1 \otimes \mathbf{w}_1)^2\big) \overset{\substack{\text{命題 4.7–(d)}\\ \text{(➡ p. 109)}}}{=} \mathrm{tr}\big(S_{X,X}(\mathbf{w}_1 \otimes \mathbf{w}_1)\big)\end{aligned}$$

となり，これらは互いに打ち消しあうから，次式を得る。

$$\mathrm{tr}\big(S_{X(E_2-\mathbf{w}_1\otimes\mathbf{w}_2),X(E_2-\mathbf{w}_1\otimes\mathbf{w}_2)}\big) = \mathrm{tr}(S_{X,X}) - \mathrm{tr}\big(S_{X,X}(\mathbf{w}_1 \otimes \mathbf{w}_1)\big)$$

最後に，この右辺第 2 項について

$$S_{X,X}(\mathbf{w}_1 \otimes \mathbf{w}_1) \overset{\substack{\text{定義 4.3}\\ \text{(➡ p. 108)}}}{=} S_{X,X}(\mathbf{w}_1 \mathbf{w}_1^\top) = (S_{X,X}\mathbf{w}_1)\mathbf{w}_1^\top \overset{\substack{\text{定義 4.3}\\ \text{(➡ p. 108)}}}{=} (S_{X,X}\mathbf{w}_1) \otimes \mathbf{w}_1$$

であり，$\mathbf{f}_1 = X\mathbf{w}_1$ だったことを思い出すと，次が成り立つ。

$$\operatorname{tr}\big(S_{X,X}(\mathbf{w}_1 \otimes \mathbf{w}_1)\big) = \operatorname{tr}\big((S_{X,X}\mathbf{w}_1) \otimes \mathbf{w}_1\big) \overset{\substack{\text{命題 4.7–(b)}\\ (\text{➲ p. 109})}}{=} \langle S_{X,X}\mathbf{w}_1, \mathbf{w}_1\rangle \overset{\substack{\text{命題 4.13–(3)}\\ (\text{➲ p. 113})}}{=} s(X\mathbf{w}_1, X\mathbf{w}_1) = s(\mathbf{f}_1, \mathbf{f}_1)$$

以上から $\operatorname{tr}\big(S_{X(E_2-\mathbf{w}_1\otimes\mathbf{w}_1),X(E_2-\mathbf{w}_1\otimes\mathbf{w}_1)}\big) = \operatorname{tr}(S_{X,X}) - s(\mathbf{f}_1, \mathbf{f}_1)$ である。

□

以下，命題 4.33 (➲ p. 133) の直前における設定のように，$S_{X,X}$ の固有値を大きい順に $\lambda_1 \geqq \lambda_2$ $(\geqq 0)$ としておいて，それぞれの固有ベクトル $\mathbf{w}_1, \mathbf{w}_2$ は $R = (\mathbf{w}_1, \mathbf{w}_2)$ が 2 次直交行列となるように選んでおいたことを思い出そう。

命題 4.33 の証明．まず「データ行列」$X(E_2 - \mathbf{w}_1 \otimes \mathbf{w}_1)$ の共分散行列 $S_{X(E_2-\mathbf{w}_1\otimes\mathbf{w}_1),X(E_2-\mathbf{w}_1\otimes\mathbf{w}_1)}$ の固有値を調べよう。補題 3–(ii) (➲ p. 8) より，$k = 1, 2$ に対して

$$\begin{aligned} S_{X(E_2-\mathbf{w}_1\otimes\mathbf{w}_1),X(E_2-\mathbf{w}_1\otimes\mathbf{w}_1)}\mathbf{w}_k &= \underbrace{S_{X,X}\mathbf{w}_k}_{=\lambda_k\mathbf{w}_k} - \lambda_1 \underbrace{\langle\mathbf{w}_k, \mathbf{w}_1\rangle}_{\substack{k=1\text{ のときは }1\text{ で}\\ k=2\text{ のときには }0}} \mathbf{w}_1 \\ &= \begin{cases} \mathbf{0}\,(=0\,\mathbf{w}_1) & \text{if } k=1, \\ \lambda_k\mathbf{w}_k & \text{if } k=2 \end{cases} \end{aligned}$$

であり，これは行列 $S_{X(E_2-\mathbf{w}_1\otimes\mathbf{w}_1),X(E_2-\mathbf{w}_1\otimes\mathbf{w}_1)}$ が直交行列 $R = (\mathbf{w}_1, \mathbf{w}_2)$ により

$$R^\top S_{X(E_2-\mathbf{w}_1\otimes\mathbf{w}_1),X(E_2-\mathbf{w}_1\otimes\mathbf{w}_1)} R = \begin{pmatrix} 0 & 0 \\ 0 & \lambda_2 \end{pmatrix}$$

という形に対角化されることを意味している。ゆえに $S_{X(E_2-\mathbf{w}_1\otimes\mathbf{w}_1),X(E_2-\mathbf{w}_1\otimes\mathbf{w}_1)}$ の固有値は λ_2 と 0 のみである ($\lambda_2 = 0$ となっていることもありうる)。

いま「データ行列」$X(E_2 - \mathbf{w}_1 \otimes \mathbf{w}_1)$ に対して定理 4.30 (➲ p. 130) を適用すると，次を得る。

$$\max_{\substack{\mathbf{w}\in\mathbb{R}^2:\\ \|\mathbf{w}\|^2=1}} \begin{pmatrix} X(E_2 - \mathbf{w}_1 \otimes \mathbf{w}_1)\mathbf{w} \\ \text{の「情報量」} \end{pmatrix} = \big(X(E_2 - \mathbf{w}_1 \otimes \mathbf{w}_1)\mathbf{w}_2 \text{ の「情報量」}\big)$$

ここで，$\langle\mathbf{w}_1, \mathbf{w}_2\rangle = 0$ であることから $(\mathbf{w}_1 \otimes \mathbf{w}_1)\mathbf{w}_2 = \langle\mathbf{w}_1, \mathbf{w}_2\rangle\mathbf{w}_1 = 0 \cdot \mathbf{w}_1 = \mathbf{0}$，ゆえに $X(E_2 - \mathbf{w}_1 \otimes \mathbf{w}_1)\mathbf{w}_2 = X\mathbf{w}_2 - (\mathbf{w}_1 \otimes \mathbf{w}_1)\mathbf{w}_2 = X\mathbf{w}_2 = \mathbf{f}_2$ となるので，次が成り立つ。

$$\begin{aligned} \max_{\substack{\mathbf{w}\in\mathbb{R}^2:\\ \|\mathbf{w}\|^2=1}} \begin{pmatrix} X(E_2 - \mathbf{w}_1 \otimes \mathbf{w}_1)\mathbf{w} \\ \text{の「情報量」} \end{pmatrix} &= (\mathbf{f}_2 \text{ の「情報量」}) \\ &= s(\mathbf{f}_2, \mathbf{f}_2) \overset{\substack{\text{命題 4.13–(3)}\\ (\text{➲ p. 113})}}{=} \langle\mathbf{w}_2,\ \underbrace{S_{X,X}\mathbf{w}_2}_{=\lambda_2\mathbf{w}_2}\ \rangle = \lambda_2 \end{aligned}$$

□

命題 4.38(➲ p. 140) の証明

一般に，自然数 k に対して，Q_k は k 次の平均化行列，P_k は k 次の中心化行列を表すことにする (定義 4.6 ➲ p. 112)。

補題 4 $P_N = \left(\begin{array}{c:c} P_n & O \\ \hdashline O & P_m \end{array}\right) + \left(\begin{array}{c:c} Q_n & O \\ \hdashline O & Q_m \end{array}\right) - Q_N$

証明. 1 の直交分割 (**定義 4.2** **➔p. 108**) の性質を用いて

$$\begin{aligned} P_N + Q_N = E_N = \left(\begin{array}{c:c} E_n & O \\ \hdashline O & E_m \end{array}\right) &= \left(\begin{array}{c:c} P_n + Q_n & O \\ \hdashline O & P_m + Q_m \end{array}\right) \\ &= \left(\begin{array}{c:c} P_n & O \\ \hdashline O & P_m \end{array}\right) + \left(\begin{array}{c:c} Q_n & O \\ \hdashline O & Q_m \end{array}\right) \end{aligned}$$

となる。この最左辺の Q_N を最右辺に移項することで従う。 □

補題 5 $Q_N \left(\begin{array}{c:c} Q_n & O \\ \hdashline O & Q_m \end{array}\right) = Q_N$ であり，特に $Q_N X = Q_N Y$ が成り立つ。

証明. 自然数 k に対して $\mathbf{1}_k = (\underbrace{1, 1, \ldots, 1}_{k\text{ 個}})^\top$ とおくと

$$Q_N = \left(\begin{array}{c:c} \frac{n}{N} Q_n & \frac{1}{N} \mathbf{1}_n \otimes \mathbf{1}_m \\ \hdashline \frac{1}{N} \mathbf{1}_n \otimes \mathbf{1}_m & \frac{m}{N} Q_m \end{array}\right)$$

と表せる。これを用いて，次の式変形が確かめられる。

$$\begin{aligned} Q_N \left(\begin{array}{c:c} Q_n & O \\ \hdashline O & Q_m \end{array}\right) &= \left(\begin{array}{c:c} \frac{n}{N} Q_n & \frac{1}{N} \mathbf{1}_n \otimes \mathbf{1}_m \\ \hdashline \frac{1}{N} \mathbf{1}_m \otimes \mathbf{1}_n & \frac{m}{N} Q_m \end{array}\right) \left(\begin{array}{c:c} Q_n & O \\ \hdashline O & Q_m \end{array}\right) \\ &= \left(\begin{array}{c:c} \frac{n}{N} Q_n^2 & \frac{1}{N} (\mathbf{1}_n \otimes \mathbf{1}_m) Q_m \\ \hdashline \frac{1}{N} (\mathbf{1}_m \otimes \mathbf{1}_n) Q_n & \frac{m}{N} Q_m^2 \end{array}\right) \end{aligned}$$

ここで Q_n，Q_m は直交射影であるから，$Q_n^2 = Q_n$，$Q_m^2 = Q_m$ が成り立つ。$Q_n \mathbf{1}_n = \mathbf{1}_n$ より，$(\mathbf{1}_m \otimes \mathbf{1}_n) Q_n = \mathbf{1}_m \mathbf{1}_n^\top Q_n = \mathbf{1}_m (Q_n \mathbf{1}_n)^\top = \mathbf{1}_m \otimes \mathbf{1}_n$ となる。同様にして $(\mathbf{1}_n \otimes \mathbf{1}_m) Q_m = (\mathbf{1}_n \otimes \mathbf{1}_m)$ となるから，上式は Q_N に等しい。
また，$Y = \left(\begin{array}{c} Q_n X_{\mathrm{A}} \\ \hdashline Q_m X_{\mathrm{B}} \end{array}\right) = \left(\begin{array}{c:c} Q_n & O \\ \hdashline O & Q_m \end{array}\right) \left(\begin{array}{c} X_{\mathrm{A}} \\ \hdashline X_{\mathrm{B}} \end{array}\right)$ が成り立つから，次が成り立つ。

$$Q_N Y = Q_N \left(\begin{array}{c:c} Q_n & O \\ \hdashline O & Q_m \end{array}\right) \left(\begin{array}{c} X_{\mathrm{A}} \\ \hdashline X_{\mathrm{B}} \end{array}\right) = Q_N \left(\begin{array}{c} X_{\mathrm{A}} \\ \hdashline X_{\mathrm{B}} \end{array}\right) = Q_N X$$

□

補題 6 $\left(\begin{array}{c:c} P_n & O \\ \hdashline O & P_m \end{array}\right) P_N = \left(\begin{array}{c:c} P_n & O \\ \hdashline O & P_m \end{array}\right)$

証明. 自然数 k に対して $\mathbf{1}_k = (\underbrace{1, 1, \ldots, 1}_{k\text{ 個}})^\top$ とおくと

$$P_N = \left(\begin{array}{c:c} E_n - \frac{n}{N} Q_n & -\frac{1}{N} \mathbf{1}_n \otimes \mathbf{1}_m \\ \hdashline -\frac{1}{N} \mathbf{1}_m \otimes \mathbf{1}_n & E_m - \frac{m}{N} Q_m \end{array}\right)$$

と表せる。これと $P_kQ_k=O$，$P_k\mathbf{1}_k=\mathbf{0}$ が成り立つことを用いると，次の式変形が確かめられる。

$$\left(\begin{array}{c:c} P_n & O \\ \hdashline O & P_m \end{array}\right)P_N=\left(\begin{array}{c:c} P_n & O \\ \hdashline O & P_m \end{array}\right)\left(\begin{array}{c:c} E_n-\frac{n}{N}Q_n & -\frac{1}{N}\mathbf{1}_n\otimes\mathbf{1}_m \\ \hdashline -\frac{1}{N}\mathbf{1}_m\otimes\mathbf{1}_n & E_m-\frac{m}{N}Q_m \end{array}\right)$$
$$=\left(\begin{array}{c:c} P_n-\frac{n}{N}P_nQ_n & -\frac{1}{N}(P_n\mathbf{1}_n)\otimes\mathbf{1}_m \\ \hdashline -\frac{1}{N}(P_m\mathbf{1}_m)\otimes\mathbf{1}_n & P_m-\frac{m}{N}P_mQ_m \end{array}\right)=\left(\begin{array}{c:c} P_n & O \\ \hdashline O & P_m \end{array}\right)$$

□

命題 4.38–(1) の証明. (1) $S_{\textsf{Total}}=\frac{1}{N}X^\top P_NX=\frac{1}{N}(P_NX)^\top(P_NX)$ であるから，まず P_NX を計算する。補題 5 (➲ p. 11) を用いて

$$P_NX=\left\{\left(\begin{array}{c:c} P_n & O \\ \hdashline O & P_m \end{array}\right)+\left(\begin{array}{c:c} Q_n & O \\ \hdashline O & Q_m \end{array}\right)-Q_N\right\}\left(\begin{array}{c} X_{\mathrm{A}} \\ \hdashline X_{\mathrm{B}} \end{array}\right)$$
$$=\left(\begin{array}{c:c} P_n & O \\ \hdashline O & P_m \end{array}\right)X+\Bigg\{\underbrace{\left(\begin{array}{c} Q_nX_{\mathrm{A}} \\ \hdashline Q_mX_{\mathrm{B}} \end{array}\right)}_{=Y}-\underbrace{Q_NX}_{\overset{\|}{Q_NY}}\Bigg\}=\left(\begin{array}{c:c} P_n & O \\ \hdashline O & P_m \end{array}\right)X+P_NY$$

であり，ゆえに次が成り立つ。

$$S_{\textsf{Total}}=\frac{1}{N}(P_NX)^\top P_NX$$
$$=\frac{1}{N}\left\{X^\top\left(\begin{array}{c:c} P_n & O \\ \hdashline O & P_m \end{array}\right)+Y^\top P_N\right\}\left\{\left(\begin{array}{c:c} P_n & O \\ \hdashline O & P_m \end{array}\right)X+P_NY\right\}$$
$$=\frac{1}{N}X^\top\left(\begin{array}{c:c} P_n & O \\ \hdashline O & P_m \end{array}\right)X+\frac{1}{N}X^\top\left(\begin{array}{c:c} P_n & O \\ \hdashline O & P_m \end{array}\right)P_NY$$
$$+\frac{1}{N}Y^\top P_N\left(\begin{array}{c:c} P_n & O \\ \hdashline O & P_m \end{array}\right)X+\frac{1}{N}Y^\top P_NY$$

この右辺の各項を計算する。

$$(\text{第 1 項})=\frac{1}{N}\left(\begin{array}{c:c} X_{\mathrm{A}}^\top & X_{\mathrm{B}}^\top \end{array}\right)\left(\begin{array}{c:c} P_n & O \\ \hdashline O & P_m \end{array}\right)\left(\begin{array}{c} X_{\mathrm{A}} \\ \hdashline X_{\mathrm{B}} \end{array}\right)$$
$$=\frac{1}{N}\left(X_{\mathrm{A}}^\top P_nX_{\mathrm{A}}+X_{\mathrm{B}}^\top P_nX_{\mathrm{B}}\right)=\frac{1}{N}\left(nS_{\mathrm{A}}+mS_{\mathrm{B}}\right)=S_{\textsf{Wthn}}$$

また，$Y=\left(\begin{array}{c:c} Q_n & O \\ \hdashline O & Q_m \end{array}\right)Y$ が成り立つことに注意して，さらに $\{P_n,Q_n\}$ と $\{P_m,Q_m\}$ が 1 の直交分割 (定義 4.2 ➲ p. 108) であることを用いると，次が得られる。

$$(\text{第 2 項})=\frac{1}{N}X^\top\left(\begin{array}{c:c} P_n & O \\ \hdashline O & P_m \end{array}\right)P_NY$$
$$\overset{\substack{\text{補題 6}\\(➲\,\text{p. 11})}}{=}\frac{1}{N}X^\top\left(\begin{array}{c:c} P_n & O \\ \hdashline O & P_m \end{array}\right)\left(\begin{array}{c:c} Q_n & O \\ \hdashline O & Q_m \end{array}\right)Y=\frac{1}{N}X^\top\left(\begin{array}{c:c} P_nQ_n & O \\ \hdashline O & P_mQ_m \end{array}\right)Y=O$$

また，(第 3 項) = (第 2 項)$^\top=O$ であり，最後に (第 4 項) $=\frac{1}{N}Y^\top P_NY=S_{Y,Y}=S_{\textsf{Btw}}$ となる。 □

命題 4.38–(2) の証明.

$$Y^\top = \left(\begin{array}{cccc:cccc} \overline{a}_1 & \overline{a}_1 & \cdots & \overline{a}_1 & \overline{b}_1 & \overline{b}_1 & \cdots & \overline{b}_1 \\ \hdashline \overline{a}_2 & \overline{a}_2 & \cdots & \overline{a}_2 & \overline{b}_2 & \overline{b}_2 & \cdots & \overline{b}_2 \end{array}\right) = \left(\begin{array}{c} x_1 \\ \hdashline x_2 \end{array}\right)$$

と表しておく。この x_1 は x_2 はそれぞれ，多くとも 2 値からなる 1 次元データとなっている。このとき，

$$\begin{aligned} \mathrm{tr}(S_{\textsf{Btw}}) &= \mathrm{tr}(S_{Y,Y}) \\ &= (1\text{ 次元データ } x_1 \text{ の分散}) + (1\text{ 次元データ } x_2 \text{ の分散}) \end{aligned}$$

であるから $\mathrm{tr}(S_{\textsf{Btw}}) = 0 \Leftrightarrow \overline{a}_1 = \overline{b}_1$ かつ $\overline{a}_2 = \overline{b}_2$ であることは明らかであろう。 □

命題 4.38–(3),(4) の証明. (3) と (4) は，定義から $S_{\textsf{Btw}} = S_{Y,Y}$，$S_{\textsf{Wthn}} = \frac{n}{N}S_{X_{\mathrm{A}},X_{\mathrm{A}}} + \frac{m}{N}S_{X_{\mathrm{B}},X_{\mathrm{B}}}$ であることと，命題 4.13–(3) (➔p. 113) より従う。 □

定理 4.39(➔p. 140) の証明

補題 7 $S_{\textsf{Btw}} = \dfrac{nm}{N^2}\begin{pmatrix} \overline{a}_1 - \overline{b}_1 \\ \overline{a}_2 - \overline{b}_2 \end{pmatrix} \otimes \begin{pmatrix} \overline{a}_1 - \overline{b}_1 \\ \overline{a}_2 - \overline{b}_2 \end{pmatrix}$

証明. $S_{\textsf{Btw}} = \frac{1}{N}Y^\top P_N Y = \frac{1}{N}(P_N Y)^\top (P_N Y)$ であるから，まず $(P_N Y)$ を計算してみよう。そこで $c_1 = \frac{1}{N}(\sum_{i=1}^n \overline{a}_1 + \sum_{j=1}^m \overline{b}_1)$，$c_2 = \frac{1}{N}(\sum_{i=1}^n \overline{a}_2 + \sum_{j=1}^m \overline{b}_2)$ とおき，自然数 k に対して $\mathbf{1}_k = (\underbrace{1, 1, \ldots, 1}_{k\text{ 個}})^\top$ とおくと

$$P_N Y = \left(\begin{array}{c:c} (\overline{a}_1 - c_1)\mathbf{1}_n & (\overline{a}_2 - c_2)\mathbf{1}_n \\ \hdashline (\overline{b}_1 - c_1)\mathbf{1}_m & (\overline{b}_2 - c_2)\mathbf{1}_m \end{array}\right)$$

と表せるが，$\overline{a}_1 - c_1 = \overline{a}_1 - (\frac{n}{N}\overline{a}_1 + \frac{m}{N}\overline{b}_1) = \frac{m}{N}(\overline{a}_1 - \overline{b}_1)$ となり，同様にして $\overline{b}_1 - c_1 = \frac{n}{N}(\overline{b}_1 - \overline{a}_1)$，$\overline{a}_2 - c_2 = \frac{m}{N}(\overline{a}_2 - \overline{b}_2)$，$\overline{b}_2 - c_2 = \frac{n}{N}(\overline{b}_2 - \overline{a}_2)$ が成り立つから

$$P_N Y = \left(\begin{array}{c:c} \frac{m}{N}(\overline{a}_1 - \overline{b}_1)\mathbf{1}_n & \frac{m}{N}(\overline{a}_2 - \overline{b}_2)\mathbf{1}_n \\ \hdashline \frac{n}{N}(\overline{b}_1 - \overline{a}_1)\mathbf{1}_m & \frac{n}{N}(\overline{b}_2 - \overline{a}_2)\mathbf{1}_m \end{array}\right) = \frac{nm}{N}\left(\begin{array}{c:c} \frac{1}{n}(\overline{a}_1 - \overline{b}_1)\mathbf{1}_n & \frac{1}{n}(\overline{a}_2 - \overline{b}_2)\mathbf{1}_n \\ \hdashline \frac{1}{m}(\overline{b}_1 - \overline{a}_1)\mathbf{1}_m & \frac{1}{m}(\overline{b}_2 - \overline{a}_2)\mathbf{1}_m \end{array}\right)$$

である。ゆえに，

$$\begin{aligned} S_{\textsf{Btw}} &= \frac{1}{N}(P_N Y)^\top (P_N Y) \\ &= \frac{(nm)^2}{N^3}\left(\begin{array}{c:c} \frac{1}{n}(\overline{a}_1 - \overline{b}_1)\mathbf{1}_n^\top & \frac{1}{m}(\overline{b}_1 - \overline{a}_1)\mathbf{1}_m^\top \\ \hdashline \frac{1}{n}(\overline{a}_2 - \overline{b}_2)\mathbf{1}_n^\top & \frac{1}{m}(\overline{b}_2 - \overline{a}_2)\mathbf{1}_m^\top \end{array}\right)\left(\begin{array}{c:c} \frac{1}{n}(\overline{a}_1 - \overline{b}_1)\mathbf{1}_n & \frac{1}{n}(\overline{a}_2 - \overline{b}_2)\mathbf{1}_n \\ \hdashline \frac{1}{m}(\overline{b}_1 - \overline{a}_1)\mathbf{1}_m & \frac{1}{m}(\overline{b}_2 - \overline{a}_2)\mathbf{1}_m \end{array}\right) \\ &= \frac{(nm)^2}{N^3}\left(\begin{array}{c:c} (\overline{a}_1 - \overline{b}_1)^2\{\frac{\langle \mathbf{1}_n, \mathbf{1}_n\rangle}{n^2} + \frac{\langle \mathbf{1}_m, \mathbf{1}_m\rangle}{m^2}\} & (\overline{a}_1 - \overline{b}_1)(\overline{a}_2 - \overline{b}_2)\{\frac{\langle \mathbf{1}_n, \mathbf{1}_n\rangle}{n^2} + \frac{\langle \mathbf{1}_m, \mathbf{1}_m\rangle}{m^2}\} \\ \hdashline (\overline{a}_2 - \overline{b}_2)(\overline{a}_1 - \overline{b}_1)\{\frac{\langle \mathbf{1}_n, \mathbf{1}_n\rangle}{n^2} + \frac{\langle \mathbf{1}_m, \mathbf{1}_m\rangle}{m^2}\} & (\overline{a}_2 - \overline{b}_2)^2\{\frac{\langle \mathbf{1}_n, \mathbf{1}_n\rangle}{n^2} + \frac{\langle \mathbf{1}_m, \mathbf{1}_m\rangle}{m^2}\} \end{array}\right) \end{aligned}$$

となる。ここで $\langle \mathbf{1}_k, \mathbf{1}_k \rangle = k$ であることを用いると，次が得られる。

$$\begin{aligned} S_{\mathsf{Btw}} &= \frac{(nm)^2}{N^3}\left(\frac{1}{n}+\frac{1}{m}\right)\left(\begin{array}{c:c} (\overline{a}_1-\overline{b}_1)^2 & (\overline{a}_1-\overline{b}_1)(\overline{a}_2-\overline{b}_2) \\ \hdashline (\overline{a}_2-\overline{b}_2)(\overline{a}_1-\overline{b}_1) & (\overline{a}_2-\overline{b}_2)^2 \end{array}\right) \\ &= \frac{nm}{N^2}\begin{pmatrix} \overline{a}_1-\overline{b}_1 \\ \overline{a}_2-\overline{b}_2 \end{pmatrix} \otimes \begin{pmatrix} \overline{a}_1-\overline{b}_1 \\ \overline{a}_2-\overline{b}_2 \end{pmatrix} \end{aligned}$$

□

関数 $\varphi(\mathbf{v}) = \dfrac{\langle \mathbf{v}, S_{\mathsf{Btw}}\mathbf{v}\rangle}{\langle \mathbf{v}, S_{\mathsf{Wthn}}\mathbf{v}\rangle}$， $\mathbf{v} \in \mathbb{R}^2 \setminus \{\mathbf{0}\}$ は，任意の 0 でない実数 c に対して

$$\varphi(c\mathbf{v}) = \frac{\langle c\mathbf{v}, S_{\mathsf{Btw}}(c\mathbf{v})\rangle}{\langle c\mathbf{v}, S_{\mathsf{Wthn}}(c\mathbf{v})\rangle} = \frac{c^2}{c^2}\frac{\langle \mathbf{v}, S_{\mathsf{Btw}}\mathbf{v}\rangle}{\langle \mathbf{v}, S_{\mathsf{Wthn}}\mathbf{v}\rangle} = \frac{\langle \mathbf{v}, S_{\mathsf{Btw}}\mathbf{v}\rangle}{\langle \mathbf{v}, S_{\mathsf{Wthn}}\mathbf{v}\rangle} = \varphi(\mathbf{v})$$

をみたすことに注意すると，関数 $\varphi(\mathbf{v})$ の最大値は領域 $C = \{\mathbf{v} \in \mathbb{R}^2 : \langle \mathbf{v}, S_{\mathsf{Wthn}}\mathbf{v}\rangle = 1\}$ における $\varphi(\mathbf{v})$ の最大値と同じになることがわかる。そこで，この領域 C における $\varphi(\mathbf{v})$ の振る舞いに注目してみよう。領域 C 内の $\mathbf{v}$ に対しては，$\varphi(\mathbf{v}) = \langle \mathbf{v}, S_{\mathsf{Btw}}\mathbf{v}\rangle$ の形をしている。そこで，関数 $f(\mathbf{v})$ を $\mathbf{v} = \begin{pmatrix} v_1 \\ v_2 \end{pmatrix} \in \mathbb{R}^2$ に対して

$$\varphi(\mathbf{v}) = \langle \mathbf{v}, S_{\mathsf{Btw}}\mathbf{v}\rangle = (S_{\mathsf{Btw}})_{11}(v_1)^2 + 2\underbrace{(S_{\mathsf{Btw}})_{12}}_{=(S_{\mathsf{Btw}})_{21}} v_1v_2 + (S_{\mathsf{Btw}})_{22}(v_2)^2$$

により定めておけば，領域 C において，$\varphi(\mathbf{v})$ の最大値と $f(\mathbf{v})$ の最大値は等しい。
領域 C 内での関数 $f(\mathbf{v})$ の最大点の候補を探すために Lagrange の未定乗数法 (**定理 3.10 ➲ p. 100**) を用いる。このために関数 g を

$$\begin{aligned} g(\mathbf{v}) &= \langle \mathbf{v}, S_{\mathsf{Wthn}}\mathbf{v}\rangle - 1 \\ &= (S_{\mathsf{Wthn}})_{11}(v_1)^2 + 2\underbrace{(S_{\mathsf{Wthn}})_{12}}_{=(S_{\mathsf{Wthn}})_{21}} v_1v_2 + (S_{\mathsf{Wthn}})_{22}(v_2)^2 - 1 \end{aligned}$$

とおけば，考えている領域は $C = \{\mathbf{v} \in \mathbb{R}^2 : g(\mathbf{v}) = 0\}$ と表せる。この領域内での $f(\mathbf{v})$ の最大点を $\mathbf{v} = \mathbf{w} = \begin{pmatrix} w_1 \\ w_2 \end{pmatrix}$ とおけば，Lagrange の未定乗数法 (**定理 3.10 ➲ p. 100**) により，適当な定数 $\lambda \in \mathbb{R}$ と共に，次の連立方程式を $(\mathbf{w}, \lambda)$ に関して解けばよい。

$$\begin{cases} \dfrac{\partial f}{\partial v_1}(\mathbf{w}) - \lambda\dfrac{\partial g}{\partial v_1}(\mathbf{w}) = 0, \\ \dfrac{\partial f}{\partial v_2}(\mathbf{w}) - \lambda\dfrac{\partial g}{\partial v_2}(\mathbf{w}) = 0, \\ g(\mathbf{w}) = 0, \end{cases}$$

このために偏微分を計算しておくと，次のとおりである。

$$\begin{aligned} \frac{\partial f}{\partial v_1}(\mathbf{v}) &= 2(S_{\mathsf{Btw}}\mathbf{v})_1, \quad \frac{\partial f}{\partial v_2}(\mathbf{v}) = 2(S_{\mathsf{Btw}}\mathbf{v})_2, \\ \frac{\partial g}{\partial v_1}(\mathbf{v}) &= 2(S_{\mathsf{Wthn}}\mathbf{v})_1, \quad \frac{\partial g}{\partial v_2}(\mathbf{v}) = 2(S_{\mathsf{Wthn}}\mathbf{v})_2 \end{aligned}$$

これを用いれば，上の連立方程式のうち上二つは $S_{\mathsf{Btw}}\mathbf{w} = \lambda S_{\mathsf{Wthn}}\mathbf{w}$，全体として

$$\lambda\mathbf{w} = (S_{\mathsf{Wthn}})^{-1}S_{\mathsf{Btw}}\mathbf{w}, \quad \langle \mathbf{w}, S_{\mathsf{Wthn}}\mathbf{w}\rangle = 1$$

と書き直せる。特に第 2 式から $\mathbf{w} \neq \mathbf{0}$ であり，ゆえに第 1 式から $\lambda \neq 0$ となることがわかる。

定理 4.39 の証明. (1)⇒(2): 関数 $\varphi(\mathbf{v})$ が $\mathbf{v} = \mathbf{w}$ ($\neq \mathbf{0}$) において最大値を取ったとすると，上の議論から，0 でないある数 λ を用いて $\lambda\mathbf{w} = (S_{\mathsf{Wthn}})^{-1}S_{\mathsf{Btw}}\mathbf{w}$ をみたさなければならない。ここで補題 7 (➲ **p. 13**) を用いると，次式を得る。

$$\begin{aligned}\lambda\mathbf{w} &= \frac{nm}{N^2}(S_{\mathsf{Wthn}})^{-1}\begin{pmatrix}\bar{a}_1-\bar{b}_1\\ \bar{a}_2-\bar{b}_2\end{pmatrix}\otimes\begin{pmatrix}\bar{a}_1-\bar{b}_1\\ \bar{a}_2-\bar{b}_2\end{pmatrix}\mathbf{w}\\ &= \frac{nm}{N^2}\left\langle\begin{pmatrix}\bar{a}_1-\bar{b}_1\\ \bar{a}_2-\bar{b}_2\end{pmatrix},\mathbf{w}\right\rangle(S_{\mathsf{Wthn}})^{-1}\begin{pmatrix}\bar{a}_1-\bar{b}_1\\ \bar{a}_2-\bar{b}_2\end{pmatrix}\end{aligned}$$

ゆえに二つのベクトル $\mathbf{w}$ と $(S_{\mathsf{Wthn}})^{-1}\begin{pmatrix}\bar{a}_1-\bar{b}_1\\ \bar{a}_2-\bar{b}_2\end{pmatrix}$ は平行である。

(2)⇒(1): 二つのベクトル $\mathbf{w}$ と $(S_{\mathsf{Wthn}})^{-1}\begin{pmatrix}\bar{a}_1-\bar{b}_1\\ \bar{a}_2-\bar{b}_2\end{pmatrix}$ が平行であると仮定すると，0 でない適当な実数 c を用いて $\mathbf{w} = c(S_{\mathsf{Wthn}})^{-1}\begin{pmatrix}\bar{a}_1-\bar{b}_1\\ \bar{a}_2-\bar{b}_2\end{pmatrix}$ と表せる。

一方で関数 $\varphi(\mathbf{v})$ は最大値をもつので，その最大点 (の一つ) を $\mathbf{w}^*$ とおくと，すでに示した (1)⇒(2) より，$\mathbf{w}^*$ と $(S_{\mathsf{Wthn}})^{-1}\begin{pmatrix}\bar{a}_1-\bar{b}_1\\ \bar{a}_2-\bar{b}_2\end{pmatrix}$ が平行でなければならない。ゆえに，0 でない適当な実数 c' を用いて $\mathbf{w}^* = c'(S_{\mathsf{Wthn}})^{-1}\begin{pmatrix}\bar{a}_1-\bar{b}_1\\ \bar{a}_2-\bar{b}_2\end{pmatrix}$ と表せる。いま，c と c' が 0 でないから $\mathbf{w} = (c/c')\mathbf{w}^*$ となり，ゆえに $\varphi(\mathbf{w}) = \varphi(\frac{c}{c'}\mathbf{w}^*) = \varphi(\mathbf{w}^*)$ となるので，関数 $\varphi(\mathbf{v})$ は $\mathbf{v} = \mathbf{w}$ において最大値をとる。 □

定理 4.40(➲ p. 141) の証明

定理 4.40(p. 141) の証明. まず文脈上，この定理 4.40 においては定理 4.39 (➲ **p. 140**) の仮定が有効であったことを注意しておく。

式変形 $n(\overline{f_\mathrm{A}}-\overline{f}) = n(\overline{f_\mathrm{A}}-(\frac{n}{N}\overline{f_\mathrm{A}}+\frac{m}{N}\overline{f_\mathrm{B}})) = \frac{nm}{N}(\overline{f_\mathrm{A}}-\overline{f_\mathrm{B}}) = -\frac{nm}{N}(\overline{f_\mathrm{B}}-\overline{f_\mathrm{A}}) = -m(\overline{f_\mathrm{B}}-\overline{f})$ であることを用いると，

$$(\overline{f_\mathrm{A}}-\overline{f})(\overline{f_\mathrm{B}}-\overline{f}) = \frac{\{n(\overline{f_\mathrm{A}}-\overline{f})\}\{m(\overline{f_\mathrm{B}}-\overline{f})\}}{nm} = -\frac{nm}{N^2}(\overline{f_\mathrm{A}}-\overline{f_\mathrm{B}})^2 \leqq 0$$

となる。ゆえに，あとはこの右辺が 0 にならないことを示せば十分。この $\overline{f_\mathrm{A}}-\overline{f_\mathrm{B}}$ という項は

$$\overline{f_\mathrm{A}}-\overline{f_\mathrm{B}} = w_1(\bar{a}_1-\bar{b}_1)+w_2(\bar{a}_2-\bar{b}_2) = \left\langle\begin{pmatrix}w_1\\ w_2\end{pmatrix},\begin{pmatrix}\bar{a}_1-\bar{b}_1\\ \bar{a}_2-\bar{b}_2\end{pmatrix}\right\rangle$$

と表せる。この $\mathbf{w} = \begin{pmatrix}w_1\\ w_2\end{pmatrix}$ については，定理 4.39–(2) (➲ **p. 140**) より零でない適当な実数 c を用いて $c\mathbf{w} = (S_{\mathsf{Wthn}})^{-1}\begin{pmatrix}\bar{a}_1-\bar{b}_1\\ \bar{a}_2-\bar{b}_2\end{pmatrix}$ と表せ，ゆえに $\begin{pmatrix}\bar{a}_1-\bar{b}_1\\ \bar{a}_2-\bar{b}_2\end{pmatrix} = c(S_{\mathsf{Wthn}})\mathbf{w}$ が成り立つ。最後に，実対称行列 S_{Wthn} が正則 (**ゆえに正定値行列**) であるという仮定と $\mathbf{w} \neq 0$ であることに注意すると

$$\overline{f_\mathrm{A}}-\overline{f_\mathrm{B}} = \left\langle\mathbf{w},\begin{pmatrix}\bar{a}_1-\bar{b}_1\\ \bar{a}_2-\bar{b}_2\end{pmatrix}\right\rangle = c\langle\mathbf{w},(S_{\mathsf{Wthn}})\mathbf{w}\rangle \neq 0$$

を得る。したがって，$(\overline{f_\mathrm{A}}-\overline{f_\mathrm{B}})^2 \neq 0$ である。 □

命題 4.42(➲ p. 143) の証明

(1)⇒(2): **Point** (➲ p. 7) および要点 B.18 (➲ p. 6) の対偶から，$S_{\mathsf{Wthn}}\mathbf{w} = \mathbf{0}$ となる $\mathbf{w} \in \mathbb{R}^2 \setminus \{\mathbf{0}\}$ が存在する。このとき

$$0 = \langle \mathbf{w}, S_{\mathsf{Wthn}}\mathbf{w} \rangle = \langle \mathbf{w}, (\tfrac{n}{N}S_{\mathrm{A}} + \tfrac{m}{N}S_{\mathrm{B}})\mathbf{w} \rangle = \frac{n}{N}\langle \mathbf{w}, S_{\mathrm{A}}\mathbf{w} \rangle + \frac{m}{N}\langle \mathbf{w}, S_{\mathrm{B}}\mathbf{w} \rangle$$

である。これを整理して，共分散行列 $S_{\mathrm{A}} = S_{X_{\mathrm{A}},X_{\mathrm{A}}}$，$S_{\mathrm{B}} = S_{X_{\mathrm{B}},X_{\mathrm{B}}}$ が非負定値であることを用いると，

$$0 \leqq \frac{n}{N}\langle \mathbf{w}, S_{\mathrm{A}}\mathbf{w} \rangle = -\frac{m}{N}\langle \mathbf{w}, S_{\mathrm{B}}\mathbf{w} \rangle \leqq 0$$

となるから $\langle \mathbf{w}, S_{\mathrm{A}}\mathbf{w} \rangle = \langle \mathbf{w}, S_{\mathrm{B}}\mathbf{w} \rangle = 0$ とならなければならない。これを分散の言葉で言い換える (**命題 4.13**➲ p. 113) と，$s(X_{\mathrm{A}}\mathbf{w}, X_{\mathrm{A}}\mathbf{w}) = s(X_{\mathrm{B}}\mathbf{w}, X_{\mathrm{B}}\mathbf{w}) = 0$ となる。ゆえに縦ベクトル

$$X_{\mathrm{A}}\mathbf{w} = \begin{pmatrix} \boldsymbol{a}_*^1 \\ \boldsymbol{a}_*^2 \\ \vdots \\ \boldsymbol{a}_*^n \end{pmatrix}\mathbf{w} = \begin{pmatrix} \boldsymbol{a}_*^1\mathbf{w} \\ \boldsymbol{a}_*^2\mathbf{w} \\ \vdots \\ \boldsymbol{a}_*^n\mathbf{w} \end{pmatrix}, \quad X_{\mathrm{B}}\mathbf{w} = \begin{pmatrix} \boldsymbol{b}_*^1\mathbf{w} \\ \boldsymbol{b}_*^2\mathbf{w} \\ \vdots \\ \boldsymbol{b}_*^m\mathbf{w} \end{pmatrix}$$

の各成分を横に並べた 1 次元データの平均をそれぞれ c_1，c_2 とすれば $\boldsymbol{a}_*^1\mathbf{w} = \boldsymbol{a}_*^2\mathbf{w} = \cdots = \boldsymbol{a}_*^n\mathbf{w} = c_1$ と $\boldsymbol{b}_*^1\mathbf{w} = \boldsymbol{b}_*^2\mathbf{w} = \cdots = \boldsymbol{b}_*^m\mathbf{w} = c_2$ が成り立つ。これは $\mathbf{w} = \begin{pmatrix} w_1 \\ w_2 \end{pmatrix}$ と書いておけば，直線 k_1: $w_1x + w_2y = c_1$ が点 $\boldsymbol{a}_*^1, \boldsymbol{a}_*^2, \ldots, \boldsymbol{a}_*^n$ をすべて通ることを意味している。同様に，直線 k_2: $w_1x + w_2y = c_2$ は点 $\boldsymbol{b}_*^1, \boldsymbol{b}_*^2, \ldots, \boldsymbol{b}_*^m$ をすべて通ることを意味している。この二つの直線 k_1 と k_2 は平行である。

(2)⇒(1): 直線 k_1 を $w_1x + w_2y = c_1$，直線 k_2 を $w_1x + w_2y = c_2$ と表しておけば，定理 4.15 (➲ p. 114) における (2)⇒(1) の証明のようにして $\mathbf{w} = \begin{pmatrix} w_1 \\ w_2 \end{pmatrix}$ が $S_{\mathrm{A}}\mathbf{w} = S_{X_{\mathrm{A}},X_{\mathrm{A}}}\mathbf{w} = \mathbf{0}$ と $S_{\mathrm{B}}\mathbf{w} = \mathbf{0}$ をみたすことがわかる。ゆえに $S_{\mathsf{Wthn}}\mathbf{w} = \frac{n}{N}S_{\mathrm{A}}\mathbf{w} + \frac{m}{N}S_{\mathrm{B}}\mathbf{w} = \mathbf{0}$ となる。いま，**Point** (➲ p. 7) および要点 B.18 (➲ p. 6) の対偶から，S_{Wthn} は正則でない。

命題 4.44(➲ p. 146) の証明

クラスタリング $G = (G_1, G_2, \ldots, G_k)$ が L の最小点であったとし，例えばクラスター G_1 に属する標本点 ω_a が最も「近い」クラスターが G_1 ではなく G_2 であったとしよう (**他の場合も以下の議論は同様である**):

$$\|\boldsymbol{x}_*^a - \overline{G}_1\| > \|\boldsymbol{x}_*^a - \overline{G}_2\| \tag{A.7}$$

このとき，$G_1' = G_1 \setminus \{\omega_a\}$，$G_2' = G_2 \cup \{\omega_a\}$，$G_3' = G_3, \ldots, G_k' = G_k$ によって新しいクラスタリング $G' = (G_1', G_2', \ldots, G_k')$ を考えてみる。(**このとき $G_1 \setminus \{\omega_a\} = \varnothing$ となることはない。実際，さもなければ $G_1 = \{\omega_a\}$，したがって $\overline{G}_1 = x_*^a$ となり，不等式 (A.7) が成り立つことはありえないからである。**) によりこしらえると

$$\begin{aligned} L(G) &= \sum_{j=1}^{k} \sum_{i:\ \omega_i \in G_j} \|\boldsymbol{x}_*^i - \overline{G}_j\|^2 \\ &= \sum_{i:\ \omega_i \in G_1} \|\boldsymbol{x}_*^i - \overline{G}_1\|^2 + \sum_{i:\ \omega_i \in G_2} \|\boldsymbol{x}_*^i - \overline{G}_2\|^2 + \sum_{j=3}^{k} \sum_{i:\ \omega_i \in G_j} \|\boldsymbol{x}_*^i - \overline{G}_j\|^2 \end{aligned} \tag{A.8}$$

この第 1 項目は

$$\sum_{i:\ \omega_i \in G_1} \|\boldsymbol{x}_*^i - \overline{G}_1\|^2 = \sum_{i:\ \omega_i \in G_1 \setminus \{\omega_a\}} \|\boldsymbol{x}_*^i - \overline{G}_1\|^2 + \|\boldsymbol{x}_*^a - \overline{G}_1\|^2$$
$$\overset{\text{(A.7)}}{>} \sum_{i:\ \omega_i \in G_1 \setminus \{\omega_a\}} \|\boldsymbol{x}_*^i - \overline{G}_1\|^2 + \underbrace{\|\boldsymbol{x}_*^a - \overline{G}_2\|^2}_{\substack{\text{これを (A.8) 最右辺} \\ \text{第 2 項にまとめる。}}}$$

であるから

$$L(G) > \sum_{\substack{i: \\ \omega_i \in G_1 \setminus \{\omega_a\}}} \|\boldsymbol{x}_*^i - \overline{G}_1\|^2 + \sum_{\substack{i: \\ \omega_i \in G_2 \cup \{\omega_a\}}} \|\boldsymbol{x}_*^i - \overline{G}_2\|^2 + \sum_{j=3}^{k} \sum_{\substack{i: \\ \omega_i \in G_j}} \|\boldsymbol{x}_*^i - \overline{G}_j\|^2$$
$$\overset{G'_j\text{ の定義}}{=} \sum_{\substack{i: \\ \omega_i \in G'_1}} \|\boldsymbol{x}_*^i - \overline{G}_1\|^2 + \sum_{\substack{i: \\ \omega_i \in G'_2}} \|\boldsymbol{x}_*^i - \overline{G}_2\|^2 + \sum_{j=3}^{k} \sum_{\substack{i: \\ \omega_i \in G'_j}} \|\boldsymbol{x}_*^i - \overline{G'_j}\|^2$$

ここで定理 2.5–(2) (➔p. 12) のことを思い出すと，この右辺の第 1 項と第 2 項について

$$\sum_{i:\ \omega_i \in G'_1} \|\boldsymbol{x}_*^i - \overline{G}_1\|^2 \geqq \sum_{i:\ \omega_i \in G'_1} \|\boldsymbol{x}_*^i - \overline{G'_1}\|^2, \quad \sum_{i:\ \omega_i \in G'_2} \|\boldsymbol{x}_*^i - \overline{G}_2\|^2 \geqq \sum_{i:\ \omega_i \in G'_2} \|\boldsymbol{x}_*^i - \overline{G'_2}\|^2$$

が成り立つから

$$L(G) > \sum_{\substack{i: \\ \omega_i \in G'_1}} \|\boldsymbol{x}_*^i - \overline{G'_1}\|^2 + \sum_{\substack{i: \\ \omega_i \in G'_2}} \|\boldsymbol{x}_*^i - \overline{G'_2}\|^2 + \sum_{j=3}^{k} \sum_{\substack{i: \\ \omega_i \in G'_j}} \|\boldsymbol{x}_*^i - \overline{G'_j}\|^2 = L(G'),$$

つまり $L(G) > L(G')$ となり，これは G が L の最小点であることに反する。

定理 4.45(➔p. 147) の証明

クラスタリング $G = (G_1, G_2, \ldots, G_k)$ の下で，標本点 ω_i が最も近いクラスター G_j のクラスター番号 (の一つ) を $j = s_i(G)$ と表すことにする。つまり，各標本点 ω_i に対して $\|\boldsymbol{x}_*^i - \overline{G}_j\| \geqq \|\boldsymbol{x}_*^i - \overline{G}_{s_i(G)}\|$ が成り立ち，さらに新しいクラスタリング $G' = (G'_1, G'_2, \ldots, G'_k)$ は $G'_j = \{\omega_i : s_i(G) = j\}$ によって与えられる。

このとき

$$L(G) = \sum_{j=1}^{k} \sum_{i:\ \omega_i \in G_j} \underbrace{\|\boldsymbol{x}_*^i - \overline{G}_j\|^2}_{\substack{\text{II}\!\vee \\ \|\boldsymbol{x}_*^i - \overline{G}_{s_i(G)}\|^2}} \geqq \sum_{j=1}^{k} \sum_{i:\ \omega_i \in G_j} \underbrace{\|\boldsymbol{x}_*^i - \overline{G}_{s_i(G)}\|^2}_{\text{これは } j \text{ によらない!}}$$
$$= \sum_{i=1}^{N} \|\boldsymbol{x}_*^i - \overline{G}_{s_i(G)}\|^2 = \sum_{j=1}^{k} \underbrace{\sum_{i:\ \omega_i \in G'_j} \|\boldsymbol{x}_*^i - \overline{G}_{s_i(G)}\|^2}_{\text{次はここに注目}}$$

ここで，定理 2.5–(2) (➔p. 12) のことを思い出すと，各クラスター番号 $j = 1, 2, \ldots, k$ に対して

$\sum_{i:\ \omega_i \in G'_j} \|\boldsymbol{x}_*^i - \overline{G}_{s_i(G)}\|^2 \geqq \sum_{i:\ \omega_i \in G'_j} \|\boldsymbol{x}_*^i - \overline{G'_j}\|^2$ が成り立つから，次の不等式が得られる。

$$L(G) \geqq \sum_{j=1}^{k} \sum_{i:\ \omega_i \in G'_j} \|\boldsymbol{x}_*^i - \overline{G'_j}\|^2 = L(G')$$

定理 7.14(➔p. 195) の証明

(2)⇒(1): $f(y \mid \theta) = \exp(a(\theta)t(y) - \psi(\theta) + b(y))$ であったとする。このとき $\mathbf{y} = (y_1, y_2, \ldots, y_n)$ に対して次が成り立つ。

$$\begin{aligned} f_n(\mathbf{y} \mid \theta) &= \prod_{i=1}^{n} f(y_i \mid \theta) = \prod_{i=1}^{n} \exp(a(\theta)t(y_i) - \psi(\theta) + b(y_i)) \\ &= \exp\left(a(\theta)\sum_{i=1}^{n} t(y_i) - n\psi(\theta)\right)\exp\left(\sum_{i=1}^{n} b(y_i)\right) \end{aligned}$$

そこで $\tau(\mathbf{y}) = \displaystyle\sum_{i=1}^{n} t(y_i)$ とおけば，無作為標本を $\mathbf{Y} = (Y_1, Y_2, \ldots, Y_n)$ と表すとき，$\tau(\mathbf{Y})$ が θ に対する 1 次元の十分統計量である。

(1)⇒(2): 標本の大きさについて $n = 2$ のときに示す。[*1] $\mathbf{Y} = (Y_1, Y_2)$ を無作為標本とし，$\tau(\mathbf{Y})$ が θ に対する十分統計量であるとし，パラメータ φ_0 を任意に固定する。すべての θ について $f_{|\theta} = f_{|\varphi_0}$ ならば, $a(\theta) = 1$, $t(y) = \log f(y \mid \varphi_0)$, $\psi(\theta) = b(y) = 0$ とおくと $f(y \mid \theta) = \exp(a(\theta) {\cdot} t(y) - \psi(\theta) + b(y))$, つまり $\{f(y \mid \theta)\}_\theta$ は指数型分布族となる。そこで以下では $f_{|\theta_0} \neq f_{|\varphi_0}$ なるパラメータ θ_0 が存在する場合を考える。

$\tau(\mathbf{Y})$ は十分統計量であるから，適当な関数 $g(\theta; t)$ を選ぶと θ の関数として $f(y \mid \theta) \propto g(\theta; \tau(\mathbf{y}))$ が成り立つから，$\sigma_\theta(y) = \log \dfrac{f(y \mid \theta)}{f(y \mid \varphi_0)}$ とおくと次が成り立つ。

$$\sigma_\theta(y_1) + \sigma_\theta(y_2) = \log \frac{f_n(\mathbf{y} \mid \theta)}{f_n(\mathbf{y} \mid \varphi_0)} = \log \frac{g(\theta; \tau(\mathbf{y}))}{g(\varphi_0; \tau(\mathbf{y}))} \ (= \Psi_\theta(\tau(\mathbf{y}))\ \text{とおく。})$$

そこでこの σ_θ や Ψ_θ に限らず，より一般にすべての x_1, x_2 について

$$\sigma(x_1) + \sigma(x_2) = \Psi(\tau(x_1, x_2))$$

をみたす連続な関数 σ と Ψ のもつ性質について調べることから始めよう。

Step 1. 点 x_0 を任意に固定する。$\sigma(y_1) = \sigma(y_2)$ かつ $\tau(x_0, y_1) < \tau(x_0, y_2)$ ならば，y が x_0 の近くを動くとき $\sigma(y)$ は定数である。

このことを示そう。条件より $y_1 \neq y_2$ でなければならないが，以下 $y_1 < y_2$ であるとして議論を進める。条件より $\tau(x_0, y_1) < \tau(x_0, y_2)$ であるが，必要なら区間 (y_1, y_2) をさらに狭く取り直すことで，条件 $\sigma(y_1) = \sigma(y_2)$ を崩さずに次が成り立つようにできる。

$$\tau(x_0, y_1) < \tau(x_0, y) < \tau(x_0, y_2), \quad y_1 < y < y_2 \tag{A.9}$$

関数 σ の連続性と，条件 $\sigma(y_1) = \sigma(y_2)$ より，開区間 (y_1, y_2) 内の点 y_0 において，$\sigma(y_0)$ は閉区間 $[y_1, y_2]$ 上での σ の最大値もしくは最小値をとる。

いま $\sigma(y_0)$ はその最小値であったとしよう (**最大値であった場合も同様の議論ができる**)。式

[*1] この証明は O. Barndorff–Nielsen and K. Pedersen (1968) 『Sufficient Data Reduction and Exponential Families』Mathematica Scandinavica **22**, pp. 197–202 に基づく。

(A.9) より次が成り立つ。

$$\underbrace{\tau(x_0, y_1)}_{\substack{=t_1\\ \text{とおく。}}} < \underbrace{\tau(x_0, y_0)}_{\substack{=t_0\\ \text{とおく。}}} < \underbrace{\tau(x_0, y_2)}_{\substack{=t_2\\ \text{とおく。}}}$$

ここで x の関数 $\tau(x, y_1)$，$\tau(x, y_0)$，$\tau(x, y_2)$ の連続性を用いる。上式において「外側二つの項に現れた点 $x = x_0$ を一斉に点 x_0 の十分近くで動かす」また「真ん中の項に現れた点 $x = x_0$ を点 x_0 の十分近くで動かす」(これらで動かす x の範囲を区間 I とする) と次が成り立つことがわかる。

$$\tau(x, y_1) < \underbrace{\tau(x_0, y_0)}_{=t_0} < \tau(x, y_2), \quad x \in I \tag{A.10}$$

$$\underbrace{\tau(x_0, y_1)}_{=t_1} < \tau(x, y_0) < \underbrace{\tau(x_0, y_2)}_{=t_2}, \quad x \in I \tag{A.11}$$

不等式 (A.10) より特に，すべての $x \in I$ に対して $t_0 = \tau(x, \alpha(x))$ となるような $\alpha(x) \in (y_1, y_2)$ が取れる (**中間値の定理**)。これより次が従い，この (b) によって **Step 1** が完了する。

(a) $\Psi(t_0) = \min_{t_1 \leqq t \leqq t_2} \Psi(t)$。

🡸 実際，

$$\begin{aligned}\Psi(t_0) = \Psi(\tau(x_0, y_0)) = \sigma(x_0) + \sigma(y_0) &= \sigma(x_0) + \min_{y_1 \leqq y \leqq y_2} \sigma(y)\\ &= \min_{y_1 \leqq y \leqq y_2} (\sigma(x_0) + \sigma(y))\\ &= \min_{y_1 \leqq y \leqq y_2} \Psi(\tau(x_0, y)) \overset{\text{(A.11)}}{=} \min_{t_1 \leqq t \leqq t_2} \Psi(t)。\end{aligned}$$

(b) すべての $x \in I$ について $\sigma(x) = \sigma(x_0)$。

🡸 実際，$x \in I$ を任意とすると次が成り立つ。

$$\begin{aligned}\sigma(x_0) + \sigma(y_0) = \Psi(\tau(x_0, y_0)) = \Psi(t_0) = \Psi(\tau(x, \alpha(x))) &= \sigma(x) + \sigma(\alpha(x))\\ &\overset{\substack{y_1 < \alpha(x) < y_2 \text{ であり，}\\ y_0 \text{ は区間 } [y_1, y_2] \text{ 上での}\\ \sigma \text{ の最小点であることより}}}{\geqq} \sigma(x) + \sigma(y_0)\\ &= \Psi(\tau(x, y_0)) \overset{\text{(a)}}{\geqq} \Psi(t_0)\end{aligned}$$

であり，以上の評価において $\Psi(t_0)$ が 2 回現れていることを考えると，これらはすべて等しくならなければならない。ゆえに $\sigma(x_0) + \sigma(y_0) = \sigma(x) + \sigma(y_0)$ であり，よって $\sigma(x_0) = \sigma(x)$ が成り立つ。

Step 2. **θ に依存する定数 $a(\theta)$，$b(\theta)$ とある θ_0 について $\sigma_\theta(y) = a(\theta)\sigma_{\theta_0}(y) + b(\theta)$ が成り立つ。**

$\sigma_{\theta_0}(y)$ が定数関数でないような θ_0 を選び固定する。するとある y_1 と y_2 について $\sigma_{\theta_0}(y_1) \neq \sigma_{\theta_0}(y_2)$ が成り立たなければならない。このとき，すべての x_0 に対して $\Psi_{\theta_0}(\tau(x_0, y_1)) = \sigma_{\theta_0}(x_0) + \sigma_{\theta_0}(y_1) \neq \sigma_{\theta_0}(x_0) + \sigma_{\theta_0}(y_2) = \Psi_{\theta_0}(\tau(x_0, y_2))$ であるから，**すべての x_0 に対して $\tau(x_0, y_1) \neq \tau(x_0, y_2)$ が成り立つ。**

また $a(\theta) = \dfrac{\sigma_\theta(y_1) - \sigma_\theta(y_2)}{\sigma_{\theta_0}(y_1) - \sigma_{\theta_0}(y_2)}$ とおくと $\sigma_\theta(y_1) - \sigma_\theta(y_2) = a(\theta) \cdot (\sigma_{\theta_0}(y_1) - \sigma_{\theta_0}(y_2))$ であり，これを整理して次を得る。

$$\sigma_\theta(y_1) - a(\theta) \cdot \sigma_{\theta_0}(y_1) = \sigma_\theta(y_2) - a(\theta) \cdot \sigma_{\theta_0}(y_2)$$

そこで新たな関数 $\sigma(y)$ を

$$\sigma(y) = \sigma_\theta(y) - a(\theta) \cdot \sigma_{\theta_0}(y)$$

により定める（これは θ にも依存する）と $\sigma(y_1) = \sigma(y_2)$ であり，任意の x_1，x_2 について次が成り立つ。

$$\begin{aligned}\sigma(x_1) + \sigma(x_2) &= \Big(\sigma_\theta(x_1) - a(\theta) \cdot \sigma_{\theta_0}(x_1)\Big) + \Big(\sigma_\theta(x_2) - a(\theta) \cdot \sigma_{\theta_0}(x_2)\Big) \\ &= \Psi_\theta(\tau(x_1, x_2)) - a(\theta) \cdot \Psi_{\theta_0}(\tau(x_1, x_2)) \\ &(= \Psi(\tau(x_1, x_2)) \text{ とおく。})\end{aligned}$$

ゆえに **Step 1** より，$\sigma(y)$ は $y = x_0$ の近くで定数関数となるが，x_0 は任意であること，連続関数 σ は区間上で定義されていることより，$\sigma(y)$ は定数関数でなければならない。ゆえに，ある定数 $b(\theta)$ について $\sigma(y) \equiv b(\theta)$ である。以上から $\sigma_\theta(y) - a(\theta) \cdot \sigma_{\theta_0}(y) \equiv b(\theta)$ となる。

Step 3. いま，$\sigma_\theta(y) - a(\theta) \cdot \sigma_{\theta_0}(y) \equiv b(\theta)$ を σ_θ の定義に従って書き直すと

$$\log \frac{f(y \mid \theta)}{f(y \mid \varphi_0)} = a(\theta) \cdot \log \frac{f(y \mid \theta_0)}{f(y \mid \varphi_0)} + b(\theta)$$

であり，ゆえに $f(y \mid \theta) = \exp\left(a(\theta) \cdot \log \frac{f(y \mid \theta_0)}{f(y \mid \varphi_0)} + b(\theta)\right) f(y \mid \varphi_0)$ となる。ここで，φ_0 と θ_0 は固定した定数であることに注意すると，これは指数型分布族の形をしていることがわかる。

定理 7.19(➲ p. 200) の証明

不偏推定量 $T = \tau(\mathbf{Y})$ を任意とする。このとき $\theta = \mathbf{E}[T \mid \theta] = \int \tau(\mathbf{y}) f_n(\mathbf{y} \mid \theta) \mathrm{d}\mathbf{y}$ が成立するので，これを θ で微分し，微分と積分を交換すると，次式が得られる。

$$1 = \int \tau(\mathbf{y}) \frac{\mathrm{d} f_n(\mathbf{y} \mid \theta)}{\mathrm{d}\theta} \mathrm{d}\mathbf{y} = \int \tau(\mathbf{y}) \left(\frac{\mathrm{d}}{\mathrm{d}\theta} \ell_n(\theta; \mathbf{y})\right) f_n(\mathbf{y} \mid \theta) \mathrm{d}\mathbf{y} = \mathbf{E}[T \cdot \left(\frac{\mathrm{d}}{\mathrm{d}\theta} \ell_n(\theta; \mathbf{Y})\right) \mid \theta]$$

一方で $s_n(\theta; \mathbf{Y}) = \sum_{i=1}^n s(\theta; Y_i)$ の期待値は 0 であるから，

$$0 = \theta \cdot \mathbf{E}[s_n(\theta; \mathbf{Y}) \mid \theta] = \mathbf{E}[\theta \cdot \left(\frac{\mathrm{d}}{\mathrm{d}\theta} \ell_n(\theta; \mathbf{Y})\right) \mid \theta]$$

となる。上の二式を辺々引くと次を得る。

$$1 = \mathbf{E}[(T - \theta) \cdot \left(\frac{\mathrm{d}}{\mathrm{d}\theta} \ell_n(\theta; \mathbf{Y})\right) \mid \theta] \overset{\substack{\text{Cauchy-Schwarz の} \\ \text{の不等式 (➲ p. 105)}}}{\leqq} \mathbf{E}[(T - \theta)^2 \mid \theta]^{\frac{1}{2}} \cdot \mathbf{E}[\left(\frac{\mathrm{d}}{\mathrm{d}\theta} \ell_n(\theta; \mathbf{Y})\right)^2 \mid \theta]^{\frac{1}{2}}$$

が成立する。両辺を二乗して整理すると，次が得られる。

$$\mathrm{Var}(T \mid \theta) \overset{\substack{\mathbf{E}[T \mid \theta] = \theta \\ \text{より}}}{=} \mathbf{E}[(T - \theta)^2 \mid \theta] \leqq \mathbf{E}[\left(\frac{\mathrm{d}}{\mathrm{d}\theta} \ell_n(\theta; \mathbf{Y})\right)^2 \mid \theta]^{-1} = \frac{1}{I_n(\theta)}$$

これにより Cramér-Rao の不等式 (7.2.1)（➲ p. 200）が示された。この等号成立条件は，上の評価に現れた Cauchy-Schwarz の不等式における等号成立条件を調べればよく，これはすべての θ に対して $T - \theta$ と $\frac{\mathrm{d}}{\mathrm{d}\theta} \ell_n(\theta; \mathbf{Y})$ が比例関係にあるとき，つまり，ある θ だけの関数 $g(\theta)$ を用いて $T - \theta = \left(\frac{\mathrm{d}}{\mathrm{d}\theta} \ell_n(\theta; \mathbf{Y})\right) \cdot g(\theta)$ と書けるときであり，かつそのときに限る。

定理 9.3(➔ p. 270) の証明

概略のみを示す。ここでは，確率変数列 $\{X_n\}_{n=1}^{\infty}$ について，「すべての正数 ε に対して $\lim_{n\to\infty} \mathbf{P}(|X_n| \geqq \varepsilon \mid \theta_0) = 0$ が成り立つ」ことを「$(X_n \mid \theta_0) = o_p(1)$」あるいは単に「$X_n = o_p(1)$」と表す。このことを指して「($\mathbf{P}(\bullet \mid \theta_0)$ の下で) X_n は 0 に確率収束する」という。一方で「すべての正数 ε に対して $\lim_{K\to\infty} \sup_n \mathbf{P}(|X_n| \geqq K \mid \theta_0) = 0$ が成り立つ」ことを「$(X_n \mid \theta_0) = O_p(1)$」あるいは単に「$X_n = O_p(1)$」と表す。このことを「($\mathbf{P}(\bullet \mid \theta_0)$ の下で) X_n は確率有界である」という。

また $X_n = o_p(1)$ かつ，ある数 $\alpha > 1$ を用いて $\sup_n \mathbf{E}[|X_n|^{\alpha} \mid \theta_0] < \infty$ が成り立つとき，$\mathbf{E}[|X_n| \mid \theta_0] = o(1)$ が成り立つことが知られている。

ここでは，次の類のやや複雑な条件がみたされていることを想定する。

- 各 $\varepsilon > 0$ に対して $\displaystyle\int f(y \mid \theta_0) \max_{\substack{\theta: \\ \|\theta-\theta_0\| \leqq \varepsilon}} \left| \frac{\partial^3 \ell(\theta; y)}{\partial\theta_i \partial\theta_j \partial\theta_k} \right|^2 \mathrm{d}y < \infty$,

最尤推定量 $\widehat{\Theta}_n$ の漸近正規性より，$\sqrt{n}(\widehat{\Theta}_n - \theta_0)$ の分布が多変量正規分布 $\mathrm{N}(\mathbf{0}, I(\theta_0)^{-1})$ に近づいていくが (必要なら分布を保ったまま $\widehat{\Theta}_n$ を取り替えることにより),

$$\sqrt{n}(\widehat{\Theta}_n - \theta_0) = (\mathrm{N}(\mathbf{0}, I(\theta_0)^{-1}) \text{ に従う確率変数}) + o_p(1) \tag{A.12}$$

とできることが知られている。また次の条件がみたされていることを想定する。

- $\sup_n \mathbf{E}\big[(\sqrt{n}\,\|\widehat{\Theta}_n - \theta_0\|)^5 \mid \theta_0 \big] < \infty$ かつ $\sup_n \mathbf{E}\big[|G(q\|f_{|\widehat{\Theta}_n})|^5 \mid \theta_0 \big] < \infty$

この条件より，

$$\mathbf{P}(\|\widehat{\Theta}_n - \theta_0\| \geqq \varepsilon \mid \theta_0) = o\left(\frac{1}{n}\right) \tag{A.13}$$

であることがわかる。

(1) まず $\mathbf{E}[G(q\|f_{|\widehat{\Theta}_n})] - \dfrac{1}{n}\mathbf{E}\big[-\log f_n(\mathbf{Y} \mid \widehat{\Theta}_n) \mid \theta_0\big] = \mathbf{E}[G(q\|f_{|\widehat{\Theta}_n})] - \mathbf{E}\big[-\log f(Y \mid \widehat{\Theta}_n) \mid \theta_0\big]$ を次のように分解する。

$$\begin{aligned}
&\mathbf{E}[G(q\|f_{|\widehat{\Theta}_n})] - \mathbf{E}\big[-\log f(Y \mid \widehat{\Theta}_n) \mid \theta_0\big] \\
&= \Big\{ \underbrace{\mathbf{E}[G(q\|f_{|\widehat{\Theta}_n})] - \mathrm{Ent}(q)}_{= \mathbf{E}[D_{\mathrm{KL}}(q\|f_{|\widehat{\Theta}_n}) \mid \theta_0]} \Big\} + \Big\{ \mathrm{Ent}(q) - \underbrace{\mathbf{E}\Big[\Big(-\log \underbrace{f(Y \mid \theta_0)}_{= q(Y) \text{ であるから…}}\Big) \mid \theta_0\Big]}_{\text{これは左の } \mathrm{Ent}(q) \text{ に等しい。}} \Big\} \\
&\qquad + \Big\{ \mathbf{E}\Big[\Big(-\log f(Y \mid \theta_0)\Big) \mid \theta_0\Big] - \mathbf{E}\Big[\Big(-\log f(Y \mid \widehat{\Theta}_n)\Big) \mid \theta_0\Big] \Big\}
\end{aligned}$$

これをまとめて，次式を得る。

$$\begin{aligned}
&\mathbf{E}\big[G(q\|f_{|\widehat{\Theta}_n}) \mid \theta_0\big] - \mathbf{E}\big[-\log f(Y \mid \widehat{\Theta}_n) \mid \theta_0\big] \\
&= \mathbf{E}\big[D_{\mathrm{KL}}(q\|f_{|\widehat{\Theta}_n}) \mid \theta_0\big] + \mathbf{E}\big[\ell(\widehat{\Theta}_n; Y) - \ell(\theta_0; Y) \mid \theta_0\big]
\end{aligned}$$

θ_0 に十分近い θ' をとり，$\varepsilon = \|\theta' - \theta_0\|$ とおくとき (θ' は，$\ell_n(\theta; \mathbf{y})$ に対する以下の Taylor 展開 (A.15) が $\|\theta - \theta_0\| < \varepsilon$ なる θ について成り立つくらいに十分 θ_0 に近くとる)，右辺に現れる二つの期待値を，

次のように $\|\widehat{\Theta}_n - \theta_0\| \geqq \varepsilon$ が成り立つ部分と，$\|\widehat{\Theta}_n - \theta_0\| < \varepsilon$ が成り立つ部分に分ける。

$$
\begin{aligned}
&\mathbf{E}\big[D_{\mathrm{KL}}(q\|f_{|\widehat{\Theta}_n}) \mid \theta_0\big] \\
&= \underbrace{\mathbf{E}\big[D_{\mathrm{KL}}(q\|f_{|\widehat{\Theta}_n})\,;\,\|\widehat{\Theta}_n - \theta_0\| \geqq \varepsilon \mid \theta_0\big]}_{=A_{\geqq\varepsilon}\ \text{とおく。}} + \underbrace{\mathbf{E}\big[D_{\mathrm{KL}}(q\|f_{|\widehat{\Theta}_n})\,;\,\|\widehat{\Theta}_n - \theta_0\| < \varepsilon \mid \theta_0\big]}_{=A_{<\varepsilon}\ \text{とおく。}},
\end{aligned}
$$

$$
\begin{aligned}
&\mathbf{E}\big[\ell(\widehat{\Theta}_n; Y) - \ell(\theta_0; Y) \mid \theta_0\big] \\
&= \underbrace{\mathbf{E}\big[\ell(\widehat{\Theta}_n; Y) - \ell(\theta_0; Y)\,;\,\|\widehat{\Theta}_n - \theta_0\| \geqq \varepsilon \mid \theta_0\big]}_{=B_{\geqq\varepsilon}\ \text{とおく。}} + \underbrace{\mathbf{E}\big[\ell(\widehat{\Theta}_n; Y) - \ell(\theta_0; Y)\,;\,\|\widehat{\Theta}_n - \theta_0\| < \varepsilon \mid \theta_0\big]}_{=B_{<\varepsilon}\ \text{とおく。}}
\end{aligned}
$$

式 (A.13) と仮定した条件より $A_{\geqq\varepsilon} + B_{\geqq\varepsilon} = o\left(\dfrac{1}{n}\right)$ であることがわかる。ゆえに，

$$
\mathbf{E}[G(q\|f_{|\widehat{\Theta}_n}) \mid \theta_0] - \mathbf{E}\big[-\log f(Y \mid \widehat{\Theta}_n) \mid \theta_0\big] = A_{<\varepsilon} + B_{<\varepsilon} + o\left(\frac{1}{n}\right) \tag{A.14}
$$

である。以降は $A_{<\varepsilon}$ と $B_{<\varepsilon}$ に注目する。

まず $A_{<\varepsilon}$ に現れる KL divergence $D_{\mathrm{KL}}(q\|f_{|\widehat{\Theta}_n}) = G(q\|f_{|\widehat{\Theta}_n}) - \mathrm{Ent}(q)$ を計算する。このうち，汎化損失は次のように表される。

$$
G(q\|f_{|\widehat{\Theta}_n}) = -\int f(y \mid \theta_0)\underbrace{\log f(y \mid \widehat{\Theta}_n)}_{=\ell(\widehat{\Theta}_n; y)}\mathrm{d}y
$$

ここに現れる対数尤度関数 $\ell(\theta; y) = \log f(y \mid \theta)$ を $\theta = \theta_0$ のまわりで 2 次まで Taylor 展開すると，次が得られる。

$$
\begin{aligned}
\ell(\theta; y) &= \ell(\theta_0; y) + \big\langle s(\theta_0; y), \theta - \theta_0\big\rangle \\
&\quad + \frac{1}{2}\Big\langle \theta - \theta_0, \Big(\mathrm{Hess}|_{\theta=\theta_0}\ell(\theta; y)\Big)(\theta - \theta_0)\Big\rangle + O(\|\theta - \theta_0\|^3)
\end{aligned} \tag{A.15}
$$

ここで $s(\theta_0; y) = \nabla|_{\theta=\theta_0}\ell(\theta; y)$ はスコア関数である。これにより次が成り立つ。

$$
\begin{aligned}
G(q\|f_{|\theta}) = &\underbrace{-\int f(y \mid \theta_0)\ell(\theta_0; y)\mathrm{d}y}_{=\ \mathrm{Ent}(q)} - \Big\langle \underbrace{\int f(y \mid \theta_0)s(\theta_0; y)\mathrm{d}y}_{=\ \mathbf{E}[s(\theta_0; Y) \mid \theta_0]\ =\ \mathbf{0}}, \theta - \theta_0\Big\rangle \\
&- \frac{1}{2}\Big\langle \theta - \theta_0, \Big(\underbrace{\int f(y \mid \theta_0)\,\mathrm{Hess}|_{\theta=\theta_0}\ell(\theta; y)\mathrm{d}y}_{=\ \mathbf{E}[\mathrm{Hess}|_{\theta=\theta_0}\ell(\theta; Y) \mid \theta_0]\ =\ -I(\theta_0)}\Big)(\theta - \theta_0)\Big\rangle + O(\|\theta - \theta_0\|^3)
\end{aligned}
$$

ゆえに，θ_0 の周りの θ について次が成り立つ。

$$
\begin{aligned}
D_{\mathrm{KL}}(q\|f_{|\theta}) &= G(q\|f_{|\theta}) - \mathrm{Ent}(q) \\
&= \frac{1}{2}\langle \theta - \theta_0, I(\theta_0)(\theta - \theta_0)\rangle + O(\|\theta - \theta_0\|^3) \\
&= \frac{1}{2}\|I(\theta_0)^{1/2}(\theta - \theta_0)\|^2 + O(\|\theta - \theta_0\|^3)
\end{aligned} \tag{A.16}
$$

いま上式の θ に MLE $\widehat{\Theta}_n$ を代入した $D_{\mathrm{KL}}(q\|f_{|\widehat{\Theta}_n})$ の期待値を計算する。式 (A.12) より，

$$\widehat{\Theta}_n-\theta_0=\frac{O_p(1)}{\sqrt{n}},\quad \text{特に}\quad \|\widehat{\Theta}_n-\theta_0\|^2=\frac{O_p(1)}{n}=\frac{o_p(1)}{\sqrt{n}},\quad \|\widehat{\Theta}_n-\theta_0\|^3=\frac{o_p(1)}{n}$$

であるが，より詳しくみると，

$$\begin{aligned}\widehat{\Theta}_n-\theta_0&=\frac{(\mathrm{N}(\mathbf{0},I(\theta_0)^{-1})\text{ に従う確率変数})}{\sqrt{n}}+\frac{o_p(1)}{\sqrt{n}}\\&=(\mathrm{N}(\mathbf{0},I_n(\theta_0)^{-1})\text{ に従う確率変数})+\frac{o_p(1)}{\sqrt{n}}\end{aligned}$$

であるから，次が成り立つ。

$$\begin{aligned}I_n(\theta_0)^{1/2}(\widehat{\Theta}_n-\theta_0)&=\left(\begin{array}{c}\mathrm{N}(\mathbf{0},I_n(\theta_0)^{1/2}I_n(\theta_0)^{-1}I_n(\theta_0)^{1/2})\\ \text{に従う確率変数}\end{array}\right)+o_p(1)\\&=(\mathrm{N}(\mathbf{0},E_d)\text{ に従う確率変数})+o_p(1)\end{aligned}$$

これにより次を得る。

$$\begin{aligned}\mathbf{E}[\|I(\theta_0)^{1/2}(\widehat{\Theta}_n-\theta_0)\|^2\mid\theta_0]&=\frac{1}{n}\mathbf{E}[\|I_n(\theta_0)^{1/2}(\widehat{\Theta}_n-\theta_0)\|^2\mid\theta_0]\\&=\frac{1}{n}\mathrm{tr}\Bigl(\underbrace{\mathrm{N}(\mathbf{0},E_d)\text{ の共分散行列}}_{=E_d}\Bigr)+o\left(\frac{1}{n}\right)\\&=\frac{d}{n}+o\left(\frac{1}{n}\right)\end{aligned}\tag{A.17}$$

これと式 (A.16) を見比べることで，次式が得られる。

$$\begin{aligned}A_{<\varepsilon}&=\mathbf{E}\bigl[D_{\mathrm{KL}}(q\|f_{|\widehat{\Theta}_n});\ \|\widehat{\Theta}_n-\theta_0\|<\varepsilon\mid\theta_0\bigr]\\&\overset{\substack{\text{(A.16)}\\(\text{➔ p. 22})}}{=}\ \frac{1}{2}\mathbf{E}[\|I(\theta_0)^{1/2}(\widehat{\Theta}_n-\theta_0)\|^2\mid\theta_0]+\mathbf{E}\Bigl[\frac{o_p(1)}{n}\mid\theta_0\Bigr]\\&\overset{\text{(A.17)}}{=}\ \frac{d}{2n}+o\left(\frac{1}{n}\right)\end{aligned}\tag{A.18}$$

次に $B_{<\varepsilon}$ に注目する。$\ell(\widehat{\Theta}_n;Y)-\ell(\theta_0;Y)$ を計算する。いま，$\|\widehat{\Theta}_n-\theta_0\|<\varepsilon$ の場合を考えているから，次の Taylor 展開が有効である。

$$\begin{aligned}\ell(\theta;Y)=\ell(\widehat{\Theta}_n;Y)&+\bigl\langle\underbrace{s(\widehat{\Theta}_n;Y)}_{=\mathbf{0}},\theta-\widehat{\Theta}_n\bigr\rangle\\&+\frac{1}{2}\bigl\langle\theta-\widehat{\Theta}_n,\Bigl(\underbrace{\mathrm{Hess}|_{\theta=\widehat{\Theta}_n}\ell(\theta;Y)}_{\substack{\text{次に，ここを}\\ \mathrm{Hess}|_{\theta=\theta_0}\ell(\theta;Y)\\ \text{に置き換えることを考える}}}\Bigr)(\theta-\widehat{\Theta}_n)\bigr\rangle+\frac{o_p(1)}{n}\end{aligned}\tag{A.19}$$

であり，さらに θ の関数 $\mathrm{Hess}\,\ell(\theta;Y)=\left(\dfrac{\partial^2\ell(\theta;Y)}{\partial\theta_i\partial\theta_j}\right)_{i,j=1}^{d}$ を点 $\theta=\theta_0$ のまわりで 1 次まで Taylor

展開すると，次のようになる。

$$\frac{\partial^2 \ell(\theta;Y)}{\partial\theta_i\partial\theta_j} = \left.\frac{\partial^2 \ell(\theta;Y)}{\partial\theta_i\partial\theta_j}\right|_{\theta=\theta_0} + \underbrace{\sum_{k=1}^{d} \left.\frac{\partial^3 \ell(\theta;Y)}{\partial\theta_i\partial\theta_j\partial\theta_k}\right|_{\theta=\theta_0} (\theta_k - (\theta_0)_k) + O_p(\|\theta-\theta_0\|^2)}_{\substack{\| \\ O_p(\|\theta-\theta_0\|)}}$$

これに $\theta = \widehat{\Theta}_n$ を代入すると，上式から次を得る。

$$\ell(\widehat{\Theta}_n;Y) - \ell(\theta_0;Y) = -\frac{1}{2}\Big\langle \theta_0 - \widehat{\Theta}_n, \Big(\underbrace{\mathrm{Hess}|_{\theta=\theta_0}\ell(\theta;Y)}_{\text{またしてもここに注目!}} \Big)(\theta_0 - \widehat{\Theta}_n)\Big\rangle + \frac{o_p(1)}{n}$$

ここで $\mathrm{Hess}|_{\theta=\theta_0}\ell(\theta;Y)$ の平均は

$$\mathbf{E}[\mathrm{Hess}|_{\theta=\theta_0}\ell(\theta;Y) \mid \theta_0] = \mathbf{E}[\mathrm{Hess}|_{\theta=\theta_0}\ell(\theta;Y) \mid \theta_0] = -I(\theta_0)$$

であり，中心極限定理から $\mathbf{P}(\bullet \mid \theta_0)$ のもとで

$$\begin{aligned}
&\frac{\Big(\mathrm{Hess}|_{\theta=\theta_0}\ell_n(\theta;\mathbf{Y})\Big)_{i,j} - \big(-I_n(\theta_0)\big)_{i,j}}{\sqrt{n}} \\
&= \frac{\displaystyle\sum_{k=1}^{n}\left(\left.\frac{\partial^2 \ell(\theta;Y_k)}{\partial\theta_i\partial\theta_j}\right|_{\theta=\theta_0} - \mathbf{E}\Big[\left.\frac{\partial^2 \ell(\theta;Y)}{\partial\theta_i\partial\theta_j}\right|_{\theta=\theta_0} \mid \theta_0\Big]\right)}{\sqrt{n}} \overset{n\to\infty}{\longrightarrow} \mathrm{N}\Big(0, \mathrm{Var}\Big(\left.\frac{\partial^2 \ell(\theta;Y)}{\partial\theta^i\partial\theta^j}\right|_{\theta=\theta_0} \mid \theta_0\Big)\Big)
\end{aligned}$$

となることがわかる。ゆえに (必要なら無作為標本 $\mathbf{Y} = (Y_1, Y_2, \ldots, Y_n)$ を分布を変えずに取り替えることで)，$n \to \infty$ のとき $\mathbf{P}(\bullet \mid \theta_0)$ のもとで次が成り立つ。

$$\mathrm{Hess}|_{\theta=\theta_0}\ell_n(\theta;\mathbf{Y}) = -I_n(\theta_0) + \sqrt{n}\cdot O_p(1) \tag{A.20}$$

式 (A.19) (➲ p. 23) において $\theta = \theta_0$ を代入し，上式 (A.20) を用いると，次が成り立つ。

$$\begin{aligned}
&\ell_n(\widehat{\Theta}_n;\mathbf{Y}) - \ell_n(\theta_0;\mathbf{Y}) \\
&\overset{\substack{\text{(A.19)}\\(➲\text{ p. 23})}}{=} -\frac{1}{2}\langle \theta_0 - \widehat{\Theta}_n, \Big(\mathrm{Hess}|_{\theta=\widehat{\Theta}_n}\ell_n(\theta;\mathbf{Y})\Big)(\theta_0 - \widehat{\Theta}_n)\rangle + o_p(1) \\
&\overset{\text{(A.20)}}{=} \frac{1}{2}\underbrace{\langle \theta_0 - \widehat{\Theta}_n, I_n(\theta_0)(\theta_0 - \widehat{\Theta}_n)\rangle}_{\substack{\| \\ \|I_n(\theta_0)^{1/2}(\theta_0 - \widehat{\Theta}_n)\|^2}} + \sqrt{n}\cdot O_p(1)\cdot\|\theta_0 - \widehat{\Theta}_n\|^2 + o_p(1) \\
&= \frac{1}{2}\|I_n(\theta_0)^{1/2}(\theta_0 - \widehat{\Theta}_n)\|^2 + o_p(1)
\end{aligned}$$

この期待値をとると，次式を得る。

$$\begin{aligned}
B_{<\varepsilon} &= \mathbf{E}\big[\ell(\widehat{\Theta}_n;Y) - \ell(\theta_0;Y);\ \|\widehat{\Theta}_n - \theta_0\| < \varepsilon \mid \theta_0\big] \\
&= \frac{1}{n}\mathbf{E}\big[\ell_n(\widehat{\Theta}_n;\mathbf{Y}) - \ell_n(\theta_0;\mathbf{Y});\ \|\widehat{\Theta}_n - \theta_0\| < \varepsilon \mid \theta_0\big] \\
&= \frac{1}{2n}\mathbf{E}\big[\|I_n(\theta_0)^{1/2}(\theta_0 - \widehat{\Theta}_n)\|^2 \mid \theta_0\big] + o\Big(\frac{1}{n}\Big) \overset{\substack{\text{(A.17)}\\(➲\text{ p. 23})}}{=} \frac{d}{2n} + o\Big(\frac{1}{n}\Big)
\end{aligned}$$

(ここで $(Y_i \mid \widehat{\Theta}_n)$ が i によらず同じ確率分布をもつという事実を暗に用いている) これと式 (A.18) (➲ p. 23) を式 (A.14) (➲ p. 22) に代入することによって (1) が示される。
(2) 上の (1) と同様にして，$\|\widehat{\Theta}_n - \theta_0\| < \varepsilon$ の部分のみ考えれば十分である。このとき，まず $G(q\|f_{|\widehat{\Theta}_n})$ の，その期待値からの偏差は

$$\begin{aligned}
G(q\|f_{|\widehat{\Theta}_n}) - \mathbf{E}[G(q\|f_{|\widehat{\Theta}_n}) \mid \theta_0] &= \Big\{G(q\|f_{|\widehat{\Theta}_n}) - \mathrm{Ent}(q)\Big\} - \Big\{\mathbf{E}[G(q\|f_{|\widehat{\Theta}_n}) \mid \theta_0] - \mathrm{Ent}(q)\Big\} \\
&= D_{\mathrm{KL}}(q\|f_{|\widehat{\Theta}_n}) - \mathbf{E}[D_{\mathrm{KL}}(q\|f_{|\widehat{\Theta}_n}) \mid \theta_0] \\
&= D_{\mathrm{KL}}(q\|f_{|\widehat{\Theta}_n}) - \mathbf{E}[D_{\mathrm{KL}}(q\|f_{|\widehat{\Theta}_n});\, \|\widehat{\Theta}_n - \theta_0\| < \varepsilon \mid \theta_0] + o_p\left(\frac{1}{n}\right) \\
&\overset{\substack{\text{(A.16) ➲ p. 22} \\ \text{(A.18) ➲ p. 23}}}{=} \frac{1}{2}\|I(\theta_0)^{1/2}(\widehat{\Theta}_n - \theta_0)\|^2 - \frac{d}{2n} + o_p\left(\frac{1}{n}\right)
\end{aligned}$$

であるから，

$$\begin{aligned}
&\mathrm{Var}\Big(G(q\|f_{|\widehat{\Theta}_n}) \mid \theta_0\Big) = \mathbf{E}[\Big(G(q\|f_{|\widehat{\Theta}_n}) - \mathbf{E}[G(q\|f_{|\widehat{\Theta}_n}) \mid \theta_0]\Big)^2 \mid \theta_0] \\
&= \mathbf{E}[\Big(G(q\|f_{|\widehat{\Theta}_n}) - \mathbf{E}[G(q\|f_{|\widehat{\Theta}_n}) \mid \theta_0]\Big)^2;\, \|\widehat{\Theta}_n - \theta_0\| < \varepsilon \mid \theta_0] + o\left(\frac{1}{n}\right) \\
&= \mathbf{E}[\left(\frac{1}{2}\|I(\theta_0)^{1/2}(\widehat{\Theta}_n - \theta_0)\|^2 - \frac{d}{2n} + o\left(\frac{1}{n}\right)\right)^2 \mid \theta_0] + o\left(\frac{1}{n}\right) \\
&= \frac{1}{4}\mathbf{E}[\|I(\theta_0)^{1/2}(\widehat{\Theta}_n - \theta_0)\|^4 \mid \theta_0] - \frac{d}{2n}\underbrace{\mathbf{E}[\|I(\theta_0)^{1/2}(\widehat{\Theta}_n - \theta_0)\|^2 \mid \theta_0]}_{\substack{\| \text{(A.17) ➲ p. 23} \\ \frac{d+o(1)}{n}}} + \frac{d^2}{4n} + o\left(\frac{1}{n}\right)
\end{aligned}$$

ここに現れた第 1 項について，式 (A.17) (➲ p. 23) の導出と同様にして，次を得ることができる。

$$\mathbf{E}[\|I_n(\theta_0)^{1/2}(\widehat{\Theta}_n - \theta_0)\|^4 \mid \theta_0] = \underbrace{\Big(\ \chi_d^2\ \text{分布の 2 次モーメント}\ \Big)}_{= d^2 + 2d} + o(1)$$

ここで，χ_d^2 分布の 2 次モーメントとは，χ_d^2 分布に従う確率変数 X に関する量 $\mathbf{E}[X^2] = d^2 + 2d$ のことをさす。上式より，次を得る。

$$\mathrm{Var}\Big(G(q\|f_{|\widehat{\Theta}_n}) \mid \theta_0\Big) = \frac{d^2 + 2d}{4n} - \frac{d^2}{2n} + \frac{d^2}{4n} + o\left(\frac{1}{n}\right) = \frac{d}{2n} + o\left(\frac{1}{n}\right)$$

命題 9.10(➲ p. 275) の証明

(1)

$$\frac{\mathbf{P}(\widehat{Q}^{(n)} = q)}{\mathbf{P}(\widehat{Q}^{(n)} = p)} = \frac{\sum_{x \in T_n(q)} \mathbf{P}(\hat{q}(x) = q)}{\sum_{x \in T_n(p)} \mathbf{P}(\hat{q}(x) = p)} = \frac{\#(T_n(q)) \cdot \prod_{l=1}^d (q_l)^{nq_l}}{\#(T_n(p)) \cdot \prod_{l=1}^d (q_l)^{np_l}}$$

$$= \frac{\dfrac{n!}{(nq_1)!(nq_2)!\cdots(nq_d)!}\displaystyle\prod_{l=1}^d (q_l)^{nq_l}}{\dfrac{n!}{(np_1)!(np_2)!\cdots(np_d)!}\displaystyle\prod_{l=1}^d (q_l)^{np_l}} = \prod_{l=1}^d \frac{(np_l)!}{(nq_l)!}(q_l)^{n(q_l - p_l)}$$

$\dfrac{m!}{n!} \geqq n^{m-n}$ ($m \geqq n$ と $m < n$ に場合分けすれば示すことができる) により，次の評価が得られる。

$$(\text{上式}) \geqq \prod_{l=1}^{d} (nq_l)^{(np_l)-(nq_l)} (q_l)^{n(q_l-p_l)} = \prod_{l=1}^{d} n^{(np_l)-(nq_l)} = n^{n\left(\sum_{l=1}^{d} p_l - \sum_{l=1}^{d} q_l\right)} = n^{n(1-1)} = 1$$

(2) $X_1, X_2, \ldots, X_n$ の独立性より，次の式変形が確かめられる。

$$\begin{aligned}
&\mathbf{P}(X_1 = x_1; X_2 = x_2; \cdots; X_n = x_n) \\
&= \prod_{i=1}^{n} \mathbf{P}(X_i = x_i) = \prod_{1 \leqq l \leqq d} \prod_{\substack{1 \leqq i \leqq n: \\ x_i = a_l}} \mathbf{P}(X_i = x_i) \\
&= \prod_{1 \leqq l \leqq d} \prod_{\substack{1 \leqq i \leqq n: \\ x_i = a_l}} q_l = \prod_{l=1}^{d} (q_l)^{n \cdot \hat{q}_l(x)} = 2^{-n \sum_{l=1}^{d} \hat{q}_l(x) \log_2 \frac{1}{q_l}} \\
&= 2^{-n \sum_{l=1}^{d} \left\{ \hat{q}_l(x) \log_2 \frac{1}{\hat{q}_l(x)} + \hat{q}_l(x) \log_2 \frac{\hat{q}_l(x)}{q_l} \right\}} = 2^{-n\{\mathrm{Ent}(\hat{q}(x)) + D(\hat{q}(x) \| q)\}}
\end{aligned}$$

(3) まず，上界について考える。独立同分布列 $X_1, X_2, \ldots, X_n$ のそれぞれが q に従うから，命題 9.10–(2) (➔ p. 275) より，

$$\begin{aligned}
1 &\geqq \sum_{x \in T_n(q)} \mathbf{P}(X_1 = x_1; X_2 = x_2; \cdots; X_n = x_n) \\
&\overset{\substack{\text{命題 9.10–(2)} \\ (➔\ \text{p. 275})}}{=} \sum_{x \in T_n(q)} 2^{-n\{\mathrm{Ent}(\hat{q}(x)) + D_{\mathrm{KL}}(\hat{q}(x) \| q)\}} \\
&\overset{\substack{\hat{q}(x) = q \text{ より}, \\ D_{\mathrm{KL}}(\hat{q}(x) \| q) = 0}}{=} \sum_{x \in T_n(q)} 2^{-n\mathrm{Ent}(q)} = 2^{-n\mathrm{Ent}(q)} \cdot \#(T_n(q))
\end{aligned}$$

なので，$\#(T_n(\pi)) \leqq 2^{n\mathrm{Ent}(q)}$ であることが示された。
次に，下界について考える。

$$\begin{aligned}
1 = \sum_{p \in \mathcal{P}_n} \mathbf{P}(\widehat{Q}^{(n)} = p) &\overset{\substack{\text{命題 9.10–(1)} \\ (➔\ \text{p. 275})}}{\leqq} \sum_{p \in \mathcal{P}_n} \mathbf{P}(\widehat{Q}^{(n)} = q) \\
&= \#\mathcal{P}_n \cdot \mathbf{P}(\widehat{Q}^{(n)} = q) \overset{\substack{\text{命題 9.10–(2)} \\ (➔\ \text{p. 275})}}{=} (\#\mathcal{P})(\#T_n(q)) 2^{-n\mathrm{Ent}(q)}
\end{aligned}$$

より，$\dfrac{1}{\#\mathcal{P}_n} 2^{n\mathrm{Ent}(q)} \leqq \#(T_n(q))$ であることが示された。
(4) 命題 9.10–(2) (➔ p. 275) から $\mathbf{P}(\widehat{Q}^{(n)} = p) = \#T_n(p) \cdot 2^{-n\{\mathrm{Ent}(p) + D_{\mathrm{KL}}(p \| q)\}}$ であることがわかり，これと命題 9.10–(3) (➔ p. 275) から成り立つことがわかる。

(5) 直接計算で示す。

$$
\begin{aligned}
\mathbf{P}(D_{\mathrm{KL}}(\widehat{Q}^{(n)}\|q)>\varepsilon) &= \sum_{\substack{p\in\mathcal{P}_n:\\ D_{\mathrm{KL}}(p\|q)>\varepsilon}} \mathbf{P}(\widehat{Q}^{(n)}=p) \\
&\overset{\substack{\text{命題 9.10–(4)}\\ (\text{➔ p. 275})}}{\leqq} \sum_{\substack{p\in\mathcal{P}_n:\\ D_{\mathrm{KL}}(p\|q)>\varepsilon}} 2^{-nD_{\mathrm{KL}}(p\|q)} < \sum_{p\in\mathcal{P}_n} 2^{-n\varepsilon} = (\#\mathcal{P}_n)\cdot 2^{-n\varepsilon}
\end{aligned}
$$

となり，命題 9.10–(5) (➔ **p.275**) の最初の不等式が示された。二つ目の不等式は，$\#\mathcal{P}_n \leqq (n+1)^d = 2^{d\log_2(n+1)}$ であることから，$\#\mathcal{P}_n\cdot 2^{-n\varepsilon} \leqq 2^{-n\left(\varepsilon-\frac{d}{n}\log_2(n+1)\right)}$ となり示される。

定理 9.16(➔ p. 278) の証明

まず，式 (9.5.3) (➔ **p. 278**) は，次のように示される。

$$
\begin{aligned}
\mathbf{P}(\widehat{Q}^{(n)}\in\Gamma) &= \mathbf{P}(\widehat{Q}^{(n)}\in\Gamma\cap\mathcal{P}_n) = \sum_{p\in\Gamma\cap\mathcal{P}_n} \mathbf{P}(\widehat{Q}^{(n)}=p) \\
&\overset{\substack{\text{命題 9.10–(4)}\\ (\text{➔ p. 275})}}{\leqq} \sum_{p\in\Gamma\cap\mathcal{P}_n} 2^{-nD_{\mathrm{KL}}(p\|q)} \leqq \#(\Gamma\cap\mathcal{P}_n)\cdot 2^{-n\inf\limits_{p\in\Gamma} D_{\mathrm{KL}}(p\|q)} \leqq \#\mathcal{P}_n\cdot 2^{-n\inf\limits_{p\in\Gamma} D_{\mathrm{KL}}(p\|q)}
\end{aligned}
$$

次に，式 (9.5.5) を示す。そこで，各 n に対して $D_{\mathrm{KL}}(p^{(n)}\|q) = \min\limits_{p\in\Gamma_n\cap\mathcal{P}_n} D_{\mathrm{KL}}(p\|q)$ をみたすようなタイプの列 $p^{(n)}\in\Gamma_n\cap\mathcal{P}_n$ $(n=1,2,3,\ldots)$ を一つ固定する。このとき，仮定より $\lim\limits_{n\to\infty} D_{\mathrm{KL}}(p^{(n)}\|q) = \lim\limits_{n\to\infty}\min\limits_{p\in\Gamma_n\cap\mathcal{P}_n} D_{\mathrm{KL}}(p\|q) = \inf\limits_{p\in\Gamma} D_{\mathrm{KL}}(p\|q)$ が成り立つ。

$$
\begin{aligned}
\mathbf{P}(\widehat{Q}_n\in\Gamma_n) &= \sum_{p\in\Gamma_n\cap\mathcal{P}_n} \mathbf{P}(\widehat{Q}^{(n)}=p) \geqq \mathbf{P}(\widehat{Q}^{(n)}=p^{(n)}) \\
&\overset{\substack{\text{命題 9.10–(4)}\\ (\text{➔ p. 275})}}{\geqq} \frac{1}{\#\mathcal{P}_n} 2^{-nD_{\mathrm{KL}}(p^{(n)}\|q)} \geqq (n+1)^{-d} 2^{-nD_{\mathrm{KL}}(p^{(n)}\|q)}
\end{aligned}
$$

が成り立つ。これと，式 (9.5.3) (➔ **p. 278**) における Γ として $\Gamma_n\cap\mathcal{P}_n$ を選ぶことで

$$
\begin{aligned}
(n+1)^{-d}2^{-nD_{\mathrm{KL}}(p^{(n)}\|q)} &\leqq \mathbf{P}(\widehat{Q}^{(n)}\in\Gamma_n) \\
&= \mathbf{P}(\widehat{Q}^{(n)}\in\Gamma_n\cap\mathcal{P}_n) \\
&\overset{\substack{\text{式 (9.5.3)}\\ (\text{➔ p. 278})}}{\leqq} \#\mathcal{P}_n\cdot 2^{-n\inf\limits_{p\in\Gamma_n\cap\mathcal{P}_n} D_{\mathrm{KL}}(p\|q)} \\
&= \#\mathcal{P}_n\cdot 2^{-n\min\limits_{p\in\Gamma_n\cap\mathcal{P}_n} D_{\mathrm{KL}}(p\|q)} = \#\mathcal{P}_n\cdot 2^{-nD_{\mathrm{KL}}(p^{(n)}\|q)}
\end{aligned}
$$

であり，これに $-\frac{1}{n}\log_2$ をつけて整理すると

$$
-\frac{\log_2\#\mathcal{P}_n}{n} + D_{\mathrm{KL}}(p^{(n)}\|q) \leqq -\frac{1}{n}\log_2\mathbf{P}(\widehat{Q}^{(n)}\in\Gamma_n) \leqq -\frac{\log_2(n+1)^{-d}}{n} + D_{\mathrm{KL}}(p^{(n)}\|q)
$$

が得られる。$n\to\infty$ のとき，この左辺と右辺は共に $\inf_{p\in\Gamma} D_{\mathrm{KL}}(p\|q)$ に収束するから，中辺にも極限が存在し，$\lim\limits_{n\to\infty} -\dfrac{1}{n}\log_2\mathbf{P}(\widehat{Q}^{(n)}\in\Gamma_n) = \inf\limits_{p\in\Gamma} D_{\mathrm{KL}}(p\|q)$ が成り立つ。

補題 9.23(➔ p. 284) の証明

いまの場合，$B_0 = \{\varphi_0\}$，$B_1 = \{\varphi_1\}$ であるから

$$\mathrm{lr}(\mathbf{y}) = \frac{\sup_{\theta\in\{\varphi_1\}} f_n(\mathbf{y} \mid \theta)}{\sup_{\theta\in\{\varphi_0\}} f_n(\mathbf{y} \mid \theta)} = \frac{f_n(\mathbf{y} \mid \varphi_1)}{f_n(\mathbf{y} \mid \varphi_0)}$$

である。任意の $\mathbf{y} \in \mathbb{R}^n$ に対して，$\phi_\kappa(\mathbf{y}) = \mathbf{1}_{[\kappa,\infty)}(\mathrm{lr}(\mathbf{y}))$ であるから，すべての検定関数 ϕ に対して次が成り立つ。

$$\begin{aligned} &f_n(\mathbf{y} \mid \varphi_1) - \kappa f_n(\mathbf{y} \mid \varphi_0) \geqq 0 \Longrightarrow \phi_\kappa(\mathbf{y}) = 1 \Longrightarrow (\phi_\kappa(\mathbf{y}) - \phi(\mathbf{y})) \geqq 0,\\ &f_n(\mathbf{y} \mid \varphi_1) - \kappa f_n(\mathbf{y} \mid \varphi_0) < 0 \Longrightarrow \phi_\kappa(\mathbf{y}) = 0 \Longrightarrow (\phi_\kappa(\mathbf{y}) - \phi(\mathbf{y})) \leqq 0 \end{aligned}$$

したがって

$$\begin{aligned} 0 &\leqq (\phi_\kappa(\mathbf{y}) - \phi(\mathbf{y}))\bigl(f_n(\mathbf{y} \mid \varphi_1) - \kappa f_n(\mathbf{y} \mid \varphi_0)\bigr)\\ &= \phi_\kappa(\mathbf{y}) f_n(\mathbf{y} \mid \varphi_1) - \phi(\mathbf{y}) f_n(\mathbf{y} \mid \varphi_1) - \kappa\bigl\{\phi_\kappa(\mathbf{y}) f_n(\mathbf{y} \mid \varphi_0) - \phi(\mathbf{y}) f_n(\mathbf{y} \mid \varphi_0)\bigr\} \end{aligned}$$

となる。片々 $\mathbf{y}$ に関して，$f(y \mid \theta)$ が p.d.f. なら積分，p.m.f. なら和をとると，

$$\begin{aligned} 0 &\leqq \mathbf{E}[\phi_\kappa(\mathbf{Y}) \mid \varphi_1] - \mathbf{E}[\phi(\mathbf{Y}) \mid \varphi_1] - \kappa\bigl\{\mathbf{E}[\phi_\kappa(\mathbf{Y}) \mid \varphi_0] - \mathbf{E}[\phi(\mathbf{Y}) \mid \varphi_0]\bigr\}\\ &\overset{\substack{\phi_\kappa(\mathbf{Y}) = \mathbf{1}_{\{\mathrm{LR}\geqq\kappa\}},\\ \phi(\mathbf{Y}) = \mathbf{1}_{\{\phi(\mathbf{Y})=1\}}\\ \text{より}}}{=} \underbrace{\mathbf{P}(\mathrm{LR} \geqq \kappa \mid \varphi_1)}_{\substack{\|\\ 1-\beta(\phi_\kappa)}} - \underbrace{\mathbf{P}(\phi(\mathbf{Y}) = 1 \mid \varphi_1)}_{\substack{\|\\ 1-\beta(\phi)}} - \kappa\bigl\{\underbrace{\mathbf{P}(\mathrm{LR} \geqq \kappa \mid \varphi_0)}_{\substack{\|\\ \alpha(\phi_\kappa)}} - \underbrace{\mathbf{P}(\phi(\mathbf{Y}) = 1 \mid \varphi_0)}_{\substack{\|\\ \alpha(\phi)}}\bigr\}\\ &= \bigl(\beta(\phi) - \beta(\phi_\kappa)\bigr) + \kappa\bigl(\alpha(\phi) - \alpha(\phi_\kappa)\bigr) \end{aligned}$$

となることがわかる。$\kappa > 0$ であることを考慮に入れると，$\alpha(\phi) - \alpha(\phi_\kappa)$ と $\beta(\phi) - \beta(\phi_\kappa)$ のいずれか一方が非正であれば，もう片方は非負とならなければならない。特に $\alpha(\phi) - \alpha(\phi_\kappa)$ が非正である場合を考えると $\alpha(\phi) \leqq \alpha(\phi_\kappa) \Rightarrow \beta(\phi) \geqq \beta(\phi_\kappa)$ となる。

定理 9.24(➔ p. 285) の証明

まず，α^* に対して，次のような $\tau^* > 0$ と $\gamma^* \in (0,1)$ を見つけることができる:

$$\alpha^* = \alpha(\phi_{\tau^*,\gamma^*}) = \mathbf{E}[\underbrace{\phi_{\tau^*,\gamma^*}}_{= \mathbf{1}_{\{\mathrm{LR}>\tau^*\}} + \gamma^* \cdot \mathbf{1}_{\{\mathrm{LR}=\tau^*\}}} \mid \varphi_0] = \mathbf{P}(\mathrm{LR} > \tau^* \mid \varphi_0) + \gamma^* \cdot \mathbf{P}(\mathrm{LR} = \tau^* \mid \varphi_0)$$

残りは，この尤度比検定 ϕ_{τ^*,γ^*} について $\beta(\phi_{\tau^*,\gamma^*}) = \min_{\substack{\phi:\mathbb{R}^n\to[0,1],\\ \alpha(\phi)\leqq\alpha^*}} \beta(\phi)$ が成り立つことを示せばよい。

この確率化検定関数 ϕ_{τ^*,γ^*} は $\alpha(\phi_{\tau^*,\gamma^*}) = \alpha$ $(\leqq \alpha)$ をみたすから，$\beta(\phi_{\tau^*,\gamma^*}) \geqq \min_{\substack{\phi:\mathbb{R}^n\to[0,1],\\ \alpha(\phi)\leqq\alpha^*}} \beta(\phi)$ であることは明らか。

そこで ϕ を，$\alpha(\phi) \leqq \alpha^*$ をみたす任意の確率化検定関数とする。このとき，補題 9.23 (➔ p. 284) の証明 (➔ p. 28) のときと同様に，$\bigl(\beta(\phi) - \beta(\phi_{\tau^*,\gamma^*})\bigr) + \tau^*\bigl(\alpha(\phi) - \underbrace{\alpha(\phi_{\tau^*,\gamma^*})}_{=\alpha^*}\bigr) \geqq 0$ が得られる。ところが $\alpha(\phi) - \alpha^* \leqq 0$ と $\tau^* > 0$ であることから $\beta(\phi) - \beta(\phi_{\tau^*,\gamma^*}) \geqq 0$ でなければならない。ϕ は $\alpha(\phi) \leqq \alpha^*$ をみたす確率化検定関数の中で任意であったから，$\min_{\substack{\phi:\mathbb{R}^n\to[0,1],\\ \alpha(\phi)\leqq\alpha^*}} \beta_\phi \geqq \beta_{\phi_{\tau^*,\gamma^*}}$ となる。

定理 9.26(➔ p. 287) の証明

$i, j \in \{0,1\}$ に対して $r(\varphi_i, \varphi_j) = r_{i,j}$ とおく。

$$\Phi(\phi) = \varphi_1 \Leftrightarrow \mathrm{H}_0 \text{ を棄却} \Leftrightarrow \phi(\mathbf{Y}) = 1,$$
$$\Phi(\phi) = \varphi_0 \Leftrightarrow \mathrm{H}_0 \text{ を棄却しない} \Leftrightarrow \phi(\mathbf{Y}) = 0$$

であることと，$\mathbf{1}_{\{\phi(\mathbf{Y})=1\}} = \phi(\mathbf{Y})$，$\mathbf{1}_{\{\phi(\mathbf{Y})=0\}} = 1 - \phi(\mathbf{Y})$ より，リスクは次のように変形できる。

$$\begin{aligned}
R(\phi) &= \mathbf{E}[r(\Theta, \Phi(\phi))] \\
&= \mathbf{E}[r(\varphi_0, \Phi(\phi)) \mid \Theta = \varphi_0]\pi_0 + \mathbf{E}[r(\varphi_1, \Phi(\phi)) \mid \Theta = \varphi_1]\pi_1 \\
&= \pi_0 \Big\{ r_{0,0} \underbrace{\mathbf{P}(\Phi(\phi) = \varphi_0 \mid \Theta = \varphi_0)}_{\| \atop \mathbf{E}[\underbrace{\mathbf{1}_{\{\phi(\mathbf{Y})=0\}}}_{=1-\phi(\mathbf{Y})} \mid \Theta = \varphi_0]} + r_{0,1} \underbrace{\mathbf{P}(\Phi(\phi) = \varphi_1 \mid \Theta = \varphi_0)}_{\| \atop \mathbf{E}[\underbrace{\mathbf{1}_{\{\phi(\mathbf{Y})=1\}}}_{=\phi(\mathbf{Y})} \mid \Theta = \varphi_0]} \Big\} \\
&\qquad + \pi_1 \Big\{ r_{1,0} \underbrace{\mathbf{P}(\Phi(\phi) = \varphi_0 \mid \Theta = \varphi_1)}_{\| \atop \mathbf{E}[\underbrace{\mathbf{1}_{\{\phi(\mathbf{Y})=0\}}}_{=1-\phi(\mathbf{Y})} \mid \Theta = \varphi_1]} + r_{1,1} \underbrace{\mathbf{P}(\Phi(\phi) = \varphi_1 \mid \Theta = \varphi_1)}_{\| \atop \mathbf{E}[\underbrace{\mathbf{1}_{\{\phi(\mathbf{Y})=1\}}}_{=\phi(\mathbf{Y})} \mid \Theta = \varphi_1]} \Big\} \\
&= \pi_0 \Big\{ r_{0,0}\mathbf{E}[1 - \phi(\mathbf{Y}) \mid \Theta = \varphi_0] + r_{0,1}\mathbf{E}[\phi(\mathbf{Y}) \mid \Theta = \varphi_0] \Big\} \\
&\quad + \pi_1 \Big\{ r_{1,0}\mathbf{E}[1 - \phi(\mathbf{Y}) \mid \Theta = \varphi_1] + r_{1,1}\mathbf{E}[\phi(\mathbf{Y}) \mid \Theta = \varphi_1] \Big\} \\
&= \pi_0 r_{0,0} + \pi_1 r_{1,0} + \pi_0(r_{0,1} - r_{0,0})\mathbf{E}[\phi(\mathbf{Y}) \mid \Theta = \varphi_0] - \pi_1(r_{1,0} - r_{1,1})\mathbf{E}[\phi(\mathbf{Y}) \mid \Theta = \varphi_1]
\end{aligned}$$

$\mathbf{Y}$ が $\mathbf{P}(\bullet \mid \Theta = \theta)$ の下で p.d.f. $f_n(\mathbf{y} \mid \theta)$ をもつ場合（**p.m.f. である場合も同様に計算できる**）は

$$\begin{aligned}
R(\phi) &= \pi_0 r_{0,0} + \pi_1 r_{1,0} \\
&\qquad + \pi_0(r_{0,1} - r_{0,0}) \int_{\mathbb{R}^n} \phi(\mathbf{y}) f_n(\mathbf{y} \mid \varphi_0) \mathrm{d}\mathbf{y} - \pi_1(r_{1,0} - r_{1,1}) \int_{\mathbb{R}^n} \phi(\mathbf{y}) f_n(\mathbf{y} \mid \varphi_1) \mathrm{d}\mathbf{y} \\
&= \pi_0 r_{0,0} + \pi_1 r_{1,0} \\
&\qquad + \int_{\mathbb{R}^n} \Big\{ \underbrace{\pi_0(r_{0,1} - r_{0,0}) f_n(\mathbf{y} \mid \varphi_0) - \pi_1(r_{1,0} - r_{1,1}) f_n(\mathbf{y} \mid \varphi_1)}_{R(\phi) \text{ を小さくするように } \phi(\mathbf{y}) \in \{0,1\} \text{ を選ぶには，} \atop \text{この部分が非正なら } \phi(\mathbf{y}) = 1,\ \text{正なら } \phi(\mathbf{y}) = 0 \text{ とすればよい。}} \Big\} \phi(\mathbf{y}) \mathrm{d}\mathbf{y}
\end{aligned}$$

と計算できる。ゆえに，これが最小になるような検定関数 ϕ は次で与えられることがわかる:

$$\phi(\mathbf{y}) = \begin{cases} 1 & \text{if } \underbrace{\pi_0(r_{0,1} - r_{0,0}) f_n(\mathbf{y} \mid \varphi_0) - \pi_1(r_{1,0} - r_{1,1}) f_n(\mathbf{y} \mid \varphi_1) \leqq 0}_{\Updownarrow \atop {\pi_1(r_{1,0} - r_{1,1}) f_n(\mathbf{y} \mid \varphi_1) \geqq \pi_0(r_{0,1} - r_{0,0}) f_n(\mathbf{y} \mid \varphi_0) \atop {\Updownarrow \atop (\mathrm{odds}(\mathbf{y}) =)\ \frac{f_n(\mathbf{y} \mid \varphi_1)}{f_n(\mathbf{y} \mid \varphi_0)} \frac{\pi_1}{\pi_0} \geqq \frac{r_{0,1} - r_{0,0}}{r_{1,0} - r_{1,1}}}}} \\ 0 & \text{otherwise} \end{cases}$$

これは閾値 $\kappa^* = \frac{r_{0,1} - r_{0,0}}{r_{1,0} - r_{1,1}}$ の事後オッズによる Bayes 仮説検定 ψ_{κ^*} に他ならない。

定理 9.27(➲ p. 289) の証明

この定理の証明は p. 30 で与えるが，その手順は次の通りである。

Step 1. 十分に大きなすべての n に対して，$\alpha^{(n)}(\phi_n) \leqq \varepsilon$ と，さらに次をみたす検定関数 $\phi_n : \mathcal{X}^n \to \{0,1\}$ を構成する。

$$\liminf_{n\to\infty}\Big\{-\frac{1}{n}\log_2\beta^{(n)}(\phi_n)\Big\} \geqq D_{\mathrm{KL}}(p_{|\mathrm{H}_0}\|p_{|\mathrm{H}_1}).$$

➔ このとき，次が成り立つことになる。

$$\lim_{n\to\infty}\Big\{-\frac{1}{n}\log_2\min_{\substack{\phi:\mathcal{X}^n\to\{0,1\},\\ \alpha^{(n)}(\phi)\leqq\varepsilon}}\beta^{(n)}(\phi)\Big\} \geqq \liminf_{n\to\infty}\Big\{-\frac{1}{n}\log_2\beta^{(n)}(\phi_n)\Big\} = D_{\mathrm{KL}}(p_{|\mathrm{H}_0}\|p_{|\mathrm{H}_1})$$

Step 2. $\alpha^{(n)}(\phi_n) \leqq \varepsilon$ である任意の検定 $\phi_n : \mathcal{X}^n \to \{0,1\}$ が，次をみたすことを示す。

$$D_{\mathrm{KL}}(p_{|\mathrm{H}_0}\|p_{|\mathrm{H}_1}) \geqq \limsup_{n\to\infty}\Big\{-\frac{1}{n}\log_2\beta^{(n)}(\phi_n)\Big\}$$

➔ このとき，特に $\beta^{(n)}(\phi_n) = \min_{\substack{\phi:\mathcal{X}^n\to\{0,1\},\\ \alpha^{(n)}(\phi)\leqq\varepsilon}}\beta^{(n)}(\phi)$ となるように検定関数 ϕ_n，$n = 1,2,3,\ldots$ を選んでおけば，

$$\begin{aligned}D_{\mathrm{KL}}(p_{|\mathrm{H}_0}\|p_{|\mathrm{H}_1}) &\geqq \limsup_{n\to\infty}\Big\{-\frac{1}{n}\log_2\beta^{(n)}(\phi_n)\Big\}\\ &= \limsup_{n\to\infty}\Big\{-\frac{1}{n}\log_2\min_{\substack{\phi:\mathcal{X}^n\to\{0,1\},\\ \alpha^{(n)}(\phi)\leqq\varepsilon}}\beta^{(n)}(\phi)\Big\}\end{aligned}$$

これらの結果を合わせることで Chernoff-Stein の定理が証明されたことになる。

Step 1. 検定関数 $\phi_n : \mathcal{X}^n \to \{0,1\}$ を次で定める。

$$\phi_n(\mathbf{y}) = \mathbf{1}_{[\delta_n,\infty)}(D_{\mathrm{KL}}(\hat{q}(\mathbf{y})\|p_{|\mathrm{H}_0})), \quad \mathbf{y}\in\mathcal{X}^n$$

ただし，$\delta_n = \dfrac{1}{n}\left(\log_2\dfrac{1}{\varepsilon} + d\log_2(n+1)\right)$，$d = \#\mathcal{X}$ であり，$\hat{q}(\mathbf{y})$ は 1 次元データ $\mathbf{y} = (y_1, y_2, \ldots, y_n)$ のタイプ (➲ p. 274) である。また，大きさ n の無作為標本 $\mathbf{Y} = (Y_1, Y_2, \ldots, Y_n)$ のタイプを $\widehat{Q}^{(n)} = (\widehat{Q}_1, \widehat{Q}_2, \ldots, \widehat{Q}_n)$ により表しておく。

検定 ϕ_n における第 1 種の誤り確率は $\alpha^{(n)}(\phi_n) \leqq \varepsilon$ をみたす。実際，次のように評価できる。

$$\begin{aligned}\alpha^{(n)}(\phi_n) &= \mathbf{P}(\underbrace{\text{検定 }\phi_n\text{ により }\mathrm{H}_0\text{ が棄却される}}_{\Updownarrow \atop D_{\mathrm{KL}}(\widehat{Q}^{(n)}\|p_{|\mathrm{H}_0})\geqq\delta_n} \mid \varphi_0)\\ &= \mathbf{P}\big(D_{\mathrm{KL}}(\widehat{Q}^{(n)}\|p_{|\mathrm{H}_0}) \geqq \delta_n \mid \varphi_0\big) \overset{\substack{\text{命題 9.10–(5)}\\(\text{➲ p. 275})}}{\leqq} 2^{-n\left(\delta_n - \frac{d}{n}\log_2(n+1)\right)} = \varepsilon\end{aligned}$$

第 2 種の誤り確率については，次のように評価できる。

$$\begin{aligned}\beta^{(n)}(\phi_n) &= \mathbf{P}\big(D_{\mathrm{KL}}(\widehat{Q}^{(n)}\|p_{|\mathrm{H}_0}) < \delta_n \mid \varphi_1\big)\\ &= \sum_{\substack{p\in\mathcal{P}_n:\\ D_{\mathrm{KL}}(p\|p_{|\mathrm{H}_0})<\delta_n}}\mathbf{P}(\widehat{Q}^{(n)} = p \mid \varphi_1) \overset{\substack{\text{命題 9.10–(4)}\\(\text{➲ p. 275})}}{\leqq} \sum_{\substack{p\in\mathcal{P}_n:\\ D_{\mathrm{KL}}(p\|p_{|\mathrm{H}_0})<\delta_n}} 2^{-nD_{\mathrm{KL}}(p\|p_{*|\varphi_1})} \leqq \#\mathcal{P}_n\cdot 2^{-nD_n^*}\end{aligned}$$

ただし，$D_n^* = \min_{\substack{p \in \mathcal{P}_n: \\ D_{\mathrm{KL}}(p \| p_{|\mathrm{H}_0}) < \delta_n}} D_{\mathrm{KL}}(p \| p_{|\mathrm{H}_1})$ である。ここで，$\lim_{n\to\infty} \delta_n = 0$ なので $\lim_{n\to\infty} D_n^* = D_{\mathrm{KL}}(p_{|\mathrm{H}_0} \| p_{|\mathrm{H}_1})$ となることから，次が成り立つ。

$$\begin{aligned}\liminf_{n\to\infty}\Big\{-\frac{1}{n}\log_2 \beta^{(n)}(\phi_n)\Big\} &\geqq \liminf_{n\to\infty}\Big\{-\frac{1}{n}\log_2 \#\mathcal{P}_n \cdot 2^{-nD_n^*}\Big\} \\ &= \liminf_{n\to\infty}\Big\{-\frac{1}{n}\log_2 \#\mathcal{P}_n + D_n^*\Big\} = D_{\mathrm{KL}}(p_{|\mathrm{H}_0} \| p_{|\mathrm{H}_1})\end{aligned}$$

Step 2. 各 n に対して，$\phi_n : \mathcal{X}^n \to \{0,1\}$ は $\alpha^{(n)}(\phi_n) \leqq \varepsilon$ をみたすような任意の検定関数とする。$\varepsilon_n = 2\dfrac{d}{n}\log_2(n+1)$ とおくと，

$$\begin{aligned}\mathbf{P}\big(D_{\mathrm{KL}}(\widehat{Q}^{(n)} \| p_{|\mathrm{H}_0}) < \varepsilon_n \mid \varphi_0\big) &= 1 - \mathbf{P}\big(D_{\mathrm{KL}}(\widehat{Q}^{(n)} \| p_{|\mathrm{H}_0}) \geqq \varepsilon_n \mid \varphi_0\big) \\ &\overset{\substack{\text{命題 9.10–(5)} \\ (\text{➲ p. 275})}}{\leqq} 1 - 2^{-n\left(\varepsilon_n - \frac{d}{n}\log_2(n+1)\right)} = 1 - 2^{-nd\log_2(n+1)} \xrightarrow{n\to\infty} 1\end{aligned}$$

であるから，十分大きな n に対してはいつでも $\mathbf{P}\big(D_{\mathrm{KL}}(\widehat{Q}^{(n)} \| p_{|\mathrm{H}_0}) < \varepsilon_n \mid \varphi_0\big) > 1 - \varepsilon$ が成り立つ。一方で，同時に $\mathbf{P}(\text{検定 } \phi_n \text{ により } \mathrm{H}_0 \text{ が棄却されない} \mid \varphi_0) = 1 - \alpha^{(n)}(\phi_n) > 1 - \varepsilon$ も成り立つから，

$$\begin{aligned}&\mathbf{P}\left(\begin{array}{c} D_{\mathrm{KL}}(\widehat{Q}^{(n)} \| p_{|\mathrm{H}_0}) < \varepsilon_n \text{ が成り立つ} \\ \text{かつ} \\ \text{検定 } \phi_n \text{ により } \mathrm{H}_0 \text{ が棄却されない} \end{array} \Bigg| \ \varphi_0\right) \\ &= \mathbf{P}\big(D_{\mathrm{KL}}(\widehat{Q}^{(n)} \| p_{|\mathrm{H}_0}) < \varepsilon_n \mid \varphi_0\big) + \mathbf{P}\big(\text{ 検定 } \phi_n \text{ により } \mathrm{H}_0 \text{ が棄却されない} \mid \varphi_0\big) \\ &\qquad - \mathbf{P}\left(\begin{array}{c} D_{\mathrm{KL}}(\widehat{Q}^{(n)} \| p_{|\mathrm{H}_0}) < \varepsilon_n \text{ が成り立つ} \\ \text{または} \\ \text{検定 } \phi_n \text{ により } \mathrm{H}_0 \text{ が棄却されない} \end{array} \Bigg| \ \varphi_0\right) \\ &> (1-\varepsilon) + (1-\varepsilon) - 1 = 1 - 2\varepsilon\end{aligned}$$

となる。

いま，$D_{\mathrm{KL}}(p^{(n)} \| p_{|\mathrm{H}_0}) < \varepsilon_n$ をみたすような n-タイプ $p^{(n)} \in \mathcal{P}_n$ で，さらに

$$\mathbf{P}\left(\begin{array}{c} D_{\mathrm{KL}}(\widehat{Q}^{(n)} \| p_{|\mathrm{H}_0}) < \varepsilon_n \text{ が成り立つ} \\ \text{かつ} \\ \text{検定 } \phi_n \text{ により } \mathrm{H}_0 \text{ が棄却されない} \end{array} \Bigg| \ \varphi_0\right) > (1-2\varepsilon)\mathbf{P}\big(\widehat{Q}^{(n)} = p^{(n)} \mid \varphi_0\big) \tag{A.21}$$

をみたすようなものが存在することがわかる。実際，さもなければ

$$\begin{aligned}1 - 2\varepsilon &< \mathbf{P}\big(D_{\mathrm{KL}}(\widehat{Q}^{(n)} \| p_{|\mathrm{H}_0}) < \varepsilon_n ; \text{検定 } \phi_n \text{ により } \mathrm{H}_0 \text{ が棄却されない} \mid \varphi_0\big) \\ &= \sum_{\substack{p \in \mathcal{P}_n: \\ D_{\mathrm{KL}}(p \| p_{|\mathrm{H}_0}) < \varepsilon_n}} \underbrace{\mathbf{P}\big(\widehat{Q}^{(n)} = p ; \text{検定 } \phi_n \text{ により } \mathrm{H}_0 \text{ が棄却されない} \mid \varphi_0\big)}_{\substack{\text{いまの背理法の仮定より} \\ \leqq (1-2\varepsilon)\mathbf{P}(\widehat{Q}^{(n)} = p \mid \varphi_0)}} \\ &\leqq (1-2\varepsilon) \sum_{\substack{p \in \mathcal{P}_n: \\ D_{\mathrm{KL}}(p \| p_{|\mathrm{H}_0}) < \varepsilon_n}} \mathbf{P}\big(\widehat{Q}^{(n)} = p \mid \varphi_0\big) = (1-2\varepsilon)\mathbf{P}\big(D_{\mathrm{KL}}(\widehat{Q}^{(n)} \| p_{|\mathrm{H}_0}) < \varepsilon_n \mid \varphi_0\big),\end{aligned}$$

つまり $\mathbf{P}\big(D_{\mathrm{KL}}(\widehat{Q}^{(n)} \| p_{|\mathrm{H}_0}) < \varepsilon_n \mid \varphi_0\big) > 1$ となり矛盾してしまうのである。

さて，同じタイプクラス (➲ p. 274) 内の各系列の確率はすべて等しいので，式 (A.21) より

$$\#\Big(T_n(p^{(n)}) \cap \Big\{\mathbf{y} \in \mathcal{X}^n : \phi_n(\mathbf{y}) = 0\Big\}\Big) > (1-2\varepsilon) \cdot \#T_n(p^{(n)})$$

であることが得られ，ゆえに式 (A.21) を仮説 H_1 の下で考えた式

$$\mathbf{P}\left(\begin{array}{c} D_{\mathrm{KL}}(\widehat{Q}^{(n)}\|p_{|\mathrm{H}_0}) < \varepsilon_n \text{ が成り立つ} \\ \text{かつ} \\ \text{検定 } \phi_n \text{ により } \mathrm{H}_0 \text{ が棄却されない} \end{array} \;\middle|\; \varphi_1\right) > (1-2\varepsilon)\mathbf{P}\big(\widehat{Q}^{(n)} = p^{(n)} \mid \varphi_1\big)$$

も得られる。このことから，十分大きな n に対して，

$$\begin{aligned} \beta^{(n)}(\phi_n) &= \mathbf{P}(\phi_n(\mathbf{Y}) = 0 \mid \varphi_1) \\ &\geqq \mathbf{P}\left(\begin{array}{c} D_{\mathrm{KL}}(\widehat{Q}^{(n)}\|p_{|\mathrm{H}_0}) < \varepsilon_n \text{ が成り立つ} \\ \text{かつ} \\ \text{検定 } \phi_n \text{ により } \mathrm{H}_0 \text{ が棄却されない} \end{array} \;\middle|\; \varphi_1\right) \\ &> (1-2\varepsilon)\mathbf{P}\big(\widehat{Q}^{(n)} = p^{(n)} \mid \varphi_1\big) \overset{\substack{\text{命題 9.10–(4)} \\ (\text{➔ p. 275})}}{\geqq} (1-2\varepsilon)\frac{1}{\#\mathcal{P}_n}2^{-nD_{\mathrm{KL}}(p^{(n)}\|p_{|\mathrm{H}_1})} \end{aligned}$$

が成り立つ。$p^{(n)} \in \mathcal{P}_n$ は $D_{\mathrm{KL}}(p^{(n)}\|p_{|\mathrm{H}_0}) < \varepsilon_n$ ($\overset{n\to\infty}{\longrightarrow} 0$) となるように選ばれていたから，次を得る。

$$\begin{aligned} \limsup_{n\to\infty}\left\{-\frac{1}{n}\log_2\beta^{(n)}(\phi_n)\right\} &\leqq \limsup_{n\to\infty}\left\{-\frac{1}{n}\log_2\Big(\frac{1-2\varepsilon}{\#\mathcal{P}_n}2^{-nD_{\mathrm{KL}}(p^{(n)}\|p_{|\mathrm{H}_1})}\Big)\right\} \\ &= \limsup_{n\to\infty}\left\{-\frac{1}{n}\log_2\frac{1-2\varepsilon}{\#\mathcal{P}_n} + D_{\mathrm{KL}}(p^{(n)}\|p_{|\mathrm{H}_1})\right\} = D_{\mathrm{KL}}(p_{|\mathrm{H}_0}\|p_{|\mathrm{H}_1}) \end{aligned}$$

定理 9.28(➔ p. 290) の証明

まず，

$$\begin{aligned} &\min_{\phi:\mathcal{X}^n\to\{0,1\}} \mathbf{P}(\text{検定 } \phi \text{ により選択された仮説は誤り}) \\ &\overset{\substack{\text{定理 9.26} \\ (\text{➔ p. 287})}}{=} \mathbf{P}(\text{検定 } \phi_{\mathrm{MAP},n} \text{ により選択された仮説は誤り}) \\ &= \pi_0\cdot\alpha^{(n)}(\phi_{\mathrm{MAP},n}) + \pi_1\cdot\beta^{(n)}(\phi_{\mathrm{MAP},n}) \\ &= \pi_0\mathbf{P}(\underbrace{\phi_{\mathrm{MAP},n}(\mathbf{Y}) = 1}_{\Leftrightarrow\ \mathrm{Odds}\,\geqq\,1} \mid \Theta = \varphi_0) + \pi_1\mathbf{P}(\underbrace{\phi_{\mathrm{MAP},n}(\mathbf{Y}) = 0}_{\Leftrightarrow\ \mathrm{Odds}\,<\,1} \mid \Theta = \varphi_1) \end{aligned}$$

であり，ここで

$$\log_2 \mathrm{Odds} = \log_2\left(\frac{f_n(\mathbf{Y}\mid\varphi_1)}{f_n(\mathbf{Y}\mid\varphi_0)}\cdot\frac{\pi_1}{\pi_0}\right) = \log_2 \underbrace{\prod_{i=1}^{n}\frac{f(Y_i\mid\varphi_1)}{f(Y_i\mid\varphi_0)}}_{=\prod_{l=1}^{d}\prod_{\substack{1\leqq i\leqq n:\\ Y_i=a_l}}\frac{p_{l|\mathrm{H}_1}}{p_{l|\mathrm{H}_0}}} + \log_2\frac{\pi_1}{\pi_0}$$

$$\overset{\#\{i:Y_i=a_l\}=\widehat{Q}_l\text{ より}}{=} \log_2\prod_{l=1}^{d}\left(\frac{p_{l|\mathrm{H}_1}}{p_{l|\mathrm{H}_0}}\right)^{n\widehat{Q}_l} + \log_2\frac{\pi_1}{\pi_0}$$

$$= n\sum_{l=1}^{d}\widehat{Q}_l\log_2\Bigl(\overbrace{\frac{\widehat{Q}_l}{p_{l|\mathrm{H}_0}}\cdot\frac{p_{l|\mathrm{H}_1}}{\widehat{Q}_l}}^{\frac{p_{l|\mathrm{H}_1}}{p_{l|\mathrm{H}_0}}\,=}\Bigr) + \log_2\frac{\pi_1}{\pi_0}$$

$$= n\Bigl(\underbrace{\sum_{l=1}^{d}\widehat{Q}_l\log_2\frac{\widehat{Q}_l}{p_{l|\mathrm{H}_0}}}_{=D_{\mathrm{KL}}(\widehat{Q}^{(n)}\|p_{|\mathrm{H}_0})} - \underbrace{\sum_{l=1}^{d}\widehat{Q}_l\log_2\frac{\widehat{Q}_l}{p_{l|\mathrm{H}_1}}}_{=D_{\mathrm{KL}}(\widehat{Q}^{(n)}\|p_{|\mathrm{H}_1})}\Bigr) - \log_2\frac{\pi_0}{\pi_1},$$

まとめて $\log_2 \mathrm{Odds} = n\bigl\{D_{\mathrm{KL}}(\widehat{Q}^{(n)}\|p_{|\mathrm{H}_0}) - D_{\mathrm{KL}}(\widehat{Q}^{(n)}\|p_{|\mathrm{H}_1})\bigr\} + \log_2\frac{\pi_1}{\pi_0}$ であるから，

$$\begin{aligned}
&\mathrm{Odds}\geqq 1 \Leftrightarrow \log_2\mathrm{Odds}\geqq 0\\
&\qquad\Leftrightarrow D_{\mathrm{KL}}(\widehat{Q}^{(n)}\|p_{|\mathrm{H}_0}) - D_{\mathrm{KL}}(\widehat{Q}^{(n)}\|p_{|\mathrm{H}_1}) \geqq \frac{1}{n}\log_2\frac{\pi_1}{\pi_0}\\
&\qquad\Leftrightarrow \widehat{Q}^{(n)} \in \left\{p\in\mathcal{P}(\mathcal{X}) : D_{\mathrm{KL}}(p\|p_{|\mathrm{H}_0}) - D_{\mathrm{KL}}(p\|p_{|\mathrm{H}_1}) \geqq \frac{1}{n}\log_2\frac{\pi_1}{\pi_0}\right\} (= \Gamma_n^{(\geqq)}\text{ とおく。})
\end{aligned}$$

が成り立つ。$\Gamma_n^{\geqq}$ の定義に現れる不等式を「$\geqq$」を，「$>$」「$\leqq$」「$<$」に置き換えたものをそれぞれ $\Gamma_n^{(>)}$, $\Gamma_n^{(\leqq)}$，$\Gamma_n^{(<)}$ により表す。また，「$\geqq \frac{1}{n}\log_2\frac{\pi_1}{\pi_0}$」を「$\geqq 0$」「$\leqq 0$」に置き換えたものをそれぞれ $\Gamma^{(\geqq)}$, $\Gamma^{(\leqq)}$ により表す。いま，

$$\begin{aligned}
&\min_{\phi:\mathcal{X}^n\to\{0,1\}} \mathbf{P}(\text{検定 }\phi\text{ により選択された仮説は誤り})\\
&= \pi_0\cdot\alpha^{(n)}(\phi_{\mathrm{MAP},n}) + \pi_1\cdot\beta^{(n)}(\phi_{\mathrm{MAP},n})\\
&= \pi_0\mathbf{P}(\mathrm{Odds}\geqq 1\mid\Theta=\varphi_0) + \pi_1\mathbf{P}(\mathrm{Odds}<1\mid\Theta=\varphi_1)\\
&= \pi_0\mathbf{P}(\widehat{Q}^{(n)}\in\Gamma_n^{(\geqq)}\mid\Theta=\varphi_0) + \pi_1\mathbf{P}(\widehat{Q}^{(n)}\in\Gamma_n^{(<)}\mid\Theta=\varphi_1)
\end{aligned}$$

ゆえに，

$$\alpha^{(n)}(\phi_{\mathrm{MAP},n}) = \mathbf{P}(\mathrm{Odds}\geqq 1\mid\Theta=\varphi_0) = \mathbf{P}(\widehat{Q}^{(n)}\in\Gamma_n^{(\geqq)}\mid\Theta=\varphi_0) \overset{\substack{\text{Sanov の定理}\\ (\text{➔ p. 278})}}{\approx} 2^{-n\min_{p\in\Gamma^{\geqq}} D_{\mathrm{KL}}(p\|p_{|\mathrm{H}_0})},$$

$$\begin{aligned}
\beta^{(n)}(\phi_{\mathrm{MAP},n}) &= \mathbf{P}(\mathrm{Odds}<1\mid\Theta=\varphi_0)\\
&= \mathbf{P}(\widehat{Q}^{(n)}\in\Gamma_n^{(<)}\mid\Theta=\varphi_1) \overset{\substack{\text{Sanov の定理}\\ (\text{➔ p. 278})}}{\approx} 2^{-n\inf_{p\in\Gamma^{(<)}} D_{\mathrm{KL}}(p\|p_{|\mathrm{H}_1})} = 2^{-n\min_{p\in\Gamma^{(\leqq)}} D_{\mathrm{KL}}(p\|p_{|\mathrm{H}_1})}
\end{aligned}$$

ゆえに式 (9.6.1) の左辺の極限が存在し，それは次のように計算できる。

$$
\begin{aligned}
&\lim_{n\to\infty}\left\{-\frac{1}{n}\log_2 \min_{\phi:\mathcal{X}^n\to\{0,1\}} \mathbf{P}(\text{検定 }\phi\text{ により選択された仮説は誤り})\right\} \\
&= \lim_{n\to\infty}\left\{-\frac{1}{n}\log_2\left(\pi_0\cdot 2^{-n\min_{p\in\Gamma^{(\geqq)}} D_{\mathrm{KL}}(p\|p_{|\mathrm{H}_0})} + \pi_1\cdot 2^{-n\min_{p\in\Gamma^{(\leqq)}} D_{\mathrm{KL}}(p\|p_{|\mathrm{H}_1})}\right)\right\}
\end{aligned}
\tag{A.22}
$$

そこで $\min_{p\in\Gamma^{(\geqq)}} D_{\mathrm{KL}}(p\|p_{|\mathrm{H}_0})$ と $\min_{p\in\Gamma^{(\leqq)}} D_{\mathrm{KL}}(p\|p_{|\mathrm{H}_1})$ を計算しよう。このためには，それぞれ次の最小化問題を解けばよい。

(a) $$\underset{p=(p_1,p_2,\ldots,p_d)\in\mathbb{R}^d}{\text{minimize}} \ \sum_{l=1}^{d} p_l \log_2\left|\frac{p_l}{p_{l|\mathrm{H}_0}}\right|$$

$$
\text{subject to}\ \begin{cases}
\underbrace{\sum_{l=1}^{d} p_l \log_2\left|\frac{p_l}{p_{l|\mathrm{H}_0}}\right| - \sum_{l=1}^{d} p_l \log_2\left|\frac{p_l}{p_{l|\mathrm{H}_1}}\right| \geqq 0}_{\Leftrightarrow \sum_{l=1}^{d} p_l \log_2\left|\frac{p_{l|\mathrm{H}_0}}{p_{l|\mathrm{H}_1}}\right| \leqq 0}\ , \\
-p_l \leqq 0, \quad l=1,2,\ldots,d, \\
p_1+p_2+\cdots+p_d = 1
\end{cases}
$$

(b) $$\underset{p=(p_1,p_2,\ldots,p_d)\in\mathbb{R}^d}{\text{minimize}} \ \sum_{l=1}^{d} p_l \log_2\left|\frac{p_l}{p_{l|\mathrm{H}_1}}\right|$$

$$
\text{subject to}\ \begin{cases}
\underbrace{\sum_{l=1}^{d} p_l \log_2\left|\frac{p_l}{p_{l|\mathrm{H}_0}}\right| - \sum_{l=1}^{d} p_l \log_2\left|\frac{p_l}{p_{l|\mathrm{H}_1}}\right| \leqq 0}_{\Leftrightarrow \sum_{l=1}^{d} p_l \log_2\left|\frac{p_{l|\mathrm{H}_1}}{p_{l|\mathrm{H}_0}}\right| \leqq 0}\ , \\
-p_l \leqq 0, \quad l=1,2,\ldots,d, \\
p_1+p_2+\cdots+p_d = 1
\end{cases}
$$

例えば最適化問題 (b) を解いてみよう。このために，$\lambda\in\mathbb{R}$，$\mu=(\mu_1,\mu_2,\ldots,\mu_d)\in\mathbb{R}^d$，$\nu\in\mathbb{R}$ に対して

$$
L(p,\lambda,\mu,\nu) = \sum_{l=1}^{d} p_l\log_2\left|\frac{p_l}{p_{l|\mathrm{H}_1}}\right| + \lambda\sum_{l=1}^{d} p_l\log_2\left|\frac{p_{l|\mathrm{H}_1}}{p_{l|\mathrm{H}_0}}\right| + \sum_{l=1}^{d}\mu_l(-p_l) + \nu\left(\sum_{l=1}^{d} p_l - 1\right)
$$

とおくと，最小化問題 (b) の解 $p^*=(p_1^*,p_2^*,\ldots,p_d^*)$ は，ある λ^*，$\mu^*=(\mu_1^*,\mu_2^*,\ldots,\mu_d^*)$，$\nu^*$ と共に，次の KKT 条件をみたす。

(i) $\dfrac{\partial L}{\partial p_l}(p^*,\lambda^*,\mu^*,\nu^*)=0, \quad l=1,2,\ldots,d,$

← **微分を計算すると**

$$
\frac{\partial L}{\partial p_l}(p,\lambda,\mu,\nu) = \log_2|p_l| - \log_2\left|(p_{l|\mathrm{H}_0})^{\lambda}(p_{l|\mathrm{H}_1})^{1-\lambda}\right| + 1 - \mu_l + \nu
$$

であるから，最適解は p^* **は次をみたす。**

$$
p_l^* = 2^{1-\mu_l^*+\nu^*}(p_{l|\mathrm{H}_0})^{\lambda^*}(p_{l|\mathrm{H}_1})^{1-\lambda^*}, \quad l=1,2,\ldots,d
$$

(ii) $p_1^*, p_2^*, \ldots, p_d^* \geqq 0,\ \sum_{l=1}^{d} p_l^* = 1$ (p^* が $\mathcal{X}$ 上の **p.m.f.** であるということ) かつ

$$D_{\mathrm{KL}}(p^* \| p_{|\mathrm{H}_0}) - D_{\mathrm{KL}}(p^* \| p_{|\mathrm{H}_1}) \leqq 0,$$

(iii) $L(p^*, \lambda^*, \mu^*, \nu^*)$ の式に現れる次の式はいずれも 0 である。

$$\lambda^* \underbrace{\sum_{l=1}^{d} p_l^* \log_2 \frac{p_{l|\mathrm{H}_1}}{p_{l|\mathrm{H}_0}}}_{\substack{\| \\ D_{\mathrm{KL}}(p^* \| p_{|\mathrm{H}_0}) - D_{\mathrm{KL}}(p^* \| p_{|\mathrm{H}_1})}}, \quad \mu_1^* p_1^*, \quad \mu_2^* p_2^*, \quad \ldots, \quad \mu_d^* p_d^*, \quad \nu^* \Bigl(\underbrace{\sum_{l=1}^{d} p_l^* - 1}_{\substack{\textbf{(ii)}\ \text{より} \\ \text{そもそも } 0}} \Bigr)$$

$l = 1, 2, \ldots, d$ を任意とする。もし $p_l^* > 0$ をみたすなら，(iii) より $\mu_l^* = 0$ とならなければならず，ゆえに (i) より $p_l^* = 2^{1+\nu^*}(p_{l|\mathrm{H}_0})^{\lambda^*}(p_{l|\mathrm{H}_1})^{1-\lambda^*}$ が成り立つ。また，$p_l^* = 0$ なら，(i) より $(p_{l|\mathrm{H}_0})^{\lambda^*}(p_{l|\mathrm{H}_1})^{1-\lambda^*} = 0$ が成り立たなければならない。以上から，すべての $l = 1, 2, \ldots, d$ に対して $p_l^* = 2^{1+\nu^*}(p_{l|\mathrm{H}_0})^{\lambda^*}(p_{l|\mathrm{H}_1})^{1-\lambda^*}$ が成り立つ。いま，(ii) より $1 = \sum_{k=1}^{d} p_k^* = 2^{1+\nu^*} \sum_{k=1}^{d} (p_{k|\mathrm{H}_0})^{\lambda^*}(p_{k|\mathrm{H}_1})^{1-\lambda^*}$ であり，よって $2^{1+\nu^*} = 1 / \sum_{k=1}^{d} (p_{k|\mathrm{H}_0})^{\lambda^*}(p_{k|\mathrm{H}_1})^{1-\lambda^*}$ となる。以上から，次を得る。

$$p_l^* = \frac{(p_{l|\mathrm{H}_0})^{\lambda^*}(p_{l|\mathrm{H}_1})^{1-\lambda^*}}{\sum_{k=1}^{d} (p_{k|\mathrm{H}_0})^{\lambda^*}(p_{k|\mathrm{H}_1})^{1-\lambda^*}} = p_l(\lambda^*), \quad l = 1, 2, \ldots, d \tag{A.23}$$

ゆえに $p^* = p(\lambda^*)$ が成り立つ。

さらに，もし $D_{\mathrm{KL}}(p(\lambda^*) \| p_{|\mathrm{H}_0}) \neq D_{\mathrm{KL}}(p(\lambda^*) \| p_{|\mathrm{H}_1})$ ならば，(iii) より $\lambda^* = 0$ とならなければならず，ゆえに (A.23) より $p^* = p(0) = p_{|\mathrm{H}_1}$ となる。これと (ii) を用いると

$$0 = D_{\mathrm{KL}}(p_{|\mathrm{H}_1} \| p_{|\mathrm{H}_1}) = D_{\mathrm{KL}}(p^* \| p_{|\mathrm{H}_1}) \overset{\text{(ii)}}{>} D_{\mathrm{KL}}(p^* \| p_{|\mathrm{H}_0}) \overset{\substack{\textbf{定理 7.24–(1)} \\ (\text{➔ } \textbf{p. 203})}}{\geqq} 0$$

となってしまい，矛盾する。よって $D_{\mathrm{KL}}(p(\lambda^*) \| p_{|\mathrm{H}_0}) = D_{\mathrm{KL}}(p(\lambda^*) \| p_{|\mathrm{H}_1})$ が成り立たつ。

同様に，最小化問題 (1) の最適解 $q^* = (q_1^*, q_2^*, \ldots, q_d^*)$ は，やはりある ρ^* を用いて次のように表せる。

$$q_l^* = \frac{(p_{l|\mathrm{H}_0})^{1-\rho^*}(p_{l|\mathrm{H}_1})^{\rho^*}}{\sum_{k=1}^{d} (p_{k|\mathrm{H}_0})^{1-\rho^*}(p_{k|\mathrm{H}_1})^{\rho^*}} = p_l(1 - \rho^*), \quad l = 1, 2, \ldots, d$$

また，やはり $D_{\mathrm{KL}}(q^* \| p_{|\mathrm{H}_0}) = D_{\mathrm{KL}}(q^* \| p_{|\mathrm{H}_1})$，つまり $D_{\mathrm{KL}}(p(1-\rho^*) \| p_{|\mathrm{H}_0}) = D_{\mathrm{KL}}(p(1-\rho^*) \| p_{|\mathrm{H}_1})$ が成り立つ。

さて，式 (A.22) (➔ **p. 34**) に戻る前に，$D_{\mathrm{KL}}(p(\lambda^*) \| p_{|\mathrm{H}_0}) = D_{\mathrm{KL}}(p(\lambda^*) \| p_{|\mathrm{H}_1})$，$D_{\mathrm{KL}}(p(1-\rho^*) \| p_{|\mathrm{H}_0}) = D_{\mathrm{KL}}(p(1-\rho^*) \| p_{|\mathrm{H}_1})$ と Chernoff 情報量の関係を見ておこう。まず，Chernoff 情報量の定義 (➔ **p. 290**) は

$$\mathrm{CI}(p_{l|\mathrm{H}_0}, p_{l|\mathrm{H}_1}) = -\min_{0 \leqq \lambda \leqq 1} \log_2 \sum_{l=1}^{d} (p_{l|\mathrm{H}_0})^{\lambda}(p_{l|\mathrm{H}_1})^{1-\lambda}$$

であった。これを計算するには，対数関数 $\log_2 x$ が単調増加であることより，非負の関数 $g(\lambda)$ を

$$g(\lambda) = \sum_{l=1}^{d} (p_{l|\mathrm{H}_0})^{\lambda}(p_{l|\mathrm{H}_1})^{1-\lambda}, \quad \lambda \in \mathbb{R}$$

とおいて，この最小値を $\log_2 x$ に放り込めばよい。微分公式 $\frac{\mathrm{d}}{\mathrm{d}\lambda} a^{\lambda} = a^{\lambda} \frac{\log_2 a}{\log_2 \mathrm{e}}$ を用いると，$g'(\lambda)$ と

$g''(\lambda)$ は次のように計算できる。

$$
\begin{aligned}
g'(\lambda) &= \sum_{l=1}^{d}(p_{l|\mathrm{H}_0})^{\lambda}(p_{l|\mathrm{H}_1})^{1-\lambda}(\log_2 p_{l|\mathrm{H}_0} - \log_2 p_{l|\mathrm{H}_1}) \\
&= \sum_{l=1}^{d}(p_{l|\mathrm{H}_0})^{\lambda}(p_{l|\mathrm{H}_1})^{1-\lambda}\log_2\frac{p_l(\lambda)}{p_{l|\mathrm{H}_1}} - \sum_{l=1}^{d}(p_{l|\mathrm{H}_0})^{\lambda}(p_{l|\mathrm{H}_1})^{1-\lambda}\log_2\frac{p_l(\lambda)}{p_{l|\mathrm{H}_0}} \\
&\overset{\substack{p_l(\lambda)\text{ の定義}\\\text{を用いた}}}{=} \left(\sum_{k=1}^{d}(p_{k|\mathrm{H}_0})^{\lambda}(p_{k|\mathrm{H}_1})^{1-\lambda}\right)\Biggl\{\underbrace{\sum_{l=1}^{d}p_l(\lambda)\log_2\frac{p_l(\lambda)}{p_{l|\mathrm{H}_1}}}_{=D_{\mathrm{KL}}(p(\lambda)\|p_{|\mathrm{H}_1})} - \underbrace{\sum_{l=1}^{d}p_l(\lambda)\log_2\frac{p_l(\lambda)}{p_{l|\mathrm{H}_0}}}_{=D_{\mathrm{KL}}(p(\lambda)\|p_{|\mathrm{H}_0})}\Biggr\},
\end{aligned}
$$

$$
g''(\lambda) = \sum_{l=1}^{d}(p_{l|\mathrm{H}_0})^{\lambda^*}(p_{l|\mathrm{H}_1})^{1-\lambda^*}\bigl(\log_2(p_{l|\mathrm{H}_0}) - \log_2(p_{l|\mathrm{H}_1})\bigr)^2 \overset{\substack{\text{仮定}\\ p_{|\mathrm{H}_0}\neq p_{|\mathrm{H}_1}\\ \text{より}}}{>} 0
$$

特に $g''(\lambda) > 0$ であるから，$g'(\lambda_0) = 0$ をみたす点 $\lambda = \lambda_0$ はただ一つしかなく，これが $g(\lambda)$ の最小点である。さらに，$g'(\lambda)$ の計算式より $g'(\lambda_0) = 0 \Leftrightarrow D_{\mathrm{KL}}(p(\lambda_0)\|p_{|\mathrm{H}_0}) = D_{\mathrm{KL}}(p(\lambda_0)\|p_{|\mathrm{H}_1})$ であることがわかる。ゆえに，$g(\lambda)$ の最小点は $\lambda_0 = \lambda^*$ で与えられることがわかった。さらに，$D_{\mathrm{KL}}(p(\lambda)\|p_{|\mathrm{H}_0}) = D_{\mathrm{KL}}(p(\lambda)\|p_{|\mathrm{H}_1})$ をみたす λ の値もまた，ただ一つしかないことになるので，特に $0 \leqq \lambda^* = 1-\rho^* \leqq 1$ であることがわかった。よって $p^* = p(\lambda^*) = p(1-\rho^*) = q^*$ である (**最小化問題 (a) と (b) の解は同じになってしまう!**)。

さて，$p(\lambda)$ の定義から，各 $l = 1, 2, \ldots, d$ について

$$
\log p_l(\lambda) = \lambda\log p_{l|\mathrm{H}_0} + (1-\lambda)\log p_{l|\mathrm{H}_1} - \log_2\sum_{k=1}^{d}(p_{k|\mathrm{H}_0})^{\lambda}(p_{k|\mathrm{H}_1})^{1-\lambda}
$$

が成り立つが，特に $\lambda = \lambda^*$ においては次のように書き直すことができる。

$$
\log p_l(\lambda^*) = \lambda^*\log p_{l|\mathrm{H}_0} + (1-\lambda^*)\log p_{l|\mathrm{H}_1} + \mathrm{CI}(p_{|\mathrm{H}_0}, p_{|\mathrm{H}_1}), \quad l = 1, 2, \ldots, d
$$

これを整理して

$$
\log\frac{p_l(\lambda^*)}{p_{l|\mathrm{H}_1}} = \lambda^*\log\frac{p_{l|\mathrm{H}_0}}{p_{l|\mathrm{H}_1}} + \mathrm{CI}(p_{|\mathrm{H}_0}, p_{|\mathrm{H}_1}), \quad l = 1, 2, \ldots, d
$$

となり，この辺々に $p_l(\lambda^*)$ を掛けたのち，$l = 1, 2, \ldots, d$ に関して和をとると，次を得る。

$$
\underbrace{\sum_{l=1}^{d}p_l(\lambda)\log\frac{p_l(\lambda^*)}{p_{l|\mathrm{H}_1}}}_{\substack{\| \\ D_{\mathrm{KL}}(p(\lambda^*\|p_{|\mathrm{H}_1})}} = \lambda^*\underbrace{\sum_{l=1}^{d}p_l(\lambda)\log\frac{p_{l|\mathrm{H}_0}}{p_{l|\mathrm{H}_1}}}_{\text{KKT 条件 (iii) より }0} + \mathrm{CI}(p_{|\mathrm{H}_0}, p_{|\mathrm{H}_1})
$$

よって $\mathrm{CI}(p_{|\mathrm{H}_0}, p_{|\mathrm{H}_1}) = D_{\mathrm{KL}}(p(\lambda^*\|p_{|\mathrm{H}_1}) = D_{\mathrm{KL}}(p(\lambda^*\|p_{|\mathrm{H}_0})$ が成り立つ。
以上から，式 (A.22) (➲ **p. 34**) により，

$$
\begin{aligned}
&\lim_{n\to\infty}\left\{-\frac{1}{n}\log_2\min_{\phi:\mathcal{X}^n\to\{0,1\}}\mathbf{P}(\text{検定 }\phi\text{ により選択された仮説は誤り})\right\} \\
&= \lim_{n\to\infty}\left\{-\frac{1}{n}\log_2\left(\pi_0\cdot 2^{-nD_{\mathrm{KL}}(p^*\|p_{|\mathrm{H}_0})} + \pi_1\cdot 2^{-nD_{\mathrm{KL}}(p^*\|p_{|\mathrm{H}_1})}\right)\right\} \\
&= D_{\mathrm{KL}}(p^*\|p_{|\mathrm{H}_0}) = D_{\mathrm{KL}}(p^*\|p_{|\mathrm{H}_1}) = \mathrm{CI}(p_{l|\mathrm{H}_0}, p_{l|\mathrm{H}_1})
\end{aligned}
$$

定理 B.16(➔「付録　行列代数」 p. 6) の証明

(1)⇒(2): 2 次正方行列 A が正則であると仮定すると，A の逆行列 A^{-1} が存在する。そこで公式 B.12 (➔ p. 4) を用いると $(\det A)(\det(A^{-1})) = \det(AA^{-1}) = \det E_2 = 1 \neq 0$ を得る。ゆえに $\det A \neq 0$ である。

(2)⇒(3): $\det A \neq 0$ と仮定すると，その逆数 $\frac{1}{\det A}$ が考えられる。そこで $X = \dfrac{1}{\det A}\begin{pmatrix} d & -b \\ -c & a \end{pmatrix}$ とおくと，次が成り立つ。

$$XA = \frac{1}{\det A}\begin{pmatrix} d & -b \\ -c & a \end{pmatrix}\begin{pmatrix} a & b \\ c & d \end{pmatrix} = \frac{1}{\det A}\begin{pmatrix} \det A & 0 \\ 0 & \det A \end{pmatrix} = E_2$$

(3)⇒(4): $XA = E_2$ と仮定すると，$(\det X)(\det A) = \det(XA) = \det E_2 = 1 \neq 0$ であるから $\det A \neq 0$ である。そこで $Y = \dfrac{1}{\det A}\begin{pmatrix} d & -b \\ -c & a \end{pmatrix}$ とおくと，次が成立する。

$$AY = \frac{1}{\det A}\begin{pmatrix} a & b \\ c & d \end{pmatrix}\begin{pmatrix} d & -b \\ -c & a \end{pmatrix} = \frac{1}{\det A}\begin{pmatrix} \det A & 0 \\ 0 & \det A \end{pmatrix} = E_2$$

特に，練習問題 B.5 (➔ p. 2) より $X = Y = \dfrac{1}{\det A}\begin{pmatrix} d & -b \\ -c & a \end{pmatrix}$ である。

(4)⇒(1): $AX = E_n$ と仮定すると，$(\det A)(\det X) = \det(AX) = \det E_2 = 1 \neq 0$ であるから $\det A \neq 0$ となる。このとき，上と同様にして次を得る。

$$A\left(\frac{1}{\det A}\begin{pmatrix} d & -b \\ -c & a \end{pmatrix}\right) = E_2 = \frac{1}{\det A}\begin{pmatrix} d & -b \\ -c & a \end{pmatrix}A$$

つまり A は正則であり，逆行列は $A^{-1} = \dfrac{1}{\det A}\begin{pmatrix} d & -b \\ -c & a \end{pmatrix}$ で与えられる。また，問 B.5 (➔ p. 2) より $X = A^{-1}$ である。

公式 B.51(➔「付録　行列代数」 p. 26) の証明

(1) $\langle \boldsymbol{x}, \overline{A}^\top \boldsymbol{y} \rangle = \boldsymbol{x}^\top \overline{\overline{A}^\top \boldsymbol{y}} = \boldsymbol{x}^\top \overline{\overline{A}}^\top \overline{\boldsymbol{y}} = \boldsymbol{x}^\top A^\top \overline{\boldsymbol{y}} = (\boldsymbol{x}^\top A^\top)\overline{\boldsymbol{y}} = (A\boldsymbol{x})^\top \overline{\boldsymbol{y}} = \langle A\boldsymbol{x}, \boldsymbol{y} \rangle$

(2) $\overline{\langle \boldsymbol{y}, \boldsymbol{x} \rangle} = \overline{\boldsymbol{y}^\top \overline{\boldsymbol{x}}} = \overline{\boldsymbol{y}}^\top \overline{\overline{\boldsymbol{x}}} = \overline{\boldsymbol{y}}^\top \boldsymbol{x} = \overline{\boldsymbol{y}}^\top (\boldsymbol{x}^\top)^\top = (\boldsymbol{x}^\top \overline{\boldsymbol{y}})^\top = (\boldsymbol{x}^\top \overline{\boldsymbol{y}})^\top = (\langle \boldsymbol{x}, \boldsymbol{y} \rangle)^\top = \langle \boldsymbol{x}, \boldsymbol{y} \rangle$ となる。ここで，最後の等号は $\langle \boldsymbol{x}, \boldsymbol{y} \rangle$ が数 (つまり $(1,1)$ 行列) であることから従う。

(3) $\langle \boldsymbol{x} + \boldsymbol{y}, \boldsymbol{z} \rangle = \sum_{k=1}^n (\boldsymbol{x} + \boldsymbol{y})_k \overline{z_k} = \sum_{k=1}^n (x_k + y_k)\overline{z_k} = \sum_{k=1}^n (x_k \overline{z_k} + y_k \overline{z_k}) = \sum_{k=1}^n x_k \overline{z_k} + \sum_{k=1}^n y_k \overline{z_k} = \langle \boldsymbol{x}, \boldsymbol{z} \rangle + \langle \boldsymbol{y}, \boldsymbol{z} \rangle$ と変形できる。また，(2) を用いて $\langle \boldsymbol{x}, \boldsymbol{y} + \boldsymbol{z} \rangle = \overline{\langle \boldsymbol{y} + \boldsymbol{z}, \boldsymbol{x} \rangle} = \overline{\langle \boldsymbol{y}, \boldsymbol{x} \rangle + \langle \boldsymbol{z}, \boldsymbol{x} \rangle} = \overline{\langle \boldsymbol{y}, \boldsymbol{x} \rangle} + \overline{\langle \boldsymbol{z}, \boldsymbol{x} \rangle} = \langle \boldsymbol{x}, \boldsymbol{y} \rangle + \langle \boldsymbol{x}, \boldsymbol{z} \rangle$ である。

(4) $\langle c\boldsymbol{x}, \boldsymbol{y} \rangle = \sum_{k=1}^n (c\boldsymbol{x})_k \overline{y_k} = \sum_{k=1}^n (cx_k)\overline{y_k} = c\sum_{k=1}^n x_k \overline{y_k} = c\langle \boldsymbol{x}, \boldsymbol{y} \rangle$ であり，また，(2) を用いて $\langle \boldsymbol{x}, c\boldsymbol{y} \rangle = \overline{\langle c\boldsymbol{y}, \boldsymbol{x} \rangle} = \overline{c\langle \boldsymbol{y}, \boldsymbol{x} \rangle} = \overline{c}\,\overline{\langle \boldsymbol{y}, \boldsymbol{x} \rangle} = \overline{c}\langle \boldsymbol{x}, \boldsymbol{y} \rangle$ が成り立つ。

(5) $\langle \boldsymbol{x}, \boldsymbol{x} \rangle = \sum_{k=1}^n x_k \overline{x_k} = \sum_{k=1}^n |x_k|^2 \geqq 0$ である。この等号が成り立つための必要十分条件は $|x_1|^2 = |x_2|^2 = \cdots = |x_n|^2 = 0$，つまり $x_1 = x_2 = \cdots = x_n = 0$ であるから $\boldsymbol{x} = \boldsymbol{0}$ である。

問 2.11(➔p. 20) の解答例

① $|\overline{x} - Q_2| \leqq \sqrt{v}$ **について。** これは成り立つ。実際，分散公式を用いると，まず次の式変形を得る。

$$|\overline{x} - Q_2| = |\,(\overline{x - Q_2})\,| = \left|\frac{1}{n}\sum_{i=1}^{n}(x_i - Q_2)\right| \leqq \frac{1}{n}\sum_{i=1}^{n}|x_i - Q_2| = L_1(Q_2) \leqq L_1(\overline{x})$$

この評価を二乗すると次が得られる。

$$(\overline{x} - Q_2)^2 \leqq \left(\frac{1}{n}\sum_{i=1}^{n}|x_i - \overline{x}|\right)^2 \leqq \frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^2 = v$$

ここで，二つ目の不等式には公式 2.10 (➔p. 19) 直後に紹介した不等式を，1 次元データ $(|x_1 - \overline{x}|, |x_2 - \overline{x}|, \ldots, |x_n - \overline{x}|)$ に対して適用した。上の不等式において辺々の平方根をとることで $|\overline{x} - Q_2| \leqq \sqrt{v}$ が成り立つことがわかる。

② $|\overline{x} - \mathrm{mode}(x)| \leqq \sqrt{2}\sqrt{v}$ **について。** これは以下の議論を通して，成立が十分に期待できる。

まず分散公式より $v_x = v_{x-\mathrm{mode}(x)} = \dfrac{1}{n}\displaystyle\sum_{i=1}^{n}(x_i - \mathrm{mode}(x))^2 - (\overline{x} - \mathrm{mode}(x))^2$ であるから

$$(\overline{x} - \mathrm{mode}(x))^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \mathrm{mode}(x))^2 - v_x$$

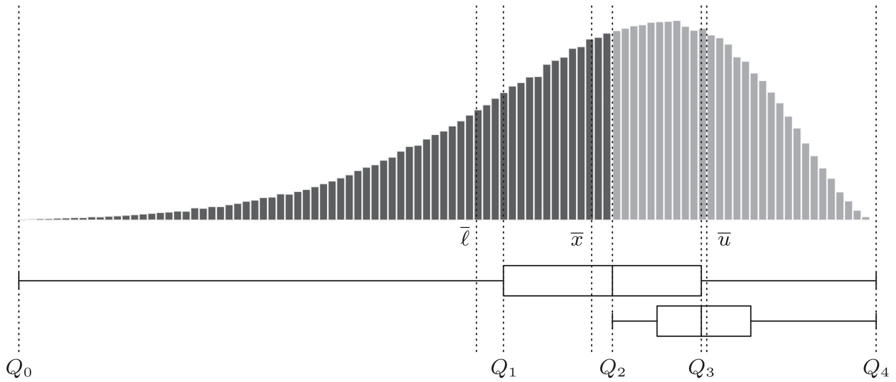
ここで，n が偶数の場合を考えよう (n が奇数の場合も同様の議論ができる)。m を自然数として，$n = 2m$ とする。x の下位のデータを $\ell = (\ell_1, \ell_2, \ldots, \ell_m)$ (小さい順に並べる。)，上位のデータを $u = (u_1, u_2, \ldots, u_m)$ (大きい順に並べる。) とおく。このとき，

$$\begin{aligned}(\overline{x} - \mathrm{mode}(x))^2 &= \frac{1}{n}\sum_{i=1}^{n}(x_i - \mathrm{mode}(x))^2 - v_x \\ &= \underbrace{\frac{1}{n}\sum_{i=1}^{m}(\ell_i - \mathrm{mode}(x))^2}_{\substack{= A \\ \text{とおく。}}} + \underbrace{\frac{1}{n}\sum_{i=1}^{m}(u_i - \mathrm{mode}(x))^2}_{\substack{= B \\ \text{とおく。}}} - v_x\end{aligned}$$

まず右辺の第 1 項に注目する。再び分散公式より，次の変形が得られる。

$$A = \frac{1}{n}\sum_{i=1}^{m}(\ell_i - \mathrm{mode}(x))^2 = \frac{1}{2m}\sum_{i=1}^{m}(\ell_i - \mathrm{mode}(x))^2 = \frac{v_\ell + (\overline{\ell} - \mathrm{mode}(x))^2}{2}$$

ここで，箱ひげ図に基づいて ℓ と u それぞれのヒストグラムに分けると，次のようになる。



1 次元データ ℓ のヒストグラム (左側) は，山が一つで右側に寄っている。したがって $\overline{\ell} \leqq Q_2(\ell) = Q_1(x) < Q_2(x) < \mathrm{mode}(x)$ (下位のデータ ℓ の中央値は，全体のデータ x の第 1 四分位数である) であると期待でき，ゆえに $|\overline{\ell} - \mathrm{mode}(x)| = \mathrm{mode}(x) - \overline{\ell}$ であろう。また 1 次元データ u のヒストグラムに注目すると，山が一つで左側に寄っている (これは x のヒストグラムの山が一つで，右端が切り立った崖のようになっていなければ期待できることではないか。) から，$\mathrm{mode}(x) \overset{\text{仮定}}{<} Q_3(x) = Q_2(u) < \overline{u}$ (上位のデータ u の中央値は，全体のデータ x の第 3 四分位数である) であることが期待できる。よって，次式が成り立つことが期待できる。

$$|\overline{\ell} - \mathrm{mode}(x)| = \mathrm{mode}(x) - \overline{\ell} < \overline{u} - \overline{\ell} = \overline{u - \ell}$$

これにより，

$$A = \frac{1}{2}\big(v_\ell + (\overline{\ell} - \mathrm{mode}(x))^2\big) < \frac{1}{2}\big(v_\ell + (\overline{u-\ell})^2\big) \overset{\substack{\text{分散}\\\text{公式}}}{=} \frac{1}{2}\left(v_\ell + \frac{1}{m}\sum_{i=1}^{m}(u_i - \ell_i)^2 - v_{u-\ell}\right)$$

ここで，右辺の第 2 項について $u_i - \ell_i = (u_i - \overline{x}) + (\overline{x} - \ell_i)$ と分解すると，

$$\begin{aligned}\frac{1}{m}\sum_{i=1}^{m}(u_i - \ell_i)^2 &= \frac{1}{m}\sum_{i=1}^{m}(u_i - \overline{x})^2 - \frac{2}{m}\sum_{i=1}^{m}(u_i - \overline{x})(\ell_i - \overline{x}) + \frac{1}{m}\sum_{i=1}^{m}(\ell_i - \overline{x})^2 \\ &\leqq \frac{1}{m}\sum_{i=1}^{m}(u_i - \overline{x})^2 + \frac{2}{m}\sum_{i=1}^{m}|u_i - \overline{x}|\cdot|\ell_i - \overline{x}| + \frac{1}{m}\sum_{i=1}^{m}(\ell_i - \overline{x})^2\end{aligned}$$

であり，相加相乗の不等式より $|u_i - \overline{x}|\cdot|\ell_i - \overline{x}| = \sqrt{(u_i - \overline{x})^2(\ell_i - \overline{x})^2} \leqq \dfrac{(u_i - \overline{x})^2 + (\ell_i - \overline{x})^2}{2}$ であるから，

$$\begin{aligned}\frac{1}{m}\sum_{i=1}^{m}(u_i - \ell_i)^2 &\leqq \frac{1}{m}\sum_{i=1}^{m}(u_i - \overline{x})^2 + \frac{2}{m}\sum_{i=1}^{m}\frac{(u_i - \overline{x})^2 + (\ell_i - \overline{x})^2}{2} + \frac{1}{m}\sum_{i=1}^{m}(\ell_i - \overline{x})^2 \\ &= \frac{2}{m}\left(\sum_{i=1}^{m}(u_i - \overline{x})^2 + \sum_{i=1}^{m}(\ell_i - \overline{x})^2\right) \leqq \frac{2\cdot n}{m}v_x = 4\cdot v_x\end{aligned}$$

一方で，u と ℓ は負の相関をもつことに注意すると，$v_{u-\ell} = v_u - 2\cdot s_{u,\ell} + v_\ell \geqq v_u + v_\ell$。よって

$$A < \frac{1}{2}(v_\ell + 4\cdot v_x - (v_u + v_\ell)) = \frac{1}{2}\big(4\cdot v_x - \underbrace{v_u}_{\geqq 0}\big) \leqq 2\cdot v_x$$

次に $B = \dfrac{1}{n}\displaystyle\sum_{i=1}^{m}(u_i - \mathrm{mode}(x))^2$ を評価する。まず x のヒストグラムの山が一つで左に寄っていることから，$\overline{x} < Q_2 \overset{\text{仮定}}{<} \mathrm{mode}(x)$ であることが期待され，このとき $|u_i - \mathrm{mode}(x)| \leqq |u_i - \overline{x}|$ $(i = 1, 2, \ldots, m)$ が成り立つ。ゆえに $B \leqq \dfrac{1}{n}\displaystyle\sum_{i=1}^{m}(u_i - \overline{x})^2 \leqq v_x$。
以上から，次の不等式が成り立つことが期待できるのである。

$$(\overline{x} - \mathrm{mode}(x))^2 = A + B - v_x \leqq 2 \cdot v_x + v_x - v_x = 2 \cdot v_x$$

問 2.15(➔p. 23) の解答例

共分散の定義より，任意の実数 a と b に対して次が成り立つ。

$$\begin{aligned}
s_{x+a1_n, y+b1_n} &= \frac{1}{n}\sum_{i=1}^{n}\big((x_i + a) - \overline{x+a}\big)\big((y_i + b) - \overline{y+b}\big) \\
&= \frac{1}{n}\sum_{i=1}^{n}\big((x_i + a) - (\overline{x} + a)\big)\big((y_i + b) - (\overline{y} + b)\big) \\
&= \frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y}) = s_{x,y}
\end{aligned}$$

また $s_{x,y} = \dfrac{1}{n}\displaystyle\sum_{i=1}^{n} x_i y_i - \overline{x}\,\overline{y}$ であるから，次が成り立つ。

$$s_{x,y} = s_{x-a1_n, y-b1_n} = \frac{1}{n}\sum_{i=1}^{n}(x_i - a)(y_i - b) - (\overline{x} - a)(\overline{y} - b)$$

特に $a = x_j$，$b = y_j$ をとると $s_{x,y} = \dfrac{1}{n}\displaystyle\sum_{i=1}^{n}(x_i - x_j)(y_i - y_j) - (\overline{x} - x_j)(\overline{y} - y_j)$ であり，これを $j = 1, 2, \ldots, n$ について平均をとると $s_{x,y} = \dfrac{1}{n^2}\displaystyle\sum_{i,j=1,2,\ldots,n}(x_i - x_j)(y_i - y_j) - s_{x,y}$。ゆえに

$$\begin{aligned}
2\,s_{x,y} &= \frac{1}{n^2}\sum_{i,j=1,2,\ldots,n}(x_i - x_j)(y_i - y_j) \\
&= \frac{1}{n^2}\sum_{\substack{i,j=1,2,\ldots,n: \\ i<j}}(x_i - x_j)(y_i - y_j) + \frac{1}{n^2}\sum_{\substack{i,j=1,2,\ldots,n: \\ i>j}}(x_i - x_j)(y_i - y_j) \\
&= \frac{2}{n^2}\sum_{\substack{i,j=1,2,\ldots,n: \\ i<j}}(x_i - x_j)(y_i - y_j)
\end{aligned}$$

両辺を 2 で割ることで $s_{x,y} = \dfrac{1}{n^2}\displaystyle\sum_{\substack{i,j=1,2,\ldots,n: \\ i<j}}(x_i - x_j)(y_i - y_j)$ が得られる。

コメント　これにより x **と** y **が共に単調増加あるいは共に単調減少ならば** $s_{x,y} \geqq 0$ **であり，片方が単調増加でもう片方が単調減少ならば** $s_{x,y} \leqq 0$ **である**ことがわかる。また前者の $x_1 \leqq x_2 \leqq \cdots \leqq x_n$，$y_1 \leqq y_2 \leqq \cdots \leqq y_n$ であるとき，$0 \leqq s_{x,y} = \dfrac{1}{n}\displaystyle\sum_{i=1}^{n} x_i y_i - \overline{x}\,\overline{y}$ より，次の不等式が得られる。

$$\frac{1}{n}\sum_{i=1}^{n} x_i y_i \geqq \overline{x}\,\overline{y}$$

これは **Chebyshev の和の不等式**として知られているものである。

問 B.3(➔p. 37) の解答例

$X \sim \mathrm{Bernoulli}(p)$ のとき，$\displaystyle\sum_{x:\ X\ \text{が取り得る値}} \mathbf{P}(X=x) = \mathbf{P}(X=1) + \mathbf{P}(X=0) = p + (1-p) = 1$。

$X \sim \mathrm{binomial}(n,p)$ のとき，

$$\begin{aligned}\sum_{x:\ X\ \text{が取り得る値}} \mathbf{P}(X=x) &= \sum_{x=0,1,2,\ldots,n} \mathbf{P}(X=x)\\ &= \sum_{x=0,1,2,\ldots,n} \binom{n}{x} p^x (1-p)^{1-x} = (p+(1-p))^n = 1\end{aligned}$$

$X \sim \mathrm{geometric}(p)$ のとき，$0<p<1$ であることに注意すると次のように計算できる。

$$\begin{aligned}\sum_{x:\ X\ \text{が取り得る値}} \mathbf{P}(X=x) &= \sum_{x=1,2,\ldots} \mathbf{P}(X=x)\\ &= \lim_{n\to\infty} \sum_{x=1,2,\ldots,n} \mathbf{P}(X=x)\\ &= \lim_{n\to\infty} \sum_{x=1,2,\ldots,n} (1-p)^{x-1} p\\ &= p\cdot\lim_{n\to\infty} \sum_{x=0,1,2,\ldots,n-1} (1-p)^x = p\cdot\lim_{n\to\infty} \frac{1-(1-p)^{n-1}}{1-(1-p)} = p\cdot\frac{1}{p} = 1\end{aligned}$$

問 2.46(➔p. 48) の証明

$N \sim \mathrm{Poisson}(\lambda)$ のとき，

$$\begin{aligned}\mathbf{E}[N] &= \sum_{k=0,1,2,\ldots} k\,\mathbf{P}(N=k) = \sum_{k=1,2,\ldots} k\,\mathbf{P}(N=k) = \sum_{k=1,2,\ldots} k\,\frac{\lambda^k}{k!}\mathrm{e}^{-\lambda}\\ &= \lambda\,\mathrm{e}^{-\lambda} \sum_{k=1,2,\ldots} \frac{\lambda^{k-1}}{(k-1)!} = \lambda\,\mathrm{e}^{-\lambda} \sum_{k=0,1,\ldots} \frac{\lambda^k}{k!} = \lambda\,\mathrm{e}^{-\lambda}\mathrm{e}^{\lambda} = \lambda\end{aligned}$$

問 2.48(➔p. 50) の解答例

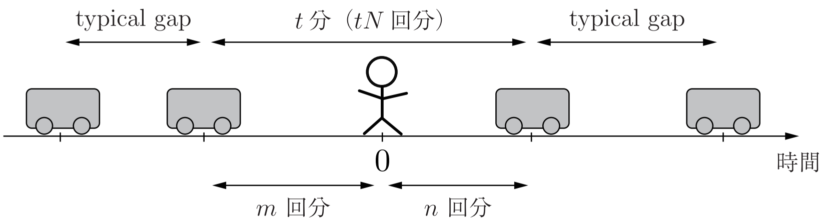
N を自然数として，時間間隔を $\delta = \dfrac{1}{N}$〔分〕刻みで考えてみよう。時刻 $\ldots, -3\delta, -2\delta, -\delta, 0, \delta, 2\delta, 3\delta, \ldots$ のそれぞれにおいて神様が確率 p で表が出るコインを投げ，表が出ればその時刻にバスはやって来ず，裏が出ればその時刻にバスが到着して客を乗せて出発すると考える。$q = 1-p$ とおく。練習問題 2.44 (➔p. 45) の考え方を踏襲すると，$p = 1 - \dfrac{\lambda}{N}$，$q = \dfrac{\lambda}{N}$ であることがわかる。
A 君 がバス停に到着した時刻 0 におけるコイン投げを「0 回目のコイン投げ」とよぶことにすると，

次が成り立つ。

$$\begin{aligned}&\mathbf{P}(n\text{ 回目に初めてバスが到着する})\\&= \underbrace{p}_{\substack{\text{時刻 }0\text{ では}\\\text{バスは来ない。}}} \times \underbrace{p^{n-1}}_{\substack{\text{それからコイン投げ}\\(n-1)\text{ 回分}\\\text{バスが来ない。}}} \times \underbrace{q}_{\substack{n\text{ 回目のコイン投げ}\\\text{の結果，バスが}\\\text{やってくる。}}} = q\cdot p^n\end{aligned}$$

同様にして次が成り立つ。

$$\mathbf{P}(\text{一個前のバスは }m\text{ 回分前に到着していた}) = q\cdot p^{m-1}$$



ゆえに次が成り立つ。

$$\begin{aligned}&\mathbf{P}\Big(\text{ A 君がバス停に到着した時刻 0 は }t\text{ 分間 }(=tN\text{ 回分})\text{ の gap 内にある }\Big)\\&= \sum_{\substack{m>0,\,n\geqq 0:\\ m+n=tN}} (q\cdot p^{m-1})(q\cdot p^n) = \sum_{\substack{m>0,\,n\geqq 0:\\ m+n=tN}} q^2\cdot p^{m+n-1} = q^2\cdot p^{tN-1} \sum_{\substack{m>0,\,n\geqq 0:\\ m+n=tN}} 1\\&= tN\cdot q^2\cdot p^{tN-1} = tN\left(\frac{\lambda}{N}\right)^2\left(1-\frac{\lambda}{N}\right)^{tN}\left(1-\frac{\lambda}{N}\right)^{-1} \overset{\substack{N\to\infty\\\text{のとき}}}{\approx} \frac{t\lambda^2}{N}\mathrm{e}^{-\lambda t}\end{aligned}$$

一方で typical gap が t 分間である確率は，時刻 $-\delta$〔分〕にバスが出発してから t 分後に次のバスが到着する確率として計算できるから，次のように表せる。(t は δ の自然数倍として考える。)

$$\begin{aligned}&\mathbf{P}(\text{typical gap が }t\text{ 分間である})\\&= \mathbf{P}(tN\text{ 回目に次のバスが到着する} \mid 1\text{ 回分前にバスが出発する})\\&= p^{tN}\cdot q = \frac{\lambda}{N}\left(1-\frac{\lambda}{N}\right)^{tN} \overset{\substack{N\to\infty\\\text{のとき}}}{\approx} \frac{\lambda}{N}\mathrm{e}^{-\lambda t}\end{aligned}$$

以上のことから，$N\to\infty$ として神様がコインを投げる時間間隔 δ を限りなく 0 に近づけるとき，λ を比例定数として次の比例関係があることがわかる。

$$\mathbf{P}\big((\text{gap around }\ )\in \mathrm{d}t\big) \propto t\cdot\mathbf{P}\big((\text{typical gap})\in\mathrm{d}t\big)$$

「dt」が t を含む微小区間であることを思い出すと，これは「A 君が t 分間前後の gap にある確率は，typical gap が t 分間前後である確率を t だけブーストしたものに比例する」ことを意味する。つまり A 君がいる gap について，この間隔が長いほど，typical gap が同程度の間隔である確率よりも大きくなる傾向にあることを示唆している。

問 2.68(➔ p. 72) の解説

ここでは，数学における基本的な道具や Stirling の公式

$$n! \approx \sqrt{2\pi n}\left(\frac{n}{\mathrm{e}}\right)^n$$

(記号「≈」については➔ p. 6) と Taylor の定理 (➔ p. 8) を用いて問 2.68 を示そう。

補題 8 二つの数列 $(a_n)_{n=1}^{\infty}$ と $(b_n)_{n=1}^{\infty}$ が $a_n \approx b_n$ をみたすとする。このとき $\lim_{n\to\infty} a_n = c$ ならば $\lim_{n\to\infty} b_n = c$ が成り立つ。

証明. 仮定のもとでは，$n \to \infty$ のとき $b_n = (b_n/a_n)a_n \to 1 \times c = c$ である。 □

上の補題 8 は，「数列 a_n の極限を直接計算することが難しい場合に極限の計算が優しい数列 b_n で，かつ $a_n \approx b_n$ となるものを見つけることができれば，a_n の極限は b_n の極限として計算してよい」ということを意味している。

補題 9 二つの実数列 $(a_n)_{n=1}^{\infty}$ と $(b_n)_{n=1}^{\infty}$ が $a_n \to 0$ かつ $(a_n)^3 b_n \to 0$ をみたすとき，

$$(1+a_n)^{b_n} \approx \exp\left\{\left(a_n - \frac{a_n^2}{2}\right)b_n\right\}$$

が成り立つ。

証明. $f(x) = \log(1+x)$，$x > -1$ とおくと $f'''(x) = \frac{2}{(1+x)^3}$ であるから，$|x| < \frac{1}{2}$ に対しては $\frac{1}{3!}|f'''(x)| < \frac{1}{3!}\frac{2}{(1-1/2)^3} = \frac{16}{6} < 3$ が成り立つ。そこで Taylor の定理を用いれば，$|x| < \frac{1}{2}$ に対して

$$\log(1+x) = x - \frac{x^2}{2} + \varepsilon(x), \qquad |\varepsilon(x)| < 3|x|^3$$

が成り立つ。いま $a_n \to 0$ なのであるから，十分大きな n に対しては常に $|a_n| < \frac{1}{2}$ が成り立ち，ゆえに

$$b_n \log(1+a_n) = \left(a_n - \frac{a_n^2}{2}\right)b_n + \varepsilon(a_n)b_n, \qquad |\varepsilon(a_n)b_n| \leqq 2|a_n^3 b_n| \to 0$$

である。したがって次が結論できる。

$$\frac{(1+a_n)^{b_n}}{\exp\left\{(a_n - \frac{a_n^2}{2})b_n\right\}} = \frac{\exp\{b_n \log(1+a_n)\}}{\exp\left\{(a_n - \frac{a_n^2}{2})b_n\right\}} = \exp\{\varepsilon(a_n)b_n\} \to 1$$

□

これで問 2.68 を証明する準備が整った。問 2.68 のときのように x を実数，n を十分大きい自然数として，さらに $k = k_n$ を

$$\frac{k_n - \frac{n}{2}}{\frac{2}{\sqrt{n}}} \leqq x \leqq \frac{(k_n+1) - \frac{n}{2}}{\frac{2}{\sqrt{n}}} \qquad \text{(A.24)}$$

$$\left(\iff \frac{2x}{\sqrt{n}} + \frac{n}{2} - 1 \leqq k_n \leqq \frac{2x}{\sqrt{n}} + \frac{n}{2}\right)$$

をみたす整数とする ($k = k_n$ は n にも依存していることに注意!)。
$n \to \infty$ のとき，$k = k_n \to \infty$ かつ $n - k \to \infty$ でもあることに注意して Stirling の公式「$n! \approx$

$\sqrt{2\pi n}(\frac{n}{\mathrm{e}})^n$」を用いると

$$
\begin{aligned}
\frac{\sqrt{n}}{2}a_n(k_n) &= \frac{\sqrt{n}}{2}\frac{1}{2^n}\binom{n}{k} = \frac{\sqrt{n}}{2}\frac{1}{2^n}\frac{n!}{k!(n-k)!} \\
&\overset{\text{Stirling の公式}}{\approx} \frac{\sqrt{n}}{2}\frac{1}{2^n}\frac{\sqrt{2\pi n}\left(\frac{n}{e}\right)^n}{\sqrt{2\pi k}\left(\frac{k}{e}\right)^k\sqrt{2\pi(n-k)}\left(\frac{n-k}{e}\right)^{n-k}} \\
&= \frac{1}{2}\sqrt{\frac{n^2}{2\pi k(n-k)}}\frac{n^n}{k^k(n-k)^{n-k}}\frac{1}{2^n} \\
&= \frac{1}{\sqrt{2\pi}} \times \frac{1}{2}\sqrt{\frac{n}{k}\left(1-\frac{k}{n}\right)^{-1}} \times \left(\frac{n}{2k}\right)^k\left(\frac{n}{2(n-k)}\right)^{n-k}
\end{aligned}
$$

となる。この最後の式に現れた二項に注目しよう。$n \to \infty$ のとき，次が成り立つ。

○ (第 2 項) $= \frac{1}{2}\sqrt{\frac{n}{k}\left(1-\frac{k}{n}\right)^{-1}} \to 1$

— 実際，$n \to \infty$ のとき (A.24) の括弧内の不等式より $\frac{k_n}{n} \to \frac{1}{2}$ に注意すると

$$
\frac{1}{2}\sqrt{\frac{n}{k}\left(1-\frac{k}{n}\right)^{-1}} \to \frac{1}{2}\sqrt{\left(\frac{1}{2}\right)^{-1}\left(1-\frac{1}{2}\right)^{-1}} = \frac{1}{2}\sqrt{4} = 1
$$

○ (第 3 項) $= \frac{1}{2}\left(\frac{n}{2k}\right)^k\left(\frac{n}{2(n-k)}\right)^{n-k} \to \mathrm{e}^{-\frac{x^2}{2}}$

— 実際，$y = \frac{k-\frac{n}{2}}{\sqrt{n}}$ とおいたとき，k の動く範囲 (A.24) から辿って y の動く範囲は

$$
\frac{x}{2} - \frac{1}{\sqrt{n}} < y = \frac{k-\frac{n}{2}}{\sqrt{n}} \leqq \frac{x}{2}
$$

となることがわかる。特に

- $y \to \frac{x}{2}$,　$\frac{2y}{\sqrt{n}} \to 0$,　$\frac{y^3}{\sqrt{n}} = (\frac{k-(n/2)}{n^{2/3}})^3 \to 0$ である。

  ∗ 最後の収束についてのみコメントしておこう。式 (A.24) の括弧内の不等式から

$$
\frac{\frac{2x}{\sqrt{n}}-1}{n^{2/3}} \leqq \frac{k-\frac{n}{2}}{n^{2/3}} \leqq \frac{\frac{2x}{\sqrt{n}}}{n^{2/3}}
$$

  であるから $\lim_{n\to\infty}\frac{k-\frac{n}{2}}{n^{2/3}} = 0$ が成り立つ。

- $(\frac{2y}{\sqrt{n}})^3(-\frac{n}{2} \pm y\sqrt{n}) = -\frac{1}{4}\frac{y^3}{\sqrt{n}} \pm 8\frac{y}{\sqrt{n}}\frac{y^3}{\sqrt{n}} \to 0$

上の二点に注意して補題 9 を適用すると

$$
\begin{aligned}
&\left(\frac{n}{2k}\right)^k\left(\frac{n}{2(n-k)}\right)^{n-k} = \left(\frac{2k}{n}\right)^{-k}\left(\frac{2(n-k)}{n}\right)^{-(n-k)} \\
&= \left(1+\frac{2y}{\sqrt{n}}\right)^{-\frac{n}{2}-y\sqrt{n}}\left(1-\frac{2y}{\sqrt{n}}\right)^{-\frac{n}{2}+y\sqrt{n}} \\
&\overset{\substack{\text{補題 9} \\ \text{(p. 43)}}}{\approx} \exp\left\{\left(\frac{2y}{\sqrt{n}} - \frac{(2y/\sqrt{n})^2}{2}\right)\left(-\frac{n}{2}-y\sqrt{n}\right)\right\} \\
&\qquad \times \exp\left\{\left(-\frac{2y}{\sqrt{n}} - \frac{(-2y/\sqrt{n})^2}{2}\right)\left(-\frac{n}{2}+y\sqrt{n}\right)\right\} \\
&= \exp\left(\sqrt{n}y - y^2 + \frac{2y^3}{\sqrt{n}}\right)\exp\left(-\sqrt{n}y - y^2 - \frac{2y^3}{\sqrt{n}}\right) = \mathrm{e}^{-2y^2} \to \mathrm{e}^{-\frac{x^2}{2}}
\end{aligned}
$$

以上から，補題 8 を適用することで $\lim_{n\to\infty}\frac{\sqrt{n}}{2}\frac{1}{2^n}\binom{n}{k} = \frac{\mathrm{e}^{-x^2/2}}{\sqrt{2\pi}}$ であることが示された。

問 4.2(➔p. 107) の解答例

まずは $n=2$ の場合を見てみよう。

$$\begin{aligned}
\mathrm{tr}(AB) &= (AB)_{11} + (AB)_{22} \\
&= (A_{11}B_{11} + A_{12}B_{21}) + (A_{21}B_{12} + A_{22}B_{22}) \\
&\overset{\substack{\text{掛け算の順序を}\\\text{ひっくり返した}}}{=} (B_{11}A_{11} + B_{21}A_{12}) + (B_{12}A_{21} + B_{22}A_{22}) \\
&\overset{\substack{\text{まとめ方}\\\text{を変えた}}}{=} (B_{11}A_{11} + B_{12}A_{21}) + (B_{21}A_{12} + B_{22}A_{22}) \\
&= (BA)_{11} + (BA)_{22} = \mathrm{tr}(BA)
\end{aligned}$$

となり，成り立つことがわかる。
同じ原理で，一般の n の場合は

$$\mathrm{tr}(AB) = \sum_{i=1}^{n}(AB)_{ii} = \sum_{i=1}^{n}\sum_{j=1}^{n}A_{ij}B_{ji} = \sum_{j=1}^{n}\sum_{i=1}^{n}B_{ji}A_{ij} = \sum_{j=1}^{n}(BA)_{jj} = \mathrm{tr}(BA)$$

という計算により確かめられる。

問 4.43(➔p. 144) の解答

N 個の標本点を k 個の空でないクラスターに分ける分け方は，$(N-k)$ 個の白丸「○」(後で白丸を 1 個ずつ各グループに追加するので，その分の白丸を k 個除いておく) と $(k-1)$ 個の仕切り「｜」

$$\underbrace{\bigcirc\bigcirc\cdots\bigcirc}_{(N-k)\,\text{個}}\quad\underbrace{|\ |\ \cdots\ |}_{(k-1)\,\text{個}}$$

を並べ替えて「|」で区切られた各区画 (これは k 個ある)

$$|\bigcirc\bigcirc|\bigcirc|\quad\cdots|\bigcirc$$

に一つずつ白丸 ○ を突っ込んだあと，この白丸たちに 1 から N までの番号を割り振る方法の数だけあるから，全部で $\binom{(N-k)+(k-1)}{k-1}N! = \binom{N-1}{k-1}N!$ 通りある。

問 5.10(➔p. 157) の解答例

(1) 部分積分公式を適用して

$$\Gamma(a+1) = \int_0^\infty x^{(a+1)-1}\mathrm{e}^{-x}\mathrm{d}x = \int_0^\infty x^a(-\mathrm{e}^{-x})'\mathrm{d}x = \left[x^a(-\mathrm{e}^{-x})\right]_0^\infty + a\int_0^\infty x^{a-1}\mathrm{e}^{-x}\mathrm{d}x = a\Gamma(a)$$

(2) $\Gamma(1) = \int_0^\infty \mathrm{e}^{-x}\mathrm{d}x = 1$ である。そこで (1) を繰り返し適用すれば，自然数 n に対して $\Gamma(n+1) = n\Gamma(n) = n(n-1)\Gamma(n-1) = \cdots = n!\cdot\Gamma(1) = n!$ が得られる。

(3) $\Gamma(\tfrac{1}{2}) = \displaystyle\int_0^\infty x^{-1/2}\mathrm{e}^{-x}\mathrm{d}x \overset{\substack{x=y^2\text{ と}\\\text{変数変換}}}{=} 2\int_0^\infty \mathrm{e}^{-y^2}\mathrm{d}y \overset{\substack{\mathrm{e}^{-y^2}\text{ は}\\\text{偶関数}}}{=} \int_{-\infty}^\infty \mathrm{e}^{-y^2}\mathrm{d}y \overset{\substack{\text{例 5.7–(1)}\\\text{(➔p. 155)}}}{=} \sqrt{\pi}$

問 5.13(➔p. 158) の解答例

(1) $x=\sin^2\theta$ と変数変換をすると $\mathrm{d}x = 2\sin\theta\cos\theta\mathrm{d}\theta$ となるから，次が成り立つ。

$$\begin{aligned}\mathrm{B}(a,b) &= \int_0^1 x^{a-1}(1-x)^{b-1}\mathrm{d}x \\ &= \int_0^{\pi/2} (\sin^2\theta)^{a-1}(\cos^2\theta)^{b-1}(2\sin\theta\cos\theta)\mathrm{d}\theta = 2\int_0^{\pi/2}(\sin\theta)^{2a-1}(\cos\theta)^{2b-1}\mathrm{d}\theta\end{aligned}$$

(2) 等式 $\Gamma(a)\Gamma(b)=\mathrm{B}(a,b)\Gamma(a+b)$ を示せばよい。

$$\begin{aligned}\Gamma(a)\Gamma(b) &= \Big(\int_0^\infty x^{a-1}\mathrm{e}^{-x}\mathrm{d}x\Big)\Big(\int_0^\infty y^{b-1}\mathrm{e}^{-y}\mathrm{d}y\Big) \\ &\overset{\substack{x=u^2,\\ y=v^2\\ \text{と変数変換}}}{=} \Big(2\int_0^\infty u^{2a-1}\mathrm{e}^{-u^2}\mathrm{d}u\Big)\Big(2\int_0^\infty v^{2b-1}\mathrm{e}^{-v^2}\mathrm{d}v\Big) = 4\int_{(0,+\infty)\times(0,+\infty)} u^{2a-1}v^{2b-1}\mathrm{e}^{-(u^2+v^2)}\mathrm{d}u\mathrm{d}v\end{aligned}$$

であり，この最後の uv-平面内の第一象限上での 2 重積分について極座標変換 $u=r\sin\theta$，$v=r\cos\theta$ を施すと，次が成り立つ。

$$\begin{aligned}&4\int_{(0,+\infty)\times(0,+\infty)} u^{2a-1}v^{2b-1}\mathrm{e}^{-(u^2+v^2)}\mathrm{d}u\mathrm{d}v \\ &= 4\int_{(0,+\infty)\times(0,\frac{\pi}{2})}(r\cos\theta)^{2a-1}(r\sin\theta)^{2b-1}\mathrm{e}^{-r^2}\underbrace{\left|\det\begin{pmatrix}\frac{\partial u}{\partial r} & \frac{\partial u}{\partial\theta}\\ \frac{\partial v}{\partial r} & \frac{\partial v}{\partial\theta}\end{pmatrix}\right|}_{=r}\mathrm{d}r\mathrm{d}\theta \\ &= \Big(\int_0^\infty (r^2)^{(a+b)-1}\mathrm{e}^{-r^2}(2r)\mathrm{d}r\Big)\underbrace{\Big(2\int_0^{\frac{\pi}{2}}(\cos\theta)^{2a-1}(\sin\theta)^{2b-1}\mathrm{d}\theta\Big)}_{=\mathrm{B}(a,b)}\end{aligned}$$

最後に $z=r^2$ と変数変換すれば，次が得られる。

$$(\text{上式}) = \mathrm{B}(a,b)\int_0^\infty z^{(a+b)-1}\mathrm{e}^{-z}\mathrm{d}z = \mathrm{B}(a,b)\Gamma(a+b)$$

問 5.23(➔p. 165) の解答例

(1)⇒(2): (1) を仮定すると，$X_1, X_2, \ldots, X_n$ は独立であるから，命題 5.5 (➔p. 153) より確率ベクトル $(X_1, X_2, \ldots, X_n)$ の p.d.f. は

$$\begin{aligned}f(x_1,x_2,\ldots,x_n) &= \underbrace{\frac{\exp\big(-\frac{(x_1-\mu_1)^2}{2\sigma_1^2}\big)}{\sqrt{2\pi\sigma_1^2}}}_{\substack{X_1\sim\mathrm{N}(\mu_1,\sigma_1^2)\\ \text{の p.d.f.}}} \times \underbrace{\frac{\exp\big(-\frac{(x_2-\mu_2)^2}{2\sigma_2^2}\big)}{\sqrt{2\pi\sigma_2^2}}}_{\substack{X_2\sim\mathrm{N}(\mu_2,\sigma_2^2)\\ \text{の p.d.f.}}} \times\cdots\times \underbrace{\frac{\exp\big(-\frac{(x_n-\mu_n)^2}{2\sigma_n^2}\big)}{\sqrt{2\pi\sigma_n^2}}}_{\substack{X_n\sim\mathrm{N}(\mu_n,\sigma_n^2)\\ \text{の p.d.f.}}} \\ &= \frac{\exp\Big(-\frac{1}{2}\Big((\frac{x_1}{\sigma_1})^2+(\frac{x_2}{\sigma_2})^2+\cdots+(\frac{x_n}{\sigma_n})^2\Big)\Big)}{\sqrt{(2\pi\sigma_1^2)(2\pi\sigma_2^2)\cdots(2\pi\sigma_n^2)}}\end{aligned}$$

で与えられる。この指数関数の肩にある 2 次形式を内積の形に書き直せば上式は

$$f(x_1,x_2,\ldots,x_n)=\frac{\exp\Bigl\{-\frac{1}{2}\Bigl\langle\begin{pmatrix}x_1-\mu_1\\x_2-\mu_2\\\vdots\\x_n-\mu_n\end{pmatrix},\begin{pmatrix}\sigma_1^2&0&\cdots&0\\0&\sigma_2^2&\cdots&0\\\vdots&\vdots&\ddots&\vdots\\0&0&\cdots&\sigma_n^2\end{pmatrix}^{-1}\begin{pmatrix}x_1-\mu_1\\x_2-\mu_2\\\vdots\\x_n-\mu_n\end{pmatrix}\Bigr\rangle\Bigr\}}{\sqrt{\det\Bigl(2\pi\begin{pmatrix}\sigma_1^2&0&\cdots&0\\0&\sigma_2^2&\cdots&0\\\vdots&\vdots&\ddots&\vdots\\0&0&\cdots&\sigma_n^2\end{pmatrix}\Bigr)}}$$

となり，これは $\mathrm{N}\Bigl(\begin{pmatrix}\mu_1\\\mu_2\\\vdots\\\mu_n\end{pmatrix},\begin{pmatrix}\sigma_1^2&0&\cdots&0\\0&\sigma_2^2&\cdots&0\\\vdots&\vdots&\ddots&\vdots\\0&0&\cdots&\sigma_n^2\end{pmatrix}\Bigr)$ の p.d.f. に他ならない。

(2)⇒(1): まず $X_1\sim\mathrm{N}(\mu_1,\sigma_1^2)$ であることを示そう。このためにまず，上の変形を逆に辿ることで確率ベクトル $(X_1,X_2,\ldots,X_n)$ の p.d.f. が

$$\begin{aligned}f(x_1,x_2,\ldots,x_n)&=\frac{\exp\Bigl\{-\frac{1}{2}\Bigl\langle\begin{pmatrix}x_1-\mu_1\\x_2-\mu_2\\\vdots\\x_n-\mu_n\end{pmatrix},\begin{pmatrix}\sigma_1^2&0&\cdots&0\\0&\sigma_2^2&\cdots&0\\\vdots&\vdots&\ddots&\vdots\\0&0&\cdots&\sigma_n^2\end{pmatrix}^{-1}\begin{pmatrix}x_1-\mu_1\\x_2-\mu_2\\\vdots\\x_n-\mu_n\end{pmatrix}\Bigr\rangle\Bigr\}}{\sqrt{\det\Bigl(2\pi\begin{pmatrix}\sigma_1^2&0&\cdots&0\\0&\sigma_2^2&\cdots&0\\\vdots&\vdots&\ddots&\vdots\\0&0&\cdots&\sigma_n^2\end{pmatrix}\Bigr)}}\\&=\frac{\exp\bigl(-\frac{(x_1-\mu_1)^2}{2\sigma_1^2}\bigr)}{\sqrt{2\pi\sigma_1^2}}\times\frac{\exp\bigl(-\frac{(x_2-\mu_2)^2}{2\sigma_2^2}\bigr)}{\sqrt{2\pi\sigma_2^2}}\times\cdots\times\frac{\exp\bigl(-\frac{(x_n-\mu_n)^2}{2\sigma_n^2}\bigr)}{\sqrt{2\pi\sigma_n^2}}\end{aligned}\tag{A.25}$$

であることに注意しよう。

いま，実数 a を任意とすると

$$\begin{aligned}\mathbf{P}(X_1\leqq a)=\mathbf{E}[\mathbf{1}_{(-\infty,a]}(X_1)]&=\int_{\mathbb{R}^n}\mathbf{1}_{(-\infty,a]}(x_1)f(x_1,x_2,\ldots,x_n)\mathrm{d}x_1\mathrm{d}x_2\cdots\mathrm{d}x_n\\&\overset{(A.25)}{=}\int_{\mathbb{R}^n}\mathbf{1}_{(-\infty,a]}(x_1)\frac{\mathrm{e}^{-\frac{(x_1-\mu_1)^2}{2\sigma_1^2}}}{\sqrt{2\pi\sigma_1^2}}\frac{\mathrm{e}^{-\frac{(x_2-\mu_2)^2}{2\sigma_2^2}}}{\sqrt{2\pi\sigma_2^2}}\cdots\frac{\mathrm{e}^{-\frac{(x_n-\mu_n)^2}{2\sigma_n^2}}}{\sqrt{2\pi\sigma_n^2}}\mathrm{d}x_1\mathrm{d}x_2\cdots\mathrm{d}x_n\end{aligned}$$

であるが，$\int_{\mathbb{R}^n}(...)\mathrm{d}x_1\cdots\mathrm{d}x_n=\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}(...)\mathrm{d}x_1\cdots\mathrm{d}x_n$ という式変形をすれば

$$
\begin{aligned}
(\text{上式}) &= \left(\int_{-\infty}^{\infty}\mathbf{1}_{(-\infty,a]}(x_1)\frac{\mathrm{e}^{-\frac{(x_1-\mu_1)^2}{2\sigma_1^2}}}{\sqrt{2\pi\sigma_1^2}}\mathrm{d}x_1\right) \\
&\qquad\times\underbrace{\left(\int_{-\infty}^{\infty}\frac{\mathrm{e}^{-\frac{(x_2-\mu_2)^2}{2\sigma_2^2}}}{\sqrt{2\pi\sigma_2^2}}\mathrm{d}x_2\right)}_{=1}\times\cdots\times\underbrace{\left(\int_{-\infty}^{\infty}\frac{\mathrm{e}^{-\frac{(x_n-\mu_n)^2}{2\sigma_n^2}}}{\sqrt{2\pi\sigma_n^2}}\mathrm{d}x_n\right)}_{=1} \\
&= \int_{-\infty}^{a}\frac{\mathrm{e}^{-\frac{(x_1-\mu_1)^2}{2\sigma_1^2}}}{\sqrt{2\pi\sigma_1^2}}\mathrm{d}x_1
\end{aligned}
$$

が得られる。これで等式 $\mathbf{P}(X_1\leqq a)=\int_{-\infty}^{a}\underbrace{\frac{\mathrm{e}^{-\frac{(x_1-\mu_1)^2}{2\sigma_1^2}}}{\sqrt{2\pi\sigma_1^2}}}_{\mathrm{N}(\mu_1,\sigma_1^2)\text{ の p.d.f.}}\mathrm{d}x_1$ が示され，ゆえに $X_1\sim\mathrm{N}(\mu_1,\sigma_1^2)$ である。他の X_i に関してもまったく同様にして $X_i\sim\mathrm{N}(\mu_i,\sigma_i^2)$ であることが示される。
いま，$X_1,X_2,\ldots,X_n$ の独立性は，式 (A.25) より確率ベクトル $(X_1,X_2,\ldots,X_n)$ の p.d.f. が

$$
f(x_1,x_2,\ldots,x_n)=\underbrace{\frac{\exp\left(-\frac{(x_1-\mu_1)^2}{2\sigma_1^2}\right)}{\sqrt{2\pi\sigma_1^2}}}_{X_1\text{ の p.d.f.}}\times\underbrace{\frac{\exp\left(-\frac{(x_2-\mu_2)^2}{2\sigma_2^2}\right)}{\sqrt{2\pi\sigma_2^2}}}_{X_2\text{ の p.d.f.}}\times\cdots\times\underbrace{\frac{\exp\left(-\frac{(x_n-\mu_n)^2}{2\sigma_n^2}\right)}{\sqrt{2\pi\sigma_n^2}}}_{X_n\text{ の p.d.f.}}
$$

となることを用いて命題 5.5 (➔p. 153) から従う。

問 5.31(➔p. 171) の解答例

(1) $G\in\mathcal{G}$ のとき，

$$
\begin{aligned}
\mathbf{E}\big[\mathbf{E}[X\mid\mathcal{G}];G\big] &\overset{\substack{\text{式 (5.1.1)}\\(\text{➔p. 152})}}{=} \mathbf{E}\big[\mathbf{1}_G\mathbf{E}[X\mid\mathcal{G}]\big] \\
&\overset{\substack{\text{公式 5.30–(7)}\\(\text{➔p. 170})}}{=} \mathbf{E}\big[\mathbf{E}[\mathbf{1}_G\cdot X\mid\mathcal{G}]\big] \\
&\overset{\substack{\text{公式 5.30–(2)}\\(\text{➔p. 170})}}{=} \mathbf{E}\big[\mathbf{E}[\mathbf{1}_G\cdot X\mid\mathcal{G}]\mid\{\varnothing,\Omega\}\big] \\
&\overset{\substack{\text{Tower property}\\(\text{➔p. 170})}}{=} \mathbf{E}[\mathbf{1}_G\cdot X\mid\{\varnothing,\Omega\}] \\
&\overset{\substack{\text{公式 5.30–(2)}\\(\text{➔p. 170})}}{=} \mathbf{E}[\mathbf{1}_G\cdot X] \overset{\substack{\text{式 (5.1.1)}\\(\text{➔p. 152})}}{=} \mathbf{E}[X;G]
\end{aligned}
$$

(2) 前問 (1) において，$G=\Omega$ $(\in\mathcal{G})$ とおけば得られる。

問 5.33(➔p. 172) の解答例

$\mathbf{E}[X\mid\sigma(A)]$ は $\sigma(A)=\sigma(\mathbf{1}_A)$-可測であるから，適当な関数 $f(x)$ を用いて $\mathbf{E}[X\mid\sigma(A)]=f(\mathbf{1}_A)$ と表せる。$f(\mathbf{1}_A)=f(0)\cdot\mathbf{1}_A+f(1)\cdot\mathbf{1}_{A^1}$ が成り立つが，この両辺に $\mathbf{1}_A$ を掛けると $f(\mathbf{1}_A)\cdot\mathbf{1}_A=f(0)\cdot\mathbf{1}_A$

となる。この両辺の期待値をとると

$$\mathbf{E}[f(\mathbf{1}_A)\cdot\mathbf{1}_A]=\mathbf{E}[f(0)\cdot\mathbf{1}_A]$$

であり，このそれぞれを計算する。左辺については

$$\begin{aligned}(\text{左辺}) &= \mathbf{E}[f(\mathbf{1}_A)\cdot\mathbf{1}_A]\\ &= \mathbf{E}[\mathbf{E}[X\mid\sigma(A)]\cdot\mathbf{1}_A]\\ &\overset{\substack{\text{式 (5.1.1)}\\(\text{➔p. 152})}}{=} \mathbf{E}[\mathbf{E}[X\mid\sigma(A)];A] \overset{\substack{\text{問 5.31–(1)}\\(\text{➔p. 171})}}{=} \mathbf{E}[X;A]\end{aligned}$$

であり，一方で右辺については，次のように変形できる。

$$(\text{右辺})=\mathbf{E}[f(0)\cdot\mathbf{1}_A]=f(0)\cdot\mathbf{E}[\mathbf{1}_A]\overset{\substack{\text{例 2.41}\\(\text{➔p. 42})}}{=}f(0)\cdot\mathbf{P}(A)$$

よって両辺を比較することで $f(0)=\dfrac{\mathbf{E}[X;A]}{\mathbf{P}(A)}=\mathbf{E}[X\mid A]$ が得られる。同様にすると $f(1)=\mathbf{E}[X\mid A^c]$ であり，ゆえに

$$\mathbf{E}[X\mid\sigma(A)]=\mathbf{E}[X\mid A]\cdot\mathbf{1}_A+\mathbf{E}[X\mid A^c]\cdot\mathbf{1}_{A^c}$$

を得る。

問 5.36(➔p. 176) の解答例

(1) 条件 $X=x$ のもとで考えると，$\mathbf{E}[Yh(X)\mid X=x]=\mathbf{E}[Yh(x)\mid X=x]=h(x)\mathbf{E}[Y\mid X=x]$。

(2) $\mathbf{E}[|g(X)-\mathbf{E}[Y\mid X]|]=0$ であることを示せばよいであろう。$\mathbf{E}[Y\mid X]$ は $\sigma(X)$-可測であるから，適当な関数 $h(x)$ を取れば $\mathbf{E}[Y\mid X]=h(X)$ とかけるはずである。いま，$G=\{g(X)-h(X)\geq 0\}$ とおくと，これは X のとる値がわかれば起きたか起きなかったかがわかるから，$G\in\sigma(X)$ が成り立つ。したがって $G^c\in\sigma(X)$ もまた成り立っている。これを用いて，次のように変形しておく。

$$\begin{aligned}&\mathbf{E}[|g(X)-\mathbf{E}[Y\mid X]|]=\mathbf{E}[|g(X)-h(X)|]\\ &=\mathbf{E}[|g(X)-h(X)|;G]+\mathbf{E}[|g(X)-h(X)|;G^c]\end{aligned}$$

ここで事象 G が起きたときは $|g(X)-h(X)|=g(X)-h(X)$，起きなければ $|g(X)-h(X)|=h(X)-g(X)$ であるから，上の式はさらに次のように変形できる。

$$\mathbf{E}[|g(X)-\mathbf{E}[Y\mid X]|]=\mathbf{E}[g(X)-h(X);G]+\mathbf{E}[h(X)-g(X)|;G^c]$$

右辺の第一項目をみてみよう。これは次のように分解できる。

$$\mathbf{E}[g(X)-h(X);G]=\mathbf{E}[g(X);G]-\mathbf{E}[h(X);G]$$

このうち，$\mathbf{E}[g(X);G]$ は次のように計算できる。

$$\begin{aligned}\mathbf{E}[g(X);G]&=\mathbf{E}[g(X);g(X)-h(X)\geqq 0]\\ &=\mathbf{E}[g(X)\cdot\mathbf{1}_{[0,\infty)}(g(X)-h(X))]\\ &=\int_{\mathbb{R}}g(x)\cdot\mathbf{1}_{[0,\infty)}(g(x)-h(x))\mathbf{P}(X\in\mathrm{d}x)\end{aligned}$$

一方で，$\mathbf{E}[h(X);G]$ は

$$
\begin{aligned}
\mathbf{E}[h(X);G] &= \mathbf{E}[\mathbf{E}[Y \mid X];G] \\
&\overset{\substack{\text{公式 5.30–(7)}\\(\text{➲ p. 170})}}{=} \mathbf{E}\big[\mathbf{E}\big[Y\cdot\mathbf{1}_G \mid X\big]\big] \\
&\overset{\substack{\text{問 5.31–(2)}\\(\text{➲ p. 171})}}{=} \mathbf{E}[Y\cdot\mathbf{1}_G] \\
&\overset{\substack{\text{公式 5.35–(2)}\\(\text{➲ p. 176})}}{=} \int_{\mathbb{R}} \mathbf{E}[Y\cdot\mathbf{1}_G \mid X=x]\mathbf{P}(X\in \mathrm{d}x) \\
&= \int_{\mathbb{R}} \mathbf{E}[Y\cdot\mathbf{1}_{[0,\infty)}(g(X)-h(X)) \mid X=x]\mathbf{P}(X\in \mathrm{d}x) \\
&\overset{(1)}{=} \int_{\mathbb{R}} \mathbf{1}_{[0,\infty)}(g(x)-h(x))\cdot\underbrace{\mathbf{E}[Y \mid X=x]}_{=g(x)}\mathbf{P}(X\in \mathrm{d}x) \\
&= \int_{\mathbb{R}} g(x)\cdot\mathbf{1}_{[0,\infty)}(g(x)-h(x))\mathbf{P}(X\in \mathrm{d}x) = \mathbf{E}[g(X);G]
\end{aligned}
$$

となるため，第一項目は消える。同様にして第二項目についても $\mathbf{E}[h(X)-g(X);G^c]=0$ であることが確かめられる。
以上より $\mathbf{E}[|g(X)-\mathbf{E}[Y \mid X]|]=0$ が成り立つ。

問 6.6(➲ p. 181) の解答例

10 時 30 分から 11 時の間の注文件数 X，11 時 30 分から 12 時の間の注文件数 Y はいずれもパラメータ $10\times\dfrac{1}{2}(\text{時間})=5$ の Poisson 分布に従うことに注意しよう。(1) $\mathbf{P}(X=0)=\mathbf{e^{-5}}$ (2) $\mathbf{P}(X=3)=\mathrm{e}^{-5}\times\dfrac{5^3}{3!}=\mathbf{20\,e^{-5}}$ (3) $\mathbf{P}(Y=7)=\mathrm{e}^{-5}\times\dfrac{5^7}{7!}=\mathbf{\dfrac{15625}{1008}e^{-5}}$

問 6.7(➲ p. 181) の解答例

(1) $T_1\sim \mathrm{E}(\texttt{rate}\ 2)$ (➲ 定義 5.3，p. 163) より，$\mathbf{P}(T_1>0.5)=\displaystyle\int_{0.5}^{\infty}2\mathrm{e}^{-2t}\,\mathrm{d}t=\Big[-\mathrm{e}^{-2t}\Big]_{0.5}^{\infty}=\mathbf{e^{-1}}$
(2) $\mathbf{P}(N_1=0)=\mathrm{e}^{-2}$ と $\mathbf{P}(T_1>3)=\displaystyle\int_3^{\infty}2\mathrm{e}^{-2t}\,dt=\mathrm{e}^{-6}$ より，$\mathbf{P}(T_1>3 \mid N_1=0)=\dfrac{\mathbf{P}(T_1>3)}{\mathbf{P}(N_1=0)}=\dfrac{\mathrm{e}^{-6}}{\mathrm{e}^{-2}}=\mathbf{e^{-4}}$
(3) Poisson 過程の独立定常増分性より，$\mathbf{P}(T_4>4 \mid T_3=2)=\mathbf{P}(T_4-T_3>2 \mid T_3=2)=\mathbf{e^{-4}}$

問 6.8(➲ p. 181) の解答例

$m=2$ の場合を証明する。$(N_t^{(1)})_{t\geqq 0}$，$(N_t^{(2)})_{t\geqq 0}$ をそれぞれ強度が λ_1,λ_2 の独立な Poisson 過程であるとする。$N_t=N_t^{(1)}+N_t^{(2)}$ とおくとき，

(1) $N_0=N_0^{(1)}+N_0^{(2)}=0$ である。

(2) $0 \leqq t_1 < t_2 < \cdots < t_k$ とするとき，

$$N_{t_i} - N_{t_{i-1}} = (N_{t_i}^{(1)} - N_{t_{i-1}}^{(1)}) + (N_{t_i}^{(2)} - N_{t_{i-1}}^{(2)})$$

となるが，仮定より $(N_t^{(1)})_{t \geqq 0}$，$(N_t^{(2)})_{t \geqq 0}$ はそれぞれ独立な Poisson 過程であるから，$\{N_{t_i}^{(1)} - N_{t_{i-1}}^{(1)}\}_{i=1}^k$，$\{N_{t_i}^{(2)} - N_{t_{i-1}}^{(2)}\}_{i=1}^k$ は独立となっている。ゆえに，$\{N_{t_i} - N_{t_{i-1}}\}_{i=1}^k$ もまた独立であるので，$(N_t)_{t \geqq 0}$ は独立増分性をもつ。

(3) $0 \leqq s < t$ とするとき，

$$N_t - N_s = (N_t^{(1)} - N_s^{(1)}) + (N_t^{(2)} - N_s^{(2)})$$

である。ここで，仮定より $N_t^{(1)} - N_s^{(1)} \sim \mathrm{Poisson}(\lambda_1(t-s))$，$N_t^{(2)} - N_s^{(2)} \sim \mathrm{Poisson}(\lambda_2(t-s))$ である。また $N_t^{(1)} - N_s^{(1)}$ と $N_t^{(2)} - N_s^{(2)}$ は独立であるので，これらの和である $N_t - N_s$ は Poisson 分布の再生性より $\mathrm{Poisson}((\lambda_1 + \lambda_2)(t-s))$ に従う。

以上 (1)，(2)，(3) より，$N_t = N_t^{(1)} + N_t^{(2)}$ で定義される確率過程 $(N_t)_{t \geqq 0}$ は強度 $\lambda_1 + \lambda_2$ の Poisson 過程である。

問 6.9(➔p. 181) の解答例

午前 9 時から午前 11 時までの間に 2 個，午後 2 時から午後 5 時までの間に 3 個の不良品が見つかる事象は独立増分性から独立となるので，求める確率は $\left(\mathrm{e}^{-4} \times \frac{4^2}{2!}\right) \times \left(\mathrm{e}^{-6} \times \frac{6^3}{3!}\right) = \mathbf{288\,e^{-10}}$

問 6.16(➔p. 186) の解答例

$Z_0 = 0$ となること，および独立増分性は明らかである。$0 \leqq s < t$ に対して，$N_s \leqq N_t$ であるので，

$$Z_t - Z_s = X_{N_s+1} + X_{N_s+2} + \cdots + X_{N_t}$$

となるが，右辺は独立な二項分布 Bernoulli(p) に従う確率変数の和であるから，$k = 0, 1, 2, \ldots$ について以下の計算が成り立つ。

$$\begin{aligned}
\mathbf{P}(Z_t - Z_s = k) &= \mathbf{P}(X_{N_s+1} + X_{N_s+2} + \cdots + X_{N_t} = k) \\
&= \sum_{\ell=k}^{\infty} \mathbf{P}(X_1 + X_2 + \cdots + X_\ell = k \mid N_t - N_s = \ell)\,\mathbf{P}(N_t - N_s = \ell) \\
&= \sum_{\ell=k}^{\infty} \mathbf{P}(X_1 + X_2 + \cdots + X_\ell = k)\,\mathbf{P}(N_t - N_s = \ell) \\
&= \sum_{\ell=k}^{\infty} \left({}_\ell\mathrm{C}_k p^k (1-p)^{\ell-k}\right) \times \left(\mathrm{e}^{-\lambda(t-s)} \frac{(\lambda(t-s))^\ell}{\ell!}\right) \\
&= \mathrm{e}^{-\lambda(t-s)} \frac{(\lambda p(t-s))^k}{k!} \sum_{\ell=k}^{\infty} \left(\frac{(\lambda(1-p)(t-s))^{\ell-k}}{(\ell-k)!}\right) \\
&= \mathrm{e}^{-\lambda(t-s)} \frac{(\lambda p(t-s))^k}{k!} \times \mathrm{e}^{\lambda(1-p)(t-s))} = \mathrm{e}^{-\lambda p(t-s)} \frac{(\lambda p(t-s))^k}{k!}
\end{aligned}$$

以上より，複合 Poisson 過程 $(Z_t)_{t \geqq 0}$ は強度 λp の Poisson 過程である。

問 6.17(➔ p. 186) の解答例

$\{X_n\}_{n=1}^{\infty}$ は独立同分布で $X_1 \sim \mathrm{N}(0,1)$ であるので，$\mathbf{E}[X_1]=0$，$\mathrm{Var}(X_1)=1$ である。よって複合 Poisson 過程の性質 (➔ 要点 6.15，p. 186) より

$$\mathbf{E}[Z_t] = \lambda t \times 0 = \mathbf{0}, \qquad \mathrm{Var}(Z_t) = \lambda t \times 1 + \lambda t \times 0^2 = \boldsymbol{\lambda t}$$

問 6.18(➔ p. 186) の解答例

生命保険会社にくる支払い請求の件数が従う Poisson 過程 $(N_t)_{t \geqq 0}$ の強度は $12 \times 0.8 = 9.6$ であることに注意する。X_n，$n = 1, 2, \ldots$，を生命保険会社にきた n 番目の支払い請求額とすると，題意よりこれらは独立同分布で，$\mathbf{E}[X_1] = 1000$，$\mathrm{Var}(X_1) = 400^2$ である。この生命保険会社の 1 週間の現金支払額は複合 Poisson 分布 $Z_1 = X_1 + X_2 + \cdots + X_{N_1}$ として表されるので，その平均と分散は次のとおり。

$$\mathbf{E}[Z_1] = 9.6 \times 1000 = \mathbf{9600}(\text{万円}), \quad \mathrm{Var}(Z_1) = 9.6 \times 400^2 + 9.6 \times 1000^2 = \mathbf{11136000}(\text{万円})^2$$

問 7.7(➔ p. 192) の解答例

$Y_1, Y_2, \ldots, Y_n$ は無作為標本，特に同分布列であるから，$\mathbf{E}[Y_1 \mid \theta] = \mathbf{E}[Y_2 \mid \theta] = \cdots = \mathbf{E}[Y_n \mid \theta]$，$\mathrm{Var}(Y_1 \mid \theta) = \mathrm{Var}(Y_2 \mid \theta) = \cdots = \mathrm{Var}(Y_n \mid \theta)$ が成り立つことに注意せよ。

(1) $\mathbf{E}[\overline{Y}_n \mid \theta] = \mathbf{E}\big[\frac{1}{n}\sum_{i=1}^{n} Y_i \mid \theta\big] = \frac{1}{n}\sum_{i=1}^{n}\mathbf{E}[Y_i \mid \theta] = \frac{1}{n}\sum_{i=1}^{n}\theta = \theta$

(2) 各 $i = 1, 2, \ldots, n$ に対して $\mathbf{E}[Y_i - \overline{Y}_n \mid \theta] = \mathbf{E}[Y_i \mid \theta] - \mathbf{E}[\overline{Y}_n \mid \theta] = 0$ である。ゆえに $\mathbf{E}[(Y_i - \overline{Y}_n)^2 \mid \theta] = \mathrm{Var}(\overline{Y}_n - Y_i \mid \theta)$ が成り立つ。さらに，

$$\begin{aligned}
\overline{Y}_n - Y_i &= \frac{Y_1 + Y_2 + \cdots + Y_n}{n} - Y_i \\
&= \frac{Y_1}{n} + \cdots + \left(1 - \frac{1}{n}\right)Y_i + \cdots + \frac{Y_n}{n}
\end{aligned}$$

であり，$Y_1, Y_2, \ldots, Y_n$ が独立であることから

$$\begin{aligned}
\mathbf{E}[U_n \mid \theta] &= \mathbf{E}\left[\frac{1}{n-1}\sum_{i=1}^{n}(Y_i - \overline{Y}_n)^2 \mid \theta\right] = \frac{1}{n-1}\sum_{i=1}^{n}\underbrace{\mathbf{E}[(Y_i - \overline{Y}_n)^2 \mid \theta]}_{\substack{\| \\ \mathrm{Var}(\overline{Y}_n - Y_i \mid \theta)}} \\
&= \frac{1}{n-1}\sum_{i=1}^{n}\mathrm{Var}\left(\frac{Y_1}{n} + \cdots + \left(1 - \frac{1}{n}\right)Y_i + \cdots + \frac{Y_n}{n} \mid \theta\right) \\
&\overset{\substack{\text{命題 5.5–(4)} \\ (➔\ \text{p. 153})}}{=} \frac{1}{n-1}\sum_{i=1}^{n}\left(\frac{\mathrm{Var}(Y_1 \mid \theta)}{n^2} + \cdots + \left(1 - \frac{1}{n}\right)^2 \mathrm{Var}(Y_i \mid \theta) + \cdots + \frac{\mathrm{Var}(Y_n \mid \theta)}{n^2}\right) \\
&= \frac{\theta}{n-1}\sum_{i=1}^{n}\underbrace{\left(\frac{1}{n^2} + \cdots + \left(1 - \frac{1}{n}\right)^2 + \cdots + \frac{1}{n^2}\right)}_{\substack{\frac{1}{n^2}\text{ が }(n-1)\text{ 個と} \\ (1-\frac{1}{n})^2 = \frac{(n-1)^2}{n^2}\text{ が 1 個}}} = \theta
\end{aligned}$$

問 7.8(➔p. 193) の解答例

$T=\tau(Y_1,Y_2,\ldots,Y_n)$ を不偏推定量とする。n 変数関数 $\widetilde{\tau}$ を次で定義する。

$$\widetilde{\tau}(y_1,y_2,\ldots,y_n)=\frac{1}{n!}\sum_{\sigma}\tau(y_{\sigma(1)},y_{\sigma(2)},\ldots,y_{\sigma(n)})$$

により定める。ただし，σ は列 $(1,2,\ldots,n)$ を並べ替える順列 $(\sigma(1),\sigma(2),\ldots,\sigma(n))$ を表しており，和はこのすべての順列 σ に渡ってとる (このような順列の総数は $n!$ 個ある)。

各順列 σ に従って $Y_1,Y_2,\ldots,Y_n$ を並べ替えた $Y_{\sigma(1)},Y_{\sigma(2)},\ldots,Y_{\sigma(n)}$ はやはり独立同分布列であるから，$\mathbf{E}[\tau(Y_{\sigma(1)},Y_{\sigma(2)},\ldots,Y_{\sigma(n)})\mid\theta]=\mathbf{E}[\tau(Y_1,Y_2,\ldots,Y_n)\mid\theta]=\mathbf{E}[T\mid\theta]=\theta$ が成り立ち，したがって $T_\sigma=\tau(Y_{\sigma(1)},Y_{\sigma(2)},\ldots,Y_{\sigma(n)})$ もまた不偏推定量である。ゆえに $\widetilde{T}=\widetilde{\tau}(Y_1,Y_2,\ldots,Y_n)$ は

$$\mathbf{E}[\widetilde{T}\mid\theta]=\frac{1}{n!}\sum_{\sigma}\underbrace{\mathbf{E}[T_\sigma\mid\theta]}_{=\theta}=\theta\cdot\underbrace{\frac{1}{n!}\sum_{\sigma}1}_{=1}=1$$

をみたし，ゆえに不偏推定量である。

分散の方は，

$$\begin{aligned}\mathrm{Var}(\widetilde{T}\mid\theta)&=\mathbf{E}[(\widetilde{T}-\theta)^2\mid\theta]=\mathbf{E}[\left(\frac{1}{n!}\sum_{\sigma}(T_\sigma-\theta)\right)^2\mid\theta]\\&\overset{\substack{\text{Cauchy-Schwarz の}\\\text{不等式 (➔p. 105)}}}{\leqq}\mathbf{E}\Big[\frac{1}{n!}\sum_{\sigma}(T_\sigma-\theta)^2\mid\theta\Big]=\frac{1}{n!}\sum_{\sigma}\underbrace{\mathbf{E}[(T_\sigma-\theta)^2\mid\theta]}_{\substack{\parallel\\\mathbf{E}[(T-\theta)^2\mid\theta]\\\parallel\\\mathrm{Var}(T\mid\theta)}}=\mathrm{Var}(T\mid\theta)\end{aligned}$$

問 7.11(➔p. 194) の解答例

式 $\displaystyle\int\exp(a(\theta)t(y)-\psi(\theta)+b(y))\,\mathrm{d}y=\int f(y\mid\theta)\mathrm{d}y=1$ を両辺 θ で微分した次の式を整理して題意が導かれる。

$$(1\text{ 階微分}):\ \underbrace{\int\big(a'(\theta)t(y)-\psi'(\theta)\big)f(y\mid\theta)\mathrm{d}y}_{\substack{\parallel\\\mathbf{E}[a'(\theta)t(Y)-\psi'(\theta)\mid\theta]}}=0$$

$$(2\text{ 階微分}):\ \underbrace{\int\Big(\big(a''(\theta)t(y)-\psi''(\theta)\big)+\big(a'(\theta)t(y)-\psi'(\theta)\big)^2\Big)f(y\mid\theta)\mathrm{d}y}_{\substack{\parallel\\\mathbf{E}[a''(\theta)t(Y)-\psi''(\theta)+\left(a'(\theta)t(Y)-\psi'(\theta)\right)^2\mid\theta]}}=0$$

問 7.28(➔ p. 205) の解答例

(1) f を平均 μ，分散 σ^2 をもつ $\mathbb{R}$ 上の p.d.f. とすると，

$$\begin{aligned}
0 \leqq D_{\mathrm{KL}}(f \| \mathrm{N}(\mu,\sigma^2)) &= \int_{-\infty}^{\infty} f(y) \log \frac{f(y)}{(2\pi\sigma^2)^{-\frac{1}{2}} \mathrm{e}^{-\frac{(y-\mu)^2}{2\sigma^2}}} \mathrm{d}y \\
&= \int_{-\infty}^{\infty} f(y)\log f(y)\mathrm{d}y - \int_{-\infty}^{\infty} f(y)\Big\{-\frac{(y-\mu)^2}{2\sigma^2} - \log\sqrt{2\pi\sigma^2}\Big\}\mathrm{d}y \\
&= \underbrace{\int_{-\infty}^{\infty} f(y)\log f(y)\mathrm{d}y}_{=-\mathrm{Ent}(f)} + \underbrace{\int_{-\infty}^{\infty}\Big\{\frac{(y-\mu)^2}{2\sigma^2}\log\sqrt{2\pi\sigma^2}\Big\}f(y)\mathrm{d}y}_{\substack{\| \\ \frac{\sigma^2}{2\sigma^2} + \log\sqrt{2\pi\sigma^2} \\ \|\ \text{練習問題 7.27–(2)} \\ \mathrm{Ent}(\mathrm{N}(\mu,\sigma^2))}} = -\mathrm{Ent}(f) + \mathrm{Ent}(\mathrm{N}(\mu,\sigma^2))
\end{aligned}$$

となるから，$\mathrm{Ent}(f) \leqq \mathrm{Ent}(\mathrm{N}(\mu,\sigma^2))$ が成り立つ。ゆえに $\mathrm{N}(\mu,\sigma^2)$ は平均 μ，分散 σ^2 をもつ $\mathbb{R}$ 上の p.d.f. の中で最大のエントロピーをもつ。

(2) 前問 (1) と同様である。

(3) 2 点集合 $\{0,1\}$ 上の p.m.f. は必ず，ある $p \in [0,1]$ についての $\mathrm{Bernoulli}(p)$ となるから，$\mathrm{Ent}(\mathrm{Bernoulli}(p))$ を最大にする p が $p = \frac{1}{2}$ で与えられることを示せばよい。まず，

$$\mathrm{Ent}(\mathrm{Bernoulli}(p)) = -p\log p - (1-p)\log(1-p)$$

であり，その微分は $\dfrac{\mathrm{d}}{\mathrm{d}p}\mathrm{Ent}(\mathrm{Bernoulli}(p)) = -\log p - 1 + \log(1-p) + 1 = \log\dfrac{1-p}{p}$ となるから，$\frac{\mathrm{d}}{\mathrm{d}p}\mathrm{Ent}(\mathrm{Bernoulli}(p)) = 0 \Leftrightarrow p = \frac{1}{2}$ である。これは最大点を与える。ゆえに $\mathrm{Bernoulli}(\frac{1}{2})$ が最大のエントロピーを与える。

問 7.30(➔ p. 207) の解答例

$p = q$ のときは両辺ともに 0 となり成立する。
p を固定して q の関数 $f(q)$ を $f(q) = p\log\dfrac{p}{q} + (1-p)\log\dfrac{1-p}{1-q} - 2|p-q|^2$ により定めると，$q \in [0,p)$ のときは

$$f'(q) = -\frac{p}{q} + \frac{1-p}{1-q} + 4(p-q) = (q-p)\Big\{\frac{1}{q(1-q)} - 4\Big\} = \underbrace{(q-p)}_{\leqq 0}\Big\{\underbrace{\frac{1}{\frac{1}{4} - (q-\frac{1}{2})^2} - 4}_{\geqq \frac{1}{4^{-1}} - 4 = 0}\Big\} \leqq 0$$

であり，同様に $q \in [p,1)$ のときは $f'(q) \geqq 0$ である。$f(p) = 0$ であるから $f(q) \geqq 0$，つまり題意が示された。

問 7.31(➔ p. 208) の解答例

(1) $x < y$ なる正数 x，y と $t \in (0,1)$ を任意とし，$z = tx + (1-t)y$ とおく。このとき $x < y < z$ が成り立つ。
$x > 0$ のとき，$f'(x) = \log x + 1$，$f''(x) = \frac{1}{x} \geqq 0$ であるから，$f'(x)$ は単調増加である。特に，関数 f の区間 $[x,z]$ での変化の割合よりも，区間 $[z,y]$ での変化の割合の方が大きい。つまり

$\frac{f(z)-f(x)}{z-x} \leqq \frac{f(y)-f(z)}{y-z}$ が成り立つ。分母を払うと $(y-z)(f(z)-f(x)) \leqq (z-x)(f(y)-f(z))$ であり，両辺を $y-x$ で割ると次を得る。

$$\underbrace{\frac{y-z}{y-x}}_{=t}\big(f(z)-f(x)\big) \leqq \underbrace{\frac{z-x}{y-x}}_{=1-t}\big(f(y)-f(z)\big)$$

これを整理すると，$f(z) \leqq tf(x)+(1-t)f(y)$ となる。関数 $f(x)$ は区間 $[0,\infty)$ 上で連続であるから，この不等式において $x \to 0$ とすることで，$x=0$ の場合も成り立つ。

(2) 数学的帰納法により示される。

(3) $s=\sum_{j=1}^n b_j$，$p_i = b_i/s \geqq 0$ とおくと $p_1+p_2+\cdots+p_n=1$ であり，

$$\begin{aligned}\sum_{i=1}^n a_i \log\frac{a_i}{b_i} &= s\sum_{i=1}^n \underbrace{\frac{b_i}{s}}_{=p_i}\underbrace{\frac{a_i}{b_i}\log\frac{a_i}{b_i}}_{=f\left(\frac{a_i}{b_i}\right)}\\ &= s\sum_{i=1}^n p_i f\left(\frac{a_i}{b_i}\right) \overset{(2)}{\geqq} sf\left(\sum_{i=1}^n p_i\frac{a_i}{b_i}\right) = \underbrace{s\left(\sum_{i=1}^n p_i\frac{a_i}{b_i}\right)}_{=\left(\sum_{i=1}^n a_i\right)}\log\Bigg(\underbrace{\sum_{i=1}^n p_i\frac{a_i}{b_i}}_{=\frac{\sum_{i=1}^n a_i}{\sum_{i=1}^n b_i}}\Bigg)\end{aligned}$$

問 7.32(➔p. 208) の解答例

(1) $x>0$ ならば $\log x < x-1$ であるから，次が成り立つ。

$$\begin{aligned}\sum_{i\in J} p_i\log\frac{p_i}{q_i} &= -\sum_{i\in J} p_i\log\frac{q_i}{p_i}\\ &\geqq -\sum_{i\in J} p_i\left(\frac{q_i}{p_i}-1\right) = \sum_{i\in J}(p_i-q_i)\end{aligned}$$

(2) 前問 (1) において $J=I$ と選ぶと $D_{\mathrm{KL}}(p\|q) = \sum_{i\in I} p_i\log\frac{p_i}{q_i} \geqq \sum_{i\in I}(p_i-q_i) = \sum_{i\in I}p_i - \sum_{i\in I}q_i = 1-1=0$ が成り立つ。

(3) $p=\sum_{i\in J}p_i$，$q=\sum_{i\in J}q_i$ とおくと，

$$\begin{aligned}D_{\mathrm{KL}}(p\|q) &= \sum_{i\in J}p_i\log\frac{p_i}{q_i} + \sum_{i\in I\setminus J}p_i\log\frac{p_i}{q_i}\\ &\overset{\substack{\text{対数和不等式}\\ \text{問 7.31–(3)}\\ (➔\text{p. 208})}}{\geqq} \underbrace{\left(\sum_{i\in J}p_i\right)\log\frac{\sum_{i\in J}p_i}{\sum_{i\in J}q_i}}_{=p\log\frac{p}{q}} + \underbrace{\left(\sum_{i\in I\setminus J}p_i\right)\log\frac{\sum_{i\in I\setminus J}p_i}{\sum_{i\in I\setminus J}q_i}}_{=(1-p)\log\frac{1-p}{1-q}}\\ &\overset{\substack{\text{問 7.30}\\ (➔\text{p. 207})}}{\geqq} 2|p-q|^2 = 2\Big|\sum_{i\in J}p_i - \sum_{i\in J}q_i\Big|^2 = 2\Big|\sum_{i\in J}(p_i-q_i)\Big|^2\end{aligned}$$

が得られる。また，次の式変形も確かめられる。

$$\sum_{i\in J}(p_i-q_i)=\sum_{i\in J}p_i-\sum_{i\in J}q_i=\Bigg(1-\sum_{i\in I\setminus J}p_i\Bigg)-\Bigg(1-\sum_{i\in I\setminus J}q_i\Bigg)$$
$$=\sum_{i\in I\setminus J}q_i-\sum_{i\in I\setminus J}p_i=\sum_{i\in I\setminus J}(q_i-p_i)$$

(4) 前問 (3) において $J=\{i\in I : p_i>q_i\}$ と選ぶと

$$D_{\mathrm{KL}}(p\|q)\overset{(3)}{\geqq}2\Big|\sum_{\substack{i\in I:\\ p_i>q_i}}(p_i-q_i)\Big|^2$$
$$\overset{(3)}{=}2\Big|\frac{1}{2}\Big(\sum_{\substack{i\in I:\\ p_i>q_i}}\underbrace{(p_i-q_i)}_{\substack{p_i>q_i\text{ より，}\\ |p_i-q_i|\text{ に等しい}}}+\sum_{\substack{i\in I:\\ p_i\leqq q_i}}\underbrace{(q_i-p_i)}_{\substack{p_i\leqq q_i\text{ より，}\\ |p_i-q_i|\text{ に等しい}}}\Big)\Big|^2=\frac{1}{2}\Big|\sum_{i\in I}|p_i-q_i|\Big|^2$$

ゆえに $\sum_{i\in I}|p_i-q_i|\leqq\sqrt{2D_{\mathrm{KL}}(p\|q)}$ が得られた。

(5) 前問 (4) より明らか。

問 7.33(➔p. 208) の解答例

(1) $x>0$ のとき $\log x<x-1$ であることを用いると，次の評価が得られる。

$$\int_B p(x)\log\frac{p(x)}{q(x)}\mathrm{d}x=-\int_B p(x)\log\frac{q(x)}{p(x)}\mathrm{d}x\geqq-\int_B p(x)\Big(\frac{q(x)}{p(x)}-1\Big)\mathrm{d}x=\int_B\{p(x)-q(x)\}\mathrm{d}x$$

(2) 前問 (1) において $B=\mathbb{R}$ と選ぶと，次の評価が得られる。

$$D_{\mathrm{KL}}(p\|q)=\int_{-\infty}^{\infty}p(x)\log\frac{p(x)}{q(x)}\mathrm{d}x\overset{(1)}{\geqq}\int_{-\infty}^{\infty}\{p(x)-q(x)\}\mathrm{d}x$$
$$=\int_{-\infty}^{\infty}p(x)\mathrm{d}x-\int_{-\infty}^{\infty}q(x)\mathrm{d}x=1-1=0$$

(3) 積分の方から計算すると，次を得る。

$$u^2\int_0^1\frac{1-t}{1+tu}\mathrm{d}t=u\int_0^1\frac{(1+u)-(1+tu)}{1+tu}\mathrm{d}t$$
$$=u\int_0^1\Big(\frac{1+u}{1+tu}-1\Big)\mathrm{d}t$$
$$=u\Big[\frac{1+u}{u}\log(1+tu)-t\Big]_{t=0}^{t=1}$$
$$=u\Big(\frac{1+u}{u}\log(1+u)-1\Big)=(1+u)\log(1+u)-u=h(u)$$

(4) $u(x)=\left(\frac{p(x)}{q(x)}-1\right)\mathbf{1}_{\{p>0\}}(x)\ (\geqq -1)$ とおくと，次の式変形が確かめられる。

$$
\begin{aligned}
D_{\mathrm{KL}}(p\|q) &= \int_{\{p>0\}} p(x)\log\frac{p(x)}{q(x)}\mathrm{d}x \\
&= \int_{\{p>0\}} \Bigg(1+\underbrace{\left(\frac{p(x)}{q(x)}-1\right)}_{=u(x)}\Bigg)\log\Bigg(1+\underbrace{\left(\frac{p(x)}{q(x)}-1\right)}_{=u(x)}\Bigg)q(x)\mathrm{d}x - \underbrace{\int_{-\infty}^{\infty}\left(\frac{p(x)}{q(x)}-1\right)q(x)\mathrm{d}x}_{=0} \\
&= \int_{-\infty}^{\infty}\underbrace{\big\{(1+u(x))\log(1+u(x))-u(x)\big\}}_{=h(u(x))}q(x)\mathrm{d}x \\
&\overset{(3)}{=} \int_{-\infty}^{\infty}\left\{u(x)^2\int_0^1\frac{1-t}{1+tu(x)}\mathrm{d}t\right\}q(x)\mathrm{d}x = \int_0^1\int_{-\infty}^{\infty}\frac{1-t}{1+t\left(\frac{p(x)}{q(x)}-1\right)}\left(\frac{p(x)}{q(x)}-1\right)^2 q(x)\mathrm{d}x\mathrm{d}t
\end{aligned}
$$

(5) 次のように評価すればよい。

$$
\begin{aligned}
&\left(\int_0^1(1-t)\mathrm{d}t\right)^2\left(\int_{-\infty}^{\infty}|u(x)|q(x)\mathrm{d}x\right)^2 = \left(\int_0^1\int_{-\infty}^{\infty}(1-t)|u(x)|q(x)\mathrm{d}x\mathrm{d}t\right)^2 \\
&= \left(\int_0^1\int_{-\infty}^{\infty}(1-t)^{\frac{1}{2}}(1+tu(x))^{\frac{1}{2}}\frac{(1-t)^{\frac{1}{2}}|u(x)|}{(1+tu(x))^{\frac{1}{2}}}q(x)\mathrm{d}x\mathrm{d}t\right)^2 \\
&\underset{\text{Cauchy-Schwarz の不等式 (➔p. 105)}}{\leqq} \left\{\int_0^1\int_{-\infty}^{\infty}(1-t)(1+tu(x))q(x)\mathrm{d}x\mathrm{d}t\right\}\left\{\int_0^1\int_{-\infty}^{\infty}\frac{1-t}{1+tu(x)}(u(x))^2q(x)\mathrm{d}x\mathrm{d}t\right\}
\end{aligned}
$$

(6) 前問 (5) を $u(x)=\frac{p(x)}{q(x)}-1\ (\geqq -1)$ に適用すると，(4) も合わせて

$$
\begin{aligned}
&\Bigg(\overbrace{\int_0^1(1-t)\mathrm{d}t}^{=\frac{1}{2}}\Bigg)^2\Bigg(\overbrace{\int_{-\infty}^{\infty}|u(x)|q(x)\mathrm{d}x}^{=\int_{-\infty}^{\infty}|p(x)-q(x)|\mathrm{d}x}\Bigg)^2 \\
&\leqq \Bigg\{\underbrace{\int_0^1\int_{-\infty}^{\infty}(1-t)(1+tu(x))q(x)\mathrm{d}x\mathrm{d}t}_{\| \atop \int_0^1\int_{-\infty}^{\infty}(1-t)\left(tp(x)+(1-t)q(x)\right)\mathrm{d}x\mathrm{d}t=\frac{1}{2}}\Bigg\}\underbrace{\left\{\int_0^1\int_{-\infty}^{\infty}\frac{1-t}{1+tu(x)}(u(x))^2q(x)\mathrm{d}x\mathrm{d}t\right\}}_{=D_{\mathrm{KL}}(p\|q)}
\end{aligned}
$$

したがって $D_{\mathrm{KL}}(p\|q)\geqq\int_{-\infty}^{\infty}|p(x)-q(x)|\mathrm{d}x$ が成り立つ。

(7) KL divergence の定義と前問 (6) より従う。

問 7.34(➔p. 209) の解答例

Taylor の定理より，次が成り立つ。

$$
\begin{aligned}
D_{\mathrm{KL}}(f_{|\theta}\|f_{|\varphi}) &= \underbrace{D_{\mathrm{KL}}(f_{|\theta}\|f_{|\theta})}_{=0} + \left(\left.\frac{\mathrm{d}}{\mathrm{d}\varphi}\right|_{\varphi=\theta}D_{\mathrm{KL}}(f_{|\theta}\|f_{|\varphi})\right)(\varphi-\theta) \\
&\quad + \frac{1}{2}\left(\left.\frac{\mathrm{d}^2}{\mathrm{d}\varphi^2}\right|_{\varphi=\theta}D_{\mathrm{KL}}(f_{|\theta}\|f_{|\varphi})\right)(\varphi-\theta)^2 + O\big((\varphi-\theta)^2\big)
\end{aligned}
$$

ここに現れた各微分係数を計算する。まず，

$$\frac{\mathrm{d}}{\mathrm{d}\varphi}D_{\mathrm{KL}}(f_{|\theta}\|f_{|\varphi}) = \frac{\mathrm{d}}{\mathrm{d}\varphi}\int f(y\mid\theta)\log\frac{f(y\mid\varphi)}{f(y\mid\varphi)}\mathrm{d}y = -\int f(y\mid\theta)\frac{\partial_\varphi f(y\mid\varphi)}{f(y\mid\varphi)}\mathrm{d}y$$

であるから，点 $\varphi=\theta$ においては，

$$\left.\frac{\mathrm{d}}{\mathrm{d}\varphi}\right|_{\varphi=\theta} D_{\mathrm{KL}}(f_{|\theta}\|f_{|\varphi}) = -\int \partial_\theta f(y\mid\theta)\mathrm{d}y = -\partial_\theta\int f(y\mid\theta)\mathrm{d}y = -\partial_\theta 1 = 0$$

となる。また，

$$\frac{\mathrm{d}^2}{\mathrm{d}\varphi^2}D_{\mathrm{KL}}(f_{|\theta}\|f_{|\varphi}) = -\int f(y\mid\theta)\partial_\varphi\frac{\partial_\varphi f(y\mid\varphi)}{f(y\mid\varphi)}\mathrm{d}y = -\int f(y\mid\theta)\partial_\varphi^2\log f(y\mid\varphi)\mathrm{d}y$$

であるから，点 $\varphi=\theta$ においては次のようになる。

$$\begin{aligned}\left.\frac{\mathrm{d}^2}{\mathrm{d}\varphi^2}\right|_{\varphi=\theta} D_{\mathrm{KL}}(f_{|\theta}\|f_{|\varphi}) &= -\int f(y\mid\theta)\partial_\theta^2\log f(y\mid\theta)\mathrm{d}y \\ &= \int \big(\partial_\theta f(y\mid\theta)\big)\big(\partial_\theta\log f(y\mid\theta)\big)\mathrm{d}y \\ &= \int \big(\partial_\theta\log f(y\mid\theta)\big)^2 f(y\mid\theta)\mathrm{d}y \overset{\substack{\text{定義 7.8}\\(\text{➔p. 196})}}{=} I(\theta)\end{aligned}$$

以上から $D_{\mathrm{KL}}(f_{|\theta}\|f_{|\varphi}) = \dfrac{1}{2}I(\theta)(\varphi-\theta)^2 + O\big((\varphi-\theta)^3\big)$ が従う。

問 7.35(➔p. 209) の解答例

大数の法則より，次が成り立つ。

$$\frac{1}{n}\log f_n(\mathbf{Y}\mid\theta) \overset{n\to\infty}{\longrightarrow} \mathbf{E}[\log f(Y\mid\theta)\mid\theta] = \int f(y\mid\theta)\log f(y\mid\theta)\mathrm{d}y = -\mathrm{Ent}(f_{|\theta})$$

問 7.36(➔p. 209) の解答例

(1) $1\leqq k\leqq d$ ならば，$\mathbf{E}[Y^k\mid\theta]$ は次のように計算できる。

$$\begin{aligned}\mathbf{E}[Y^k\mid\theta] &= \int_{-\infty}^{\infty} y^k f(y\mid\theta)\mathrm{d}y = \int_{-\infty}^{\infty}\underbrace{Z(\theta)^{-1}\frac{\partial}{\partial\theta_k}\exp\left(\sum_{l=1}^{d}\theta_l y^l\right)}_{\substack{\| \\ \frac{\partial}{\partial\theta_k}f(y\mid\theta)+\frac{\partial\log Z(\theta)}{\partial\theta_k}f(y\mid\theta)}}\mathrm{d}y \\ &= \frac{\partial}{\partial\theta_k}\underbrace{\int_{-\infty}^{\infty} f(y\mid\theta)\mathrm{d}y}_{=1} + \frac{\partial\log Z(\theta)}{\partial\theta_k}\underbrace{\int_{-\infty}^{\infty} f(y\mid\theta)\mathrm{d}y}_{=1} = \frac{\partial\log Z(\theta)}{\partial\theta_k}\end{aligned}$$

(2) $\dfrac{\partial}{\partial\theta_k}f(y\mid\theta) = y^k f(y\mid\theta) + f(y\mid\theta)Z(\theta)\dfrac{\partial}{\partial\theta_k}Z(\theta)^{-1} = y^k f(y\mid\theta) - f(y\mid\theta)\dfrac{\partial}{\partial\theta_k}\log Z(\theta)$ であ

るから $\dfrac{\partial}{\partial\theta_k}\ell(\theta;y)=y^k-\dfrac{\partial}{\partial\theta_k}\log Z(\theta)$ が成り立つ。これを用いて，$I(\theta)_{k,l}$ は次のように計算できる。

$$
\begin{aligned}
I(\theta)_{k,l} &= \mathbf{E}\left[\left(\frac{\partial}{\partial\theta_k}\ell(\theta;Y)\right)\left(\frac{\partial}{\partial\theta_l}\ell(\theta;Y)\right)\mid\theta\right]\\
&= \mathbf{E}\left[\left(Y^k-\frac{\partial}{\partial\theta_k}\log Z(\theta)\right)\left(Y^l-\frac{\partial}{\partial\theta_l}\log Z(\theta)\right)\mid\theta\right]\\
&= \mathbf{E}[Y^{k+l}\mid\theta]-\mathbf{E}[Y^k\mid\theta]\frac{\partial}{\partial\theta_l}\log Z(\theta)-\mathbf{E}[Y^l\mid\theta]\frac{\partial}{\partial\theta_k}\log Z(\theta)+\left(\frac{\partial}{\partial\theta_k}\log Z(\theta)\right)\left(\frac{\partial}{\partial\theta_l}\log Z(\theta)\right)\\
&\overset{(1)}{=} \underbrace{\mathbf{E}[Y^{k+l}\mid\theta]}_{\substack{k+l\leqq d\text{ なら}\\ \frac{\partial}{\partial\theta_{k+l}}\log Z(\theta)\\ \text{と表せる。}}}-\left(\frac{\partial}{\partial\theta_k}\log Z(\theta)\right)\left(\frac{\partial}{\partial\theta_l}\log Z(\theta)\right)
\end{aligned}
$$

(3) まず，汎化損失 $G(f_{|\theta_0}\|f_{|\theta})$ を次のように計算しておく。

$$
\begin{aligned}
G(f_{|\theta_0}\|f_{|\theta}) &= -\int_{-\infty}^{\infty} f(y\mid\theta_0)\underbrace{\log f(y\mid\theta)}_{=\sum_{k=1}^{d}\theta_k y^k-\log Z(\theta)}\mathrm{d}y\\
&= \log Z(\theta)\underbrace{\int_{-\infty}^{\infty} f(y\mid\theta_0)\mathrm{d}y}_{=1}-\sum_{k=1}^{d}\theta_k\underbrace{\int_{-\infty}^{\infty} y^k f(y\mid\theta_0)\mathrm{d}y}_{=\left.\frac{\partial}{\partial\theta_k}\right|_{\theta=\theta_0}\log Z(\theta)} = \log Z(\theta)-\sum_{k=1}^{d}\theta_k\left.\frac{\partial}{\partial\theta_k}\right|_{\theta=\theta_0}\log Z(\theta)
\end{aligned}
$$

特に $\mathrm{Ent}(f_{|\theta_0})=\log Z(\theta_0)-\sum_{k=1}^{d}(\theta_0)_k\left.\dfrac{\partial}{\partial\theta_k}\right|_{\theta=\theta_0}\log Z(\theta)$ である。これを用いると，次式を得る。

$$
D_{\mathrm{KL}}(f_{|\theta_0}\|f_{|\theta})=\log\frac{Z(\theta)}{Z(\theta_0)}-\sum_{k=1}^{d}(\theta_k-(\theta_0)_k)\left.\frac{\partial}{\partial\theta_k}\right|_{\theta=\theta_0}\log Z(\theta)
$$

(4) $\log Z(\theta)$ の 2 階微分は次のように計算できる。

$$
\begin{aligned}
\frac{\partial^2}{\partial\theta_k\partial\theta_j}\log Z(\theta) &\overset{(1)}{=} \frac{\partial}{\partial\theta_j}\int_{-\infty}^{\infty} y^k Z(\theta)^{-1}\exp\left(\sum_{l=1}^{d}\theta_l y^l\right)\mathrm{d}y\\
&= -\int_{-\infty}^{\infty} y^k\underbrace{\frac{\frac{\partial}{\partial\theta_j}Z(\theta)}{Z(\theta)^2}\exp\left(\sum_{l=1}^{d}\theta_l y^l\right)}_{=\left(\frac{\partial}{\partial\theta_j}\log Z(\theta)\right)f(y\mid\theta)}\mathrm{d}y+\underbrace{\int_{-\infty}^{\infty} y^{j+k}Z(\theta)^{-1}\exp\left(\sum_{l=1}^{d}\theta_l y^l\right)\mathrm{d}y}_{=\mathbf{E}[Y^{j+k}\mid\theta]}\\
&= -\left(\frac{\partial}{\partial\theta_j}\log Z(\theta)\right)\underbrace{\int_{-\infty}^{\infty} y^k f(y\mid\theta)\mathrm{d}y}_{=\frac{\partial}{\partial\theta_k}\log Z(\theta)}+\mathbf{E}[Y^{j+k}\mid\theta]\overset{(2)}{=} I(\theta)_{j,k}
\end{aligned}
$$

(5) $\nabla \log \tau(\theta) = \nabla \log Z(\theta)$ であることを示せばよい。実際に計算してみると，次のようになる。

$$\frac{\partial}{\partial\theta_k}\log\tau(\theta) = \frac{\partial}{\partial\theta_k}\int_0^1 \sum_{l=1}^d \big(\theta_l - (\theta_0)_l\big)\ \underbrace{\mathbf{E}[Y^l \mid \theta_0 + t(\theta-\theta_0)]}_{=\left.\frac{\partial}{\partial\theta_l}\right|_{\theta_0+t(\theta-\theta_0)}\log Z(\theta)}\ \mathrm{d}t$$

$$= \sum_{l=1}^d \underbrace{\left(\frac{\partial}{\partial\theta_k}\big(\theta_l - (\theta_0)_l\big)\right)}_{\substack{l\text{ が }1,2,\ldots,d\text{ と動くとき，}\\ l=k\text{ のときのみ }1\text{ で}\\ l\neq k\text{ のときは }0}} \int_0^1 \left.\frac{\partial}{\partial\theta_l}\right|_{\theta_0+t(\theta-\theta_0)}\log Z(\theta)\,\mathrm{d}t + \int_0^1 \sum_{l=1}^d \big(\theta_l - (\theta_0)_l\big)\underbrace{\frac{\partial}{\partial\theta_k}\left.\frac{\partial}{\partial\theta_l}\right|_{\theta_0+t(\theta-\theta_0)}\log Z(\theta)\mathrm{d}t}_{\substack{\| \\ t\left.\frac{\partial}{\partial\theta_l}\right|_{\theta_0+t(\theta-\theta_0)}\frac{\partial\log Z(\theta)}{\partial\theta_k}}}$$

$$= \int_0^1 \left.\frac{\partial\log Z(\theta)}{\partial\theta_k}\right|_{\theta_0+t(\theta-\theta_0)}\mathrm{d}t + \int_0^1 \underbrace{t\sum_{l=1}^d \big(\theta_l - (\theta_0)_l\big)\left.\frac{\partial}{\partial\theta_l}\right|_{\theta_0+t(\theta-\theta_0)}\frac{\partial\log Z(\theta)}{\partial\theta_k}}_{=\frac{\mathrm{d}}{\mathrm{d}t}\left.\frac{\partial}{\partial\theta_k}\right|_{\theta_0+t(\theta-\theta_0)}\log Z(\theta)}\mathrm{d}t$$

$$= \int_0^1 \left.\frac{\partial\log Z(\theta)}{\partial\theta_k}\right|_{\theta_0+t(\theta-\theta_0)}\mathrm{d}t + \int_0^1 t\frac{\mathrm{d}}{\mathrm{d}t}\left.\frac{\partial\log Z(\theta)}{\partial\theta_k}\right|_{\theta_0+t(\theta-\theta_0)}\mathrm{d}t$$

$$\overset{\substack{\text{部分}\\\text{積分}}}{=} \left[t\cdot\left.\frac{\partial\log Z(\theta)}{\partial\theta_k}\right|_{\theta_0+t(\theta-\theta_0)}\right]_{t=0}^{t=1} = \frac{\partial\log Z(\theta)}{\partial\theta_k}$$

(6) (5) より $\log\dfrac{Z(\theta)}{Z(\theta_0)} = \log\tau(\theta)$ であり，ゆえに (3) より，次を得る。

$$D_{\mathrm{KL}}(f_{|\theta_0}\|f_{|\theta}) = \sum_{k=1}^d \big(\theta_k - (\theta_0)_k\big)\left\{\int_0^1 \mathbf{E}[Y^k \mid \theta_0 + t(\theta-\theta_0)]\mathrm{d}t - \mathbf{E}[Y^k \mid \theta_0]\right\}$$

以下，この右辺の第 2 因子目を計算する。

$$\int_0^1 \mathbf{E}[Y^k \mid \theta_0 + t(\theta-\theta_0)]\mathrm{d}t - \mathbf{E}[Y^k \mid \theta_0]$$

$$= \underbrace{\mathbf{E}[Y^k \mid \theta] - \mathbf{E}[Y^k \mid \theta_0]}_{\substack{\| \\ \int_0^1 \frac{\mathrm{d}}{\mathrm{d}t}\mathbf{E}[Y^k \mid \theta_0+t(\theta-\theta_0)]\mathrm{d}t}} - \int_0^1 t\frac{\mathrm{d}}{\mathrm{d}t}\mathbf{E}[Y^k \mid \theta_0 + t(\theta-\theta_0)]\mathrm{d}t$$

$$= \int_0^1 (1-t)\frac{\mathrm{d}}{\mathrm{d}t}\mathbf{E}[Y^k \mid \theta_0 + t(\theta-\theta_0)]\mathrm{d}t$$

$$= \sum_{l=1}^d \big(\theta_k - (\theta_0)_k\big)\int_0^1 (1-t)\left.\frac{\partial}{\partial\theta_l}\right|_{\theta_0+t(\theta-\theta_0)}\mathbf{E}[Y^k \mid \theta]\mathrm{d}t = \sum_{l=1}^d \big(\theta_k - (\theta_0)_k\big)\int_0^1 (1-t)I(\theta_0 + t(\theta-\theta_0))_{k,l}\mathrm{d}t$$

問 7.39(➔ p. 212) の解答例

$\hat{\mu}$ を定数として $(y_1, y_2, \ldots, y_n)$ が $\frac{1}{n}\sum_{i=1}^n y_i = \hat{\mu}$ をみたすように自由に動くとき，$(y_1, y_2, \ldots, y_n)$ の最初の $n-1$ 個 $y_1, y_2, \ldots, y_{n-1}$ を勝手に指定すると，残る y_n の値が決まり，$y_n = n\cdot\hat{\mu} - (y_1 + y_2 + \cdots + y_{n-1})$ となる ($n \geqq 3$ であるから，少なくとも y_1 と y_2 の二つの変数は自由に動くことができるこ

とに注意!)。このとき，式 $\sum_{i=1}^{n} g(y_i - \hat{\mu}) = 0$ を変数 y_1 について偏微分すると

$$
\begin{aligned}
0 &= \frac{\partial}{\partial y_1}\Big\{ g(y_1 - \hat{\mu}) + g(y_2 - \hat{\mu}) + \cdots + g(y_{n-1} - \hat{\mu}) + g\big(\overbrace{(n-1)\cdot\hat{\mu} - (y_1 + y_2 + \cdots + y_{n-1})}^{y_n - \hat{\mu} =}\big)\Big\} \\
&= g'(y_1 - \hat{\mu}) - g'(y_n - \hat{\mu}),
\end{aligned}
$$

つまり $g'(y_1 - \hat{\mu}) = g'(y_n - \hat{\mu})$ が成り立つ。同様にしてやはり $g'(y_2 - \hat{\mu}) = g'(y_n - \hat{\mu})$ が成り立ち，したがって $g'(y_1 - \hat{\mu}) = g'(y_2 - \hat{\mu})$ でなければならない。y_1 と y_2 は任意であるから，これは関数 $y \mapsto g'(y - \hat{\mu})$ が定数関数であることを意味する。ゆえに，ある定数 a について $g'(y - \hat{\mu}) = a$ が成り立つ。これを y に関して積分すると積分定数を b として $g(y - \hat{\mu}) = a \cdot y + b$ と表せるが，条件から

$$0 = \sum_{i=1}^{n} g(y_i - \hat{\mu}) = a \sum_{i=1}^{n} y_i + n \cdot b = an \cdot \hat{\mu} + n \cdot b$$

であり，ゆえに $b = -a \cdot \hat{\mu}$ でなければならない。よって $g(y - \hat{\mu}) = a \cdot y - a \cdot \hat{\mu} = a(y - \hat{\mu})$ となる。

問 8.14(➔p. 243) の解答例

(1) θ が位置パラメータであるとすると，$f(y \mid \theta) = h(y - \theta)$ であるような関数 $h(y)$ が存在する。任意の関数 g に対して，$\partial_y g(y - \theta) = -\partial_\theta g(y - \theta)$ であることに注意せよ。特に，$\ell(\theta; y) = \log f(y \mid \theta)$ についても，ある関数 g を用いて $(\partial_\theta \ell(\theta; y))^2 = g(y - \theta)$ と表せるから，

$$
\begin{aligned}
\partial_\theta I(\theta) &= \partial_\theta \int_{-\infty}^{\infty} (\partial_\theta \ell(\theta; y))^2 h(y - \theta) \mathrm{d}y \\
&= \underbrace{\int_{-\infty}^{\infty} \Big(\underbrace{\partial_\theta (\partial_\theta \ell(\theta; y))^2}_{\substack{\| \\ -\partial_y (\partial_\theta \ell(\theta; y))^2}} \Big) h(y - \theta) \mathrm{d}y}_{\substack{\| \ \text{部分積分} \\ \int_{-\infty}^{\infty} (\partial_\theta \ell(\theta; y))^2 h'(y - \theta) \mathrm{d}y}} - \int_{-\infty}^{\infty} (\partial_\theta \ell(\theta; y))^2 h'(y - \theta) \mathrm{d}y = 0
\end{aligned}
$$

であるから $I(\theta)$ は定数，特に $\pi(\theta) \propto 1$ である。
(2) θ は尺度パラメータであるから，ある関数 $g(y)$ と定数 m を用いて $f(y \mid \theta) = \frac{1}{\theta} g\big(\frac{y-m}{\theta}\big)$ と表せる。$x = \frac{y-m}{\theta}$ とおくとき，

$$
\begin{aligned}
\partial_\theta \ell(\theta; y) &= \frac{\partial_\theta f(y \mid \theta)}{f(y \mid \theta)} \\
&= \frac{-\frac{1}{\theta^2} g\big(\frac{y-m}{\theta}\big) - \frac{1}{\theta^2} \frac{y-m}{\theta} g'\big(\frac{y-m}{\theta}\big)}{\frac{1}{\theta} g\big(\frac{y-m}{\theta}\big)} = \frac{-\frac{1}{\theta^2} g(x) - \frac{1}{\theta^2} x g'(x)}{\frac{1}{\theta} g(x)} = -\frac{1}{\theta}\left(1 + x \frac{g'(x)}{g(x)}\right)
\end{aligned}
$$

$\mathbf{P}(\bullet \mid \theta)$ の下で Y が p.d.f. $f(y \mid \theta)$ をもつとき，$X = \frac{Y-m}{\theta}$ は g を p.d.f. にもつ (**X の分布は θ によらない!**)。このとき，

$$I(\theta) = \int_{-\infty}^{\infty} \big(\partial_\theta \ell(\theta; y)\big)^2 f(y \mid \theta) \mathrm{d}\theta = \mathbf{E}[\big(\partial_\theta \ell(\theta; Y)\big)^2 \mid \theta] = \frac{1}{\theta^2} \underbrace{\mathbf{E}[\left(1 + X \frac{g'(X)}{g(X)}\right)^2 \mid \theta]}_{\theta \text{ によらない定数!}}$$

となり，ゆえに $\pi(\theta) \propto \theta^{-1}$ が成り立つ。

問 8.18(➔p. 255) の解答例

Bayes の公式から

$$
\begin{aligned}
\pi(\theta \mid \mathbf{y}^{\mathrm{obs}}) &\propto \left(\prod_{i=1}^{n} \theta^{y_i^{\mathrm{obs}}}(1-\theta)^{1-y_i^{\mathrm{obs}}}\right)\pi(\theta) \\
&= \theta^{\sum_{i=1}^{n} y_i^{\mathrm{obs}}}(1-\theta)^{\sum_{i=1}^{n}(1-y_i^{\mathrm{obs}})}\big(c\cdot \mathrm{beta}(N,N)(\theta)+(1-c)\mathrm{beta}(M,M)(\theta)\big) \\
&= \theta^{h_{\mathrm{obs}}}(1-\theta)^{t_{\mathrm{obs}}}\left(c\frac{\theta^{N-1}(1-\theta)^{N-1}}{\mathrm{B}(N,N)}+(1-c)\frac{\theta^{M-1}(1-\theta)^{M-1}}{\mathrm{B}(M,M)}\right) \\
&= c\frac{\theta^{N+h_{\mathrm{obs}}-1}(1-\theta)^{N+t_{\mathrm{obs}}-1}}{\mathrm{B}(N,N)}+(1-c)\frac{\theta^{M+h_{\mathrm{obs}}-1}(1-\theta)^{M+t_{\mathrm{obs}}-1}}{\mathrm{B}(M,M)} \\
&= \frac{c\cdot \mathrm{B}(N+h_{\mathrm{obs}},N+t_{\mathrm{obs}})}{\mathrm{B}(N,N)}\underbrace{\frac{\theta^{N+h_{\mathrm{obs}}-1}(1-\theta)^{N+t_{\mathrm{obs}}-1}}{\mathrm{B}(N+h_{\mathrm{obs}},N+t_{\mathrm{obs}})}}_{\substack{\parallel \\ \mathrm{beta}(N+h_{\mathrm{obs}},N+t_{\mathrm{obs}})(\theta)}} \\
&\qquad + \frac{(1-c)\mathrm{B}(M+h_{\mathrm{obs}},M+t_{\mathrm{obs}})}{\mathrm{B}(M,M)}\underbrace{\frac{\theta^{M+h_{\mathrm{obs}}-1}(1-\theta)^{M+t_{\mathrm{obs}}-1}}{\mathrm{B}(M+h_{\mathrm{obs}},M+t_{\mathrm{obs}})}}_{\substack{\parallel \\ \mathrm{beta}(M+h_{\mathrm{obs}},M+t_{\mathrm{obs}})(\theta)}} \\
&= \frac{\left(\begin{array}{l} c\cdot \mathrm{B}(N+h_{\mathrm{obs}},N+t_{\mathrm{obs}})\mathrm{B}(M,M)\mathrm{beta}(N+h_{\mathrm{obs}},N+t_{\mathrm{obs}}) \\ \qquad +(1-c)\cdot \mathrm{B}(M+h_{\mathrm{obs}},M+t_{\mathrm{obs}})\mathrm{B}(N,N)\mathrm{beta}(M+h_{\mathrm{obs}},M+t_{\mathrm{obs}}) \end{array}\right)}{\mathrm{B}(N,N)\mathrm{B}(M,M)} \\
&\propto c\cdot \mathrm{B}(N+h_{\mathrm{obs}},N+t_{\mathrm{obs}})\mathrm{B}(M,M)\mathrm{beta}(N+h_{\mathrm{obs}},N+t_{\mathrm{obs}})(\theta) \\
&\qquad + (1-c)\cdot \mathrm{B}(M+h_{\mathrm{obs}},M+t_{\mathrm{obs}})\mathrm{B}(N,N)\mathrm{beta}(M+h_{\mathrm{obs}},M+t_{\mathrm{obs}})(\theta)
\end{aligned}
$$

そこで

$$
\begin{aligned}
P &= c\cdot \mathrm{B}(N+h_{\mathrm{obs}},N+t_{\mathrm{obs}})\mathrm{B}(M,M), \\
Q &= (1-c)\cdot \mathrm{B}(M+h_{\mathrm{obs}},M+t_{\mathrm{obs}})\mathrm{B}(N,N)
\end{aligned}
$$

とおくと

$$
\pi(\theta \mid \mathbf{y}^{\mathrm{obs}}) \propto P\cdot \underbrace{\mathrm{beta}(N+h_{\mathrm{obs}},N+t_{\mathrm{obs}})(\theta)}_{\text{p.d.f. であるから積分したら 1}} + Q\cdot \underbrace{\mathrm{beta}(M+h_{\mathrm{obs}},M+t_{\mathrm{obs}})(\theta)}_{\text{p.d.f. であるから積分したら 1}}
$$

となる。ゆえに，次を得る。

$$
\pi(\theta \mid \mathbf{y}^{\mathrm{obs}}) = \frac{P}{P+Q}\mathrm{beta}(N+h_{\mathrm{obs}},N+t_{\mathrm{obs}})(\theta)+\frac{Q}{P+Q}\mathrm{beta}(M+h_{\mathrm{obs}},M+t_{\mathrm{obs}})(\theta)
$$

ここに現れた係数については，次のように計算できる。

$$
\begin{aligned}
\frac{P}{P+Q} &= \frac{c\cdot \mathrm{B}(N+h_{\mathrm{obs}},N+t_{\mathrm{obs}})\mathrm{B}(M,M)}{c\cdot \mathrm{B}(N+h_{\mathrm{obs}},N+t_{\mathrm{obs}})\mathrm{B}(M,M)+(1-c)\cdot \mathrm{B}(M+h_{\mathrm{obs}},M+t_{\mathrm{obs}})\mathrm{B}(N,N)} \\
&= \frac{c}{c+(1-c)\dfrac{\mathrm{B}(M+h_{\mathrm{obs}},M+t_{\mathrm{obs}})\mathrm{B}(N,N)}{\mathrm{B}(N+h_{\mathrm{obs}},N+t_{\mathrm{obs}})\mathrm{B}(M,M)}} = \frac{c}{c+(1-c)A} = \widetilde{c}, \\
\frac{Q}{P+Q} &= 1-\frac{P}{P+Q} = 1-\widetilde{c}
\end{aligned}
$$

問 8.19(➔p. 255) の解答例

$\mathrm{Beta}(K,L)$ の期待値は $\frac{K}{K+L}$ であるから，事後平均 $\widehat{\theta}_{\mathrm{true}}$ は次のように計算できる。

$$\begin{aligned}
\widehat{\theta}_{\mathrm{true}} &= \int_0^1 \theta\cdot\pi_n(\theta \mid \mathbf{y}^{\mathrm{obs}})\mathrm{d}\theta \\
&= \widetilde{c}_n \underbrace{\int_0^1 \theta\cdot\mathrm{beta}(N+\theta_{\mathrm{true}}n, N+(1-\theta_{\mathrm{true}})n)(\theta)\mathrm{d}\theta}_{=\frac{N+\theta_{\mathrm{true}}n}{\left(N+\theta_{\mathrm{true}}n\right)+\left(N+(1-\theta_{\mathrm{true}})n\right)}} \\
&\qquad + (1-\widetilde{c}_n)\underbrace{\int_0^1 \theta\cdot\theta\mathrm{beta}(M+\theta_{\mathrm{true}}n, M+(1-\theta_{\mathrm{true}})n)(\theta)\mathrm{d}\theta}_{=\frac{M+\theta_{\mathrm{true}}n}{\left(M+\theta_{\mathrm{true}}n\right)+\left(M+(1-\theta_{\mathrm{true}})n\right)}} \\
&= \widetilde{c}_n\frac{N+\theta_{\mathrm{true}}n}{2N+n} + (1-\widetilde{c}_n)\frac{M+\theta_{\mathrm{true}}n}{2M+n}
\end{aligned}$$

問 8.20(➔p. 257)

離散分布に対する汎化損失とエントロピーを定義に基づいて計算するのみである。

問 8.21(➔p. 258) の解答例

Stirling の公式 $\Gamma(x) \approx \sqrt{\frac{2\pi}{x}}\left(\frac{x}{\mathrm{e}}\right)^x$ より

$$\mathrm{B}(N,N) = \frac{\Gamma(N)\Gamma(N)}{\Gamma(2N)} \approx \frac{\frac{2\pi}{N}\left(\frac{N}{\mathrm{e}}\right)^{2N}}{\sqrt{\frac{2\pi}{2N}}\left(\frac{2N}{\mathrm{e}}\right)^{2N}} = \frac{(n\text{ によらない定数})}{2^{2an}\sqrt{n}},$$

$$\mathrm{B}(M+\theta_{\mathrm{true}}n, M+(1-\theta_{\mathrm{true}})n) = \frac{\Gamma(M+\theta_{\mathrm{true}}n)\Gamma(M+(1-\theta_{\mathrm{true}})n)}{\Gamma(2M+n)}$$

$$\approx \frac{\sqrt{\frac{(2\pi)^2}{(M+\theta_{\mathrm{true}}n)(M+(1-\theta_{\mathrm{true}})n)})}\left(\frac{M+\theta_{\mathrm{true}}n}{\mathrm{e}}\right)^{M+\theta_{\mathrm{true}}n}\left(\frac{M+(1-\theta_{\mathrm{true}})n}{\mathrm{e}}\right)^{M+(1-\theta_{\mathrm{true}})n}}{\sqrt{\frac{2\pi}{(2M+n)}}\left(\frac{2M+n}{\mathrm{e}}\right)^{2M+n}}$$

$$\approx \frac{(n\text{ によらない定数})}{\sqrt{n}}\left(\frac{M+\theta_{\mathrm{true}}n}{2M+n}\right)^{M+\theta_{\mathrm{true}}n}\left(\frac{M+(1-\theta_{\mathrm{true}})n}{2M+n}\right)^{M+(1-\theta_{\mathrm{true}})n}$$

であり，同様にして次を得る。

$$\mathrm{B}(N+\theta_{\mathrm{true}}n, N+(1-\theta_{\mathrm{true}})n) = \frac{\Gamma(N+\theta_{\mathrm{true}}n)\Gamma(N+(1-\theta_{\mathrm{true}})n)}{\Gamma(2N+n)}$$

$$\approx \frac{\sqrt{\frac{(2\pi)^2}{(N+\theta_{\mathrm{true}}n)(N+(1-\theta_{\mathrm{true}})n)})}\left(\frac{N+\theta_{\mathrm{true}}n}{\mathrm{e}}\right)^{N+\theta_{\mathrm{true}}n}\left(\frac{N+(1-\theta_{\mathrm{true}})n}{\mathrm{e}}\right)^{N+(1-\theta_{\mathrm{true}})n}}{\sqrt{\frac{2\pi}{2N+n}}\left(\frac{2N+n}{\mathrm{e}}\right)^{2N+n}}$$

$$\approx \frac{(n\text{ によらない定数})}{\sqrt{n}}\left(\frac{a+\theta_{\mathrm{true}}}{2a+1}\right)^{(a+\theta_{\mathrm{true}})n}\left(\frac{a+(1-\theta_{\mathrm{true}})}{2a+1}\right)^{(a+(1-\theta_{\mathrm{true}}))n}$$

以上から，次が成り立つことがわかる。

$$
\begin{aligned}
A_n &= \frac{\mathrm{B}(N,N)\mathrm{B}(M+\theta_{\mathrm{true}}n, M+(1-\theta_{\mathrm{true}})n)}{\mathrm{B}(N+\theta_{\mathrm{true}}n, N+(1-\theta_{\mathrm{true}})n)\mathrm{B}(M,M)} \\
&\approx \frac{(n\,\text{によらない定数})}{\sqrt{n}} \frac{\left(\dfrac{M+\theta_{\mathrm{true}}n}{2M+n}\right)^{M+\theta_{\mathrm{true}}n}\left(\dfrac{M+(1-\theta_{\mathrm{true}})n}{2M+n}\right)^{M+(1-\theta_{\mathrm{true}})n}}{2^{2an}\left(\dfrac{a+\theta_{\mathrm{true}}}{2a+1}\right)^{(a+\theta_{\mathrm{true}})n}\left(\dfrac{a+(1-\theta_{\mathrm{true}})}{2a+1}\right)^{(a+(1-\theta_{\mathrm{true}}))n}}
\end{aligned}
$$

問 8.22(➔p. 261) の解答例

(1) 逆に条件 $(\mathrm{A})_{0,\frac{1}{2},0}$ が成り立つとき，ある確率 $\mathbf{P}$ の世界で

$$
\mathbf{E}[\Theta^2] = \int_0^1 \theta^2 \mathbf{P}(\Theta \in \mathrm{d}\theta) = \mathbf{P}(Y_1 = 0; Y_2 = 0) \overset{\substack{\text{いまの場合}\\ c=0\\ \text{より}}}{=} 0,
$$

ゆえに $\mathbf{P}(\Theta = 0) = 1$ とならなければならない。特に $\displaystyle\int_0^1 \theta \mathbf{P}(\Theta \in \mathrm{d}\theta) = \mathbf{E}[\Theta] = 0$ である。ところが，このとき

$$
0 = \int_0^1 \theta\mu(\mathrm{d}\theta) - \int_0^1 \theta^2 \mathbf{P}(\Theta \in \mathrm{d}\theta) = \int_0^1 \theta(1-\theta)\mathbf{P}(\Theta \in \mathrm{d}\theta) = \mathbf{P}(Y_1 = 0; Y_2 = 1) \overset{\substack{\text{いまの場合}\\ b=\frac{1}{2}\\ \text{より}}}{=} \frac{1}{2},
$$

となり矛盾してしまう。ゆえに $(\mathrm{A})_{0,\frac{1}{2},0}$ は成り立たない。

(2) 条件 $(\mathrm{A})_{a,b,c}$ が成り立つとし，ある確率 $\mathbf{P}$ の世界で $\mu(\mathrm{d}\theta) = \mathbf{P}(\Theta \in \mathrm{d}\theta)$ とおくと，次を得る。

$$
\begin{aligned}
a &= \mathbf{P}(Y_1 = 0; Y_2 = 0) = \int_0^1 (1-\theta)^2 \mu(\mathrm{d}\theta), \\
b &= \mathbf{P}(Y_1 = 0; Y_2 = 1) = \mathbf{P}(Y_1 = 1; Y_2 = 0) = \int_0^1 \theta(1-\theta)\mu(\mathrm{d}\theta), \\
c &= \mathbf{P}(Y_1 = 1; Y_2 = 1) = \mathbf{P}\int_0^1 \theta^2 \mu(\mathrm{d}\theta)
\end{aligned}
$$

同様に，条件 $(\mathrm{A})_{a',b',c'}$ が成り立つとすると，ある確率 $\mathbf{P}'$ の世界で $\mu'(\mathrm{d}\theta) = \mathbf{P}'(\Theta \in \mathrm{d}\theta)$ とおくと次が成り立つ。

$$
\begin{aligned}
a' &= \mathbf{P}'(Y_1 = 0; Y_2 = 0) = \int_0^1 (1-\theta)^2 \mu'(\mathrm{d}\theta), \\
b' &= \mathbf{P}'(Y_1 = 0; Y_2 = 1) = \mathbf{P}(Y_1 = 1; Y_2 = 0) = \int_0^1 \theta(1-\theta)\mu'(\mathrm{d}\theta), \\
c' &= \mathbf{P}'(Y_1 = 1; Y_2 = 1) = \mathbf{P}\int_0^1 \theta^2 \mu'(\mathrm{d}\theta)
\end{aligned}
$$

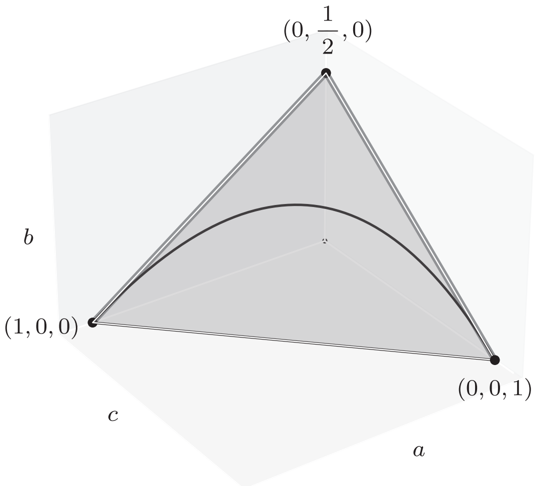
点 (a,b,c) と点 (a',b',c') を結ぶ線分上の点は，ある $\lambda \in (0,1)$ を用いて $a_\lambda = \lambda a + (1-\lambda)a'$，$b_\lambda = \lambda b + (1-\lambda)b'$，$c_\lambda = \lambda c + (1-\lambda)c'$ の 3 つ組 $(a_\lambda, b_\lambda, c_\lambda)$ により表すことができる。いま，

$\mathbf{P}_\lambda = \lambda\mathbf{P} + (1-\lambda)\mathbf{P}'$ とおくと，$\mathbf{P}_\lambda$ もまた，一つの確率の測り方となる。この $\mathbf{P}_\lambda$ の世界では

$$\begin{aligned}
a_\lambda &= \lambda a + (1-\lambda)a' \\
&= \lambda\int_0^1 (1-\theta)^2\mu(\mathrm{d}\theta) + (1-\lambda)\int_0^1 \theta^2\mu'(\mathrm{d}\theta) \\
&= \lambda\int_0^1 (1-\theta)^2\mathbf{P}(\Theta\in\mathrm{d}\theta) + (1-\lambda)\int_0^1 \theta^2\mathbf{P}'(\Theta\in\mathrm{d}\theta) \\
&= \int_0^1 (1-\theta)^2\underbrace{\big(\lambda\mathbf{P}(\Theta\in\mathrm{d}\theta) + (1-\lambda)\mathbf{P}'(\Theta\in\mathrm{d}\theta)\big)}_{=\mathbf{P}_\lambda(\Theta\in\mathrm{d}\theta)}
\end{aligned}$$

であり，同様に $b_\lambda = \displaystyle\int_0^1 \theta(1-\theta)\mathbf{P}_\lambda(\Theta\in\mathrm{d}\theta)$, $c_\lambda = \displaystyle\int_0^1 \theta^2\mathbf{P}_\lambda(\Theta\in\mathrm{d}\theta)$ が成り立ち，ゆえに $(\mathrm{A})_{a_\lambda,b_\lambda,c_\lambda}$ が成り立つ。

(3) p が $0\leqq p\leqq 1$ の範囲を動くとき，点 $(a,b,c) = (p^2, p(1-p), (1-p)^2)$ の軌跡を図示すると，次のようになる。



また，式変形 $\displaystyle\int_0^1 (\theta-p)^2\mu(\mathrm{d}\theta) = \int_0^1 \theta^2\mu(\mathrm{d}\theta) - 2p\int_0^1 \theta\mu(\mathrm{d}\theta) + p^2$ に注意すると，次の同値性がわかる。

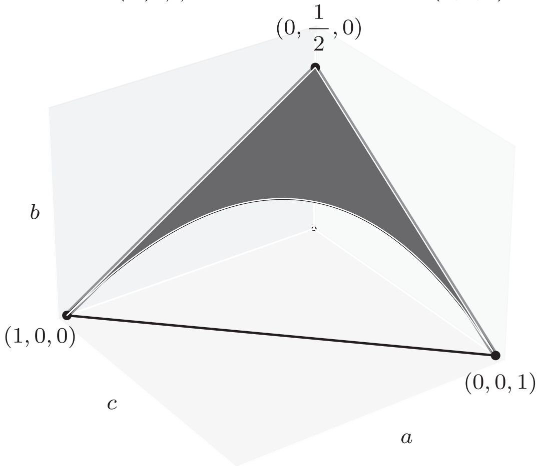
$$(\mathrm{A})_{p^2,p(1-p),(1-p)^2} \iff \left\{\begin{array}{l}
\text{ある確率 } \mathbf{P} \text{ について} \\
\quad \mathbf{P}(Y_1=0)=\mathbf{P}(Y_2=0)=p, \\
\quad \mathbf{P}(Y_1=1)=\mathbf{P}(Y_2=1)=1-p, \\
\quad \mathbf{P}(\Theta=p)=1, \\
\quad Y_1 \text{ と } Y_2 \text{ は独立である。}
\end{array}\right.$$

確率 p で裏が出るコインを 2 回投げるなど，右のような状況は確かにあるので，上の同値性から $(\mathrm{A})_{p^2,p(1-p),(1-p)^2}$ が成り立つ。

(4) 条件 $(\mathrm{A})_{a,b,c}$ が成り立つような (a,b,c) の集まりは，(2) より凸集合をなすから，(3) で描いた弧とその弦で囲まれる領域内の (a,b,c) について，$(\mathrm{A})_{a,b,c}$ が成り立つ。一方で，この外側に属する (a',b',c') については，$(\mathrm{A})_{a',b',c'}$ は成り立たない。なぜなら，逆に $(\mathrm{A})_{a',b',c'}$ が成り立ったとすると，(2) のときのようにして，点 (a,b,c) と点 (a',b',c') を結ぶ線分上のすべての点 $(a_\lambda,b_\lambda,c_\lambda)$ について $(\mathrm{A})_{a_\lambda,b_\lambda,c_\lambda}$ が成り立たなければならないが，特にこの線分と (3) で描いた弧は 1 点で交わる。この点を $(\mathrm{A})_{a_\lambda,b_\lambda,c_\lambda} = (\mathrm{A})_{p^2,p(1-p),(1-p)^2}$ と表すとき，(2) のときのようにして $1 = \mathbf{P}_\lambda(\Theta=p) = \lambda\mathbf{P}(\Theta=p) + (1-\lambda)\mathbf{P}'(\Theta=p)$ が成り立たなければならないが，これは $\mathbf{P}(\Theta=p) = \mathbf{P}'(\Theta=p) = 1$ を意味

する。ゆえに，いま考えた線分の端点 (a, b, c) と (a', b', c') が共に $(\mathrm{A})_{p^2, p(1-p), (1-p)^2}$ に一致しなければならないが，これはそれぞれの点の選び方に反する。

以上から，$(\mathrm{A})_{a,b,c}$ が成り立たないような (a, b, c) は，下図の黒く塗りつぶされた部分をなす。



問 9.15(➔ p. 277) の解答例

$$\mathbf{P}(|\overline{X}_n - p| \geqq a) = \mathbf{P}(\overline{X}_n - p \geqq a) + \mathbf{P}(\overline{X}_n - p \leqq -a) \tag{A.26}$$

であるが，まず，上側の裾 (式 (**A.26**) の右辺第 **1** 項) について

$$\mathbf{P}(\overline{X}_n \geqq p + a) = \mathbf{P}\left(\sum_{i=1}^{n} X_i \geqq n(p+a)\right) \overset{\substack{\textbf{Chernoff}\\\textbf{の不等式}\\(➔\,\textbf{p. 277})}}{\leqq} \inf_{t>0} \mathrm{e}^{-n(p+a)t} \mathbf{E}\left[\mathrm{e}^{t \sum_{i=1}^{n} X_i}\right]$$

が成り立つ。ここで，右辺に現れた期待値について

$$\mathbf{E}\left[\mathrm{e}^{t \sum_{i=1}^{n} X_i}\right] = \mathbf{E}[\mathrm{e}^{tX_1}]\mathbf{E}[\mathrm{e}^{tX_2}] \cdots \mathbf{E}[\mathrm{e}^{tX_n}] = \bigl(1 + p(\mathrm{e}^t - 1)\bigr)^n \overset{\substack{\textbf{不等式}\\ 1+y \leqq \mathrm{e}^y\\ \textbf{より}}}{\leqq} \mathrm{e}^{np(\mathrm{e}^t - 1)}$$

となるから，次の評価が得られる。

$$\mathbf{P}(\overline{X}_n \geqq p + a) \leqq \inf_{t>0} \underbrace{\mathrm{e}^{-n(p+a)t} \cdot \mathrm{e}^{np(\mathrm{e}^t - 1)}}_{\substack{\shortparallel \\ \mathrm{e}^{-n\{(p+a)t - p(\mathrm{e}^t - 1)\}}}}$$

$$\overset{\substack{t = \log(1 + \frac{a}{p})\\ \textbf{とおいた}}}{\leqq} \mathrm{e}^{-n\left\{(p+a)\log\left(1 + \frac{a}{p}\right) - p(\mathrm{e}^{\log\left(1 + \frac{a}{p}\right)} - 1)\right\}}$$

$$= \mathrm{e}^{-n\left\{(p+a)\log\left(1 + \frac{a}{p}\right) - a\right\}} \overset{\substack{\textbf{不等式}\\ \log(1+y) \geqq \frac{2y}{y+2}\\ \textbf{より}}}{\leqq} \mathrm{e}^{-n\left\{(p+a)\frac{2\frac{a}{p}}{\frac{a}{p} + 2} - a\right\}} = \mathrm{e}^{-n\frac{a^2}{2p+a}}$$

次に，下側の裾 (式 (A.26) の右辺第 2 項) については，上側の裾の場合と同様に次の評価が得られる。

$$
\begin{aligned}
\mathbf{P}(\overline{X}_n \leqq p-a) &\overset{\substack{\text{Chernoff}\\ \text{の不等式}\\ (➔\text{p. 277})}}{\leqq} \min_{t>0} \mathrm{e}^{n(p-a)t}\,\mathrm{e}^{np(\mathrm{e}^{-t}-1)} = \min_{t>0} \mathrm{e}^{-n\{p(1-\mathrm{e}^{-t})-(p-a)t\}} \\
&\overset{\substack{t=-\log(1-\frac{a}{p})\\ \text{とおいた}}}{\leqq} \mathrm{e}^{-n\left\{p\left(1-\mathrm{e}^{\log\left(1-\frac{a}{p}\right)}\right)+(p-a)\log\left(1-\frac{a}{p}\right)\right\}} \\
&= \mathrm{e}^{-n\left\{a+p\left(1-\frac{a}{p}\right)\log\left(1-\frac{a}{p}\right)\right\}} \overset{\substack{\text{不等式}\\ (1-y)\log(1-y)\geqq\frac{1}{2}y^2-y\\ \text{より}}}{\leqq} \mathrm{e}^{-n\left\{a+p\left(\frac{1}{2}\frac{a^2}{p^2}-\frac{a}{p}\right)\right\}} = \mathrm{e}^{-n\frac{a^2}{2p}}
\end{aligned}
$$

ここで，$a \in (0,p)$ なので $\frac{a^2}{2p+a}-\frac{a^2}{3p}=a^2\frac{p-a}{3p(2p+a)} \geqq 0$ となり，$\mathrm{e}^{-n\frac{a^2}{2p+a}} \leqq \mathrm{e}^{-n\frac{a^2}{3p}}$ であることに注意すれば，$\mathrm{e}^{-n\frac{a^2}{2p}} \leqq \mathrm{e}^{-n\frac{a^2}{3p}}$ であることと合わせると，次が成り立つことがわかる。

$$
\mathbf{P}(|\overline{X}_n-p| \geqq a) \leqq \mathrm{e}^{-n\frac{a^2}{2p+a}}+\mathrm{e}^{-n\frac{a^2}{2p}} \leqq \mathrm{e}^{-n\frac{a^2}{3p}}+\mathrm{e}^{-n\frac{a^2}{3p}} = 2\mathrm{e}^{-n\frac{a^2}{3p}}
$$

問 9.17(➔p. 280) の解答例

(1) KKT 条件は次によって与えられる。

(i) $\dfrac{\partial L}{\partial p_l}(p^*,\lambda^*,\mu^*,\nu^*)=0,\ \ l=1,2,\ldots,d,$

(ii) $p_1^*,p_2^*,\ldots,p_d^* \geqq 0,\ \ \displaystyle\sum_{l=1}^d p_l^*=1,\ \ \text{かつ}\ \sum_{l=1}^d g_j(a_l)p_l \geqq c_j,\ \ j=1,2,\ldots,m,$

(iii) $L(p^*,\lambda^*,\mu^*,\nu^*)$ の式に現れる次の式はいずれも 0 である。

$$
\begin{aligned}
&\lambda_1^*p_1^*,\quad \ldots,\quad \lambda_d^*p_d^*,\\
&\mu_1^*\left(c_1-\sum_{l=1}^d g_1(a_l)p_l^*\right),\quad \ldots,\quad \mu_m^*\left(c_m-\sum_{l=1}^d g_m(a_l)p_l^*\right),\\
&\nu^*\left(\sum_{l=1}^d p_l^*-1\right)
\end{aligned}
$$

(i) については，

$$
0=\frac{\partial L}{\partial p_l}(p^*,\lambda^*,\mu^*,\nu^*)=\log\left|\frac{p_l^*}{q_l}\right|+1-\lambda_l^*-\sum_{j=1}^m \mu_j^* g_j(a_l)+\nu^*
$$

であることと，(iii) の最後の項が 0 であることは (ii) に現れる条件 $\sum_{l=1}^d p_l^*=1$ から自動的に従うことから，題意が確かめられる。

(2) $l=1,2,\ldots,d$ を任意とすると。KKT 条件より $p_l^*=q_l\cdot\exp\left(\lambda_l^*+\displaystyle\sum_{j=1}^m \mu_j^* g_j(a_l)\right)\cdot\mathrm{e}^{-(1+\nu^*)}$ であるから，特に $p_l^*=0 \Leftrightarrow q_l=0$ となる。さらに $p_l^*>0$ と仮定すると，再び KKT 条件より $\lambda_l^*=0$ とならなければならず，ゆえに

$$
p_l^*=q_l\cdot\exp\left(\sum_{j=1}^m \mu_j^* g_j(a_l)\right)\cdot\mathrm{e}^{-(1+\nu^*)} \tag{A.27}
$$

である。「$p_l^* = 0 \Leftrightarrow q_l = 0$」であったから，この等式は $p_l^* > 0$ という条件なしで，すべての $l = 1, 2, \ldots, d$ について成立する。よって $1 = \sum_{l=1}^{d} p_l^* = \mathrm{e}^{-(1+\nu^*)} \sum_{l=1}^{d} q_l \exp\left(\sum_{j=1}^{m} \mu_j^* g_j(a_l)\right)$ であり，式を整理して $\mathrm{e}^{1+\nu^*} = \sum_{l=1}^{d} q_l \exp\left(\sum_{j=1}^{m} \mu_j^* g_j(a_l)\right)$ を得る。これを式 (A.27) に代入して，次式を得る。

$$p_l^* = \frac{q_l \exp\left(\sum_{j=1}^{m} \mu_j^* g_j(a_l)\right)}{\sum_{k=1}^{d} q_k \exp\left(\sum_{j=1}^{m} \mu_j^* g_j(a_k)\right)}, \quad l = 1, 2, \ldots, d$$

問 9.18(➔p. 281) の解答例

(1) 問 9.17 (➔p. 280) より従う。

(2) 仮定より $\mu^* > 0$ であるから，$\sum_{l=1}^{d} p_l^* = a$ でなければならないことに注意する。題意を示すには $\inf_{x \geqq a} \Lambda(x) = D_{\mathrm{KL}}(p^* \| q)$ が成り立つことを示せばよい。$M(\mu) = \sum_{l=1}^{d} q_l \cdot \mathrm{e}^{\mu a_l}$ とおくと，(1) より $p_l^* = \dfrac{q_l \cdot \mathrm{e}^{\mu^* a_l}}{M(\mu^*)}$ であるから，次の式変形ができる。

$$D_{\mathrm{KL}}(p^* \| q) = \sum_{l=1}^{d} p_l^* \log \frac{p_l^*}{q_l} = \sum_{l=1}^{d} p_l^* \log \frac{\mathrm{e}^{\mu^* a_l}}{M(\mu^*)} = \mu^* \underbrace{\sum_{l=1}^{d} a_l p_l^*}_{=a} - \log M(\mu^*)$$

次に，$\Lambda(x)$ が $x \geqq a$ の範囲で単調増加であることを確認しておこう。そのために $\Lambda(x)$ の別表現を与えておく。$\Lambda(x)$ を計算するには，μ の関数 $f(\mu) = \mu x - \log \sum_{l=1}^{d} q_l \cdot \mathrm{e}^{\mu a_l}$ の最大点を求めればよく，これは

$$0 = f'(\mu) = \frac{\mathrm{d}}{\mathrm{d}\mu}\left\{\mu x - \log \sum_{l=1}^{d} q_l \cdot \mathrm{e}^{\mu a_l}\right\} = x - \sum_{l=1}^{d} a_l \frac{q_l \cdot \mathrm{e}^{\mu a_l}}{M(\mu)}$$

をみたさなければならない。この右辺は $\mu = 0$ において $f'(0) = x - \sum_{l=1}^{d} a_l q_l$ となるが，$x \geqq a$ よりこの値は $a - \sum_{l=1}^{d} a_l q_l$ 以上である。いま $\mu^* > 0$ の仮定と (1) から $q \neq p^*$ であり，ゆえに $q \notin \Gamma$ であることにより，この値は非負である。よって $f'(0) \geqq 0$ である。また $f'(\mu)$ を微分すると

$$f''(\mu) = \left(\sum_{l=1}^{d} a_l \frac{q_l \cdot \mathrm{e}^{\mu a_l}}{M(\mu)}\right)^2 - \sum_{l=1}^{d} (a_l)^2 \frac{q_l \cdot \mathrm{e}^{\mu a_l}}{M(\mu)} \leqq 0$$

であることがわかる。したがって上の式 $f'(\mu) = 0$ を μ について解き，その解を $\mu = \mu(x)$ と表すと，この値は非負とならなければならない。このとき $\Lambda(x) = \mu(x)x - \log \sum_{l=1}^{d} q_l \cdot \mathrm{e}^{\mu(x) a_l}$ と表現できる。これを x について微分すると

$$\Lambda'(x) = \mu'(x)x + \mu(x) - \mu'(x) \underbrace{\sum_{l=1}^{d} a_l \frac{q_l \cdot \mathrm{e}^{\mu(x) a_l}}{M(\mu(x))}}_{=x} = \mu(x) \geqq 0$$

であり，ゆえに $\Lambda(x)$ は単調増加関数である。したがって $\inf_{x \geqq a} \Lambda(x) = \Lambda(a)$ である。

最後に，$\Lambda(a)$ を計算しよう。このためには μ の関数 $\mu a - \log\sum_{l=1}^{d} q_l \cdot \mathrm{e}^{\mu a_l}$ の最大点を求めればよく，これは

$$0 = \frac{\mathrm{d}}{\mathrm{d}\mu}\left\{\mu a - \log\sum_{l=1}^{d} q_l \cdot \mathrm{e}^{\mu a_l}\right\} = a - \sum_{l=1}^{d} a_l \frac{q_l \cdot \mathrm{e}^{\mu a_l}}{M(\mu)}$$

を μ について解けばよいが，これはいまの場合 $\mu = \mu^*$ のときにみたされる。ゆえに，次式を得る。

$$\inf_{x \geqq a} \Lambda(x) = \Lambda(a) = \mu^* a - \log \underbrace{\sum_{l=1}^{d} q_l \cdot \mathrm{e}^{\mu^* a_l}}_{= M(\mu^*)} = D_{\mathrm{KL}}(p^* \| q) = \inf_{p \in \Gamma} D_{\mathrm{KL}}(p \| q)$$

問 B.3(➔「付録　行列代数」p. 1) の解答

(1) δ_{ij} (2) a_i

問 B.5(➔「付録　行列代数」p. 2) の解答例

$X = XE_n = X(AY) = (XA)Y = E_nY = Y$

問 B.11(➔「付録　行列代数」p. 4) の解答

(1) $\det\begin{pmatrix} 1 & 2 \\ 3 & 2 \end{pmatrix} = 1 \cdot 2 - 2 \cdot 3 = -4$ (2) $\det\begin{pmatrix} 1 & 2 \\ 3 & 6 \end{pmatrix} = 1 \cdot 6 - 2 \cdot 3 = 0$

(3) $\det E_2 = \det\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = 1 \cdot 1 - 0 \cdot 0 = 1$

(4) $\det(-E_2) = \det\left(-\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}\right) = \det\begin{pmatrix} -1 & 0 \\ 0 & -1 \end{pmatrix} = 1$

問 B.13(➔「付録　行列代数」p. 5) の解答例

公式 B.12 (➔p. 4) より $(\det X)(\det A) = \det(XA) = \det E_2 = 1 \neq 0$ であるから $\det A \neq 0$ となる。

問 B.32(➔「付録　行列代数」p. 15) の解答例

(1) $\begin{pmatrix} c \\ c \\ c \end{pmatrix} = c\begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$ (2) $\begin{pmatrix} a \\ b+c \\ 0 \end{pmatrix} = a\begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} + b\begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} + c\begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}$ と表せるが，同じベクトルが現れたときは，それで括っておこう。すると $\begin{pmatrix} a \\ b+c \\ 0 \end{pmatrix} = a\begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} + (b+c)\begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}$ と表現できる。b と c が任意パラメータであることから，$b+c$ を一つの任意パラメータと考えるのである。こうして現れるパラメータを含まないベクトルたちは，どれか一つが他の一次結合として表現できないように選ぶ習慣をつけるとよい。

(3) $\begin{pmatrix} a+3c \\ 2b+a \\ c+a+2b \end{pmatrix} = a\begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} + 2b\begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix} + c\begin{pmatrix} 3 \\ 0 \\ 1 \end{pmatrix}$

問 B.39(➔「付録　行列代数」p. 21) の解答例

$\det(A-\lambda E_2) = \det\begin{pmatrix} -\lambda & 1 \\ 0 & -\lambda \end{pmatrix} = \lambda^2$ であるから，A の固有方程式は $\lambda^2 = 0$ を解くと，固有値は $\lambda = 0$ しかないことがわかる。しかし A は対角行列ではないので，例 B.38 (➔p. 20) の事情により A は対角化可能でない。

問 B.44(➔「付録　行列代数」p. 23) の解答

(1) A_{ji} (2) $\overline{A_{ij}}$ (3) $\overline{A_{ji}}$ (4) $\overline{A_{ji}}$

問 B.46(➔「付録　行列代数」p. 24) の解答例

(1) $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ のとき，$\det(A^\top) = \det\begin{pmatrix} a & b \\ c & d \end{pmatrix}^\top = \det\begin{pmatrix} a & c \\ b & d \end{pmatrix} = ad - cb = ad - bc = \det\begin{pmatrix} a & b \\ c & d \end{pmatrix} = \det A$ である。

(2) $X = (A^{-1})^\top$ とおくと，$A^\top X = A^\top (A^{-1})^\top \overset{\text{公式 B.45 (➔p. 24)}}{=} (A^{-1}A)^\top = E_2^\top = E_2$ と計算できる。ゆえに，定理 B.16 (➔p. 6) より $A^\top$ は正則であり，さらに $(A^\top)^{-1} = X = (A^{-1})^\top$ である。

問 B.49(➔「付録　行列代数」p. 25) の解答例

(1) n に関する数学的帰納法による。R_2 が 2 次直交行列であることは例 B.48 (➔p.24) の通り。R_k が k 次直交行列であるとすると，

$$
\begin{aligned}
R_{k+1}^\top R_{k+1} &= \left(\begin{pmatrix} R_k & \mathbf{0} \\ \mathbf{0} & 1 \end{pmatrix}\begin{pmatrix} E_{k-1} & \mathbf{0} \\ \mathbf{0} & R_2(\theta_{k-1}) \end{pmatrix}\right)^\top \left(\begin{pmatrix} R_k & \mathbf{0} \\ \mathbf{0} & 1 \end{pmatrix}\begin{pmatrix} E_{k-1} & \mathbf{0} \\ \mathbf{0} & R_2(\theta_k) \end{pmatrix}\right) \\
&= \begin{pmatrix} E_{k-1} & \mathbf{0} \\ \mathbf{0} & R_2(\theta_k) \end{pmatrix}^\top \begin{pmatrix} R_k & \mathbf{0} \\ \mathbf{0} & 1 \end{pmatrix}^\top \begin{pmatrix} R_k & \mathbf{0} \\ \mathbf{0} & 1 \end{pmatrix}\begin{pmatrix} E_{k-1} & \mathbf{0} \\ \mathbf{0} & R_2(\theta_k) \end{pmatrix} \\
&= \begin{pmatrix} E_{k-1} & \mathbf{0} \\ \mathbf{0} & R_2(\theta_k)^\top \end{pmatrix}\begin{pmatrix} R_k^\top & \mathbf{0} \\ \mathbf{0} & 1 \end{pmatrix}\begin{pmatrix} R_k & \mathbf{0} \\ \mathbf{0} & 1 \end{pmatrix}\begin{pmatrix} E_{k-1} & \mathbf{0} \\ \mathbf{0} & R_2(\theta_k) \end{pmatrix} \\
&= \begin{pmatrix} E_{k-1} & \mathbf{0} \\ \mathbf{0} & R_2(\theta_k)^\top \end{pmatrix}\begin{pmatrix} R_k^\top R_k & \mathbf{0} \\ \mathbf{0} & 1 \end{pmatrix}\begin{pmatrix} E_{k-1} & \mathbf{0} \\ \mathbf{0} & R_2(\theta_k) \end{pmatrix} \\
&= \begin{pmatrix} E_{k-1} & \mathbf{0} \\ \mathbf{0} & R_2(\theta_k)^\top \end{pmatrix}\begin{pmatrix} E_k & \mathbf{0} \\ \mathbf{0} & 1 \end{pmatrix}\begin{pmatrix} E_{k-1} & \mathbf{0} \\ \mathbf{0} & R_2(\theta_k) \end{pmatrix}
\end{aligned}
$$

ここで $\begin{pmatrix} E_k & \mathbf{0} \\ \mathbf{0} & 1 \end{pmatrix} = E_{k+1}$ であるから

$$
\begin{aligned}
R_{k+1}^\top R_{k+1} &= \begin{pmatrix} E_{k-1} & \mathbf{0} \\ \mathbf{0} & R_2(\theta_k)^\top \end{pmatrix}\begin{pmatrix} E_{k-1} & \mathbf{0} \\ \mathbf{0} & R_2(\theta_k) \end{pmatrix} \\
&= \begin{pmatrix} E_{k-1} & \mathbf{0} \\ \mathbf{0} & R_2(\theta_k)^\top R_2(\theta_k) \end{pmatrix} = \begin{pmatrix} E_{k-1} & \mathbf{0} \\ \mathbf{0} & E_2 \end{pmatrix} = E_{k+1}
\end{aligned}
$$

となり，やはり R_{k+1} もまた $(k+1)$ 次直交行列となる。
(2) R_k の列ベクトル表示を $R_k=(\mathbf{u}_1^{(k)},\mathbf{u}_2^{(k)},\cdots,\mathbf{u}_k^{(k)})$ とするとき，$\mathbf{u}_k^{(k)}$ を計算すればよい。

$$
\begin{aligned}
R_{k+1} &= \begin{pmatrix} R_k & \mathbf{0} \\ \mathbf{0} & 1 \end{pmatrix} \left(\begin{array}{c:cc} E_{k-1} & \multicolumn{2}{c}{\mathbf{0}} \\ \hdashline \mathbf{0} & \cos\theta_k & \sin\theta_k \\ & -\sin\theta_k & \cos\theta_k \end{array}\right) \\
&= \left(\begin{array}{ccc:cc} \mathbf{u}_1^{(k)} & \cdots & \mathbf{u}_{k-1}^{(k)} & \mathbf{u}_k^{(k)} & \mathbf{0} \\ \hdashline 0 & \cdots & 0 & 0 & 1 \end{array}\right) \left(\begin{array}{c:cc} E_{k-1} & \multicolumn{2}{c}{\mathbf{0}} \\ \hdashline \mathbf{0} & \cos\theta_k & \sin\theta_k \\ & -\sin\theta_k & \cos\theta_k \end{array}\right) \\
&= \left(\begin{array}{ccc:cc} \mathbf{u}_1^{(k)} & \cdots & \mathbf{u}_{k-1}^{(k)} & \mathbf{u}_k^{(k)}(\cos\theta_k) & \mathbf{u}_k^{(k)}(\sin\theta_k) \\ \hdashline 0 & \cdots & 0 & -\sin\theta_k & \cos\theta_k \end{array}\right)
\end{aligned}
$$

であり，これが $R_{k+1}=(\ \mathbf{u}_1^{(k+1)}\ \ \cdots\ \ \mathbf{u}_{k-1}^{(k+1)}\ \ \mathbf{u}_k^{(k+1)}\ \ \mathbf{u}_{k+1}^{(k+1)}\ )$ と等しいのであるから，特に最も右側に位置する列ベクトルを比較すると $\mathbf{u}_{k+1}^{(k+1)}=\left(\begin{array}{c} \mathbf{u}_k^{(k)}(\sin\theta_k) \\ \hdashline \cos\theta_k \end{array}\right)$ とならなければならない。
$\mathbf{u}_2^{(2)}=\begin{pmatrix} \sin\theta_1 \\ \cos\theta_1 \end{pmatrix}$ であるから，

$$
\mathbf{u}_3^{(3)}=\left(\begin{array}{c} \mathbf{u}_2^{(2)}(\sin\theta_2) \\ \hdashline \cos\theta_2 \end{array}\right)=\left(\begin{array}{c} \begin{pmatrix} \sin\theta_1 \\ \cos\theta_1 \end{pmatrix}(\sin\theta_2) \\ \hdashline \cos\theta_2 \end{array}\right)=\begin{pmatrix} (\sin\theta_1)(\sin\theta_2) \\ (\cos\theta_1)(\sin\theta_2) \\ \cos\theta_2 \end{pmatrix}
$$

$$
\mathbf{u}_4^{(4)}=\left(\begin{array}{c} \mathbf{u}_3^{(3)}(\sin\theta_3) \\ \hdashline \cos\theta_3 \end{array}\right)=\left(\begin{array}{c} \begin{pmatrix} (\sin\theta_1)(\sin\theta_2) \\ (\cos\theta_1)(\sin\theta_2) \\ \cos\theta_2 \end{pmatrix}(\sin\theta_3) \\ \hdashline \cos\theta_3 \end{array}\right)=\begin{pmatrix} (\sin\theta_1)(\sin\theta_2)(\sin\theta_3) \\ (\cos\theta_1)(\sin\theta_2)(\sin\theta_3) \\ (\cos\theta_2)(\sin\theta_3) \\ \cos\theta_3 \end{pmatrix}
$$

を繰り返して帰納的に題意を得る。

問 B.55(➔「付録　行列代数」p. 27) の解答例

(1) $\mathbf{u}=\begin{pmatrix} u_1 \\ u_2 \end{pmatrix}$ とおくとき，条件より $1=\|\mathbf{u}\|^2=(u_1)^2+(u_2)^2$，つまり $(u_1)^2+(u_2)^2=1$ をみたす。そこで座標平面における (u_2,u_1) が表す点の偏角を θ とおけば $u_1=\sin\theta$，$u_2=\cos\theta$ とかける。ゆえに $\mathbf{u}=\begin{pmatrix} \sin\theta \\ \cos\theta \end{pmatrix}$ である。
(2) $\mathbf{u}=(u_1,u_2,\ldots,u_n)^\top$ とおくとき，条件より $1=\|\mathbf{u}\|^2=(u_1)^2+(u_2)^2+\cdots+(u_n)^2$，つまり $\sum_{k=1}^n(u_k)^2=1$ が成り立つ。
特に $-1\leqq u_n\leqq 1$ であるから，$u_n=\cos\theta_{n-1}$ となる実数 θ_{n-1} が取れる。すると $\sum_{k=1}^{n-1}(u_k)^2=1-(u_n)^2=1-\cos^2\theta_{n-1}=\sin^2\theta_{n-1}$ が成り立つ。
特に $-|\sin\theta_{n-1}|\leqq u_{n-1}\leqq|\sin\theta_{n-1}|$ が成り立つので，うまく実数 θ_{n-2} を選べば $u_{n-1}=(\cos\theta_{n-2})(\sin\theta_{n-2})$ が成り立つ。すると $\sum_{k=1}^{n-2}(u_k)^2=\sin^2\theta_{n-1}-(u_{n-1})^2=\sin^2\theta_{n-1}(1-\cos^2\theta_{n-2})=(\sin^2\theta_{n-2})(\sin^2\theta_{n-1})$ が成り立つ。
特に $-|\sin\theta_{n-2}\sin\theta_{n-1}|\leqq u_{n-2}\leqq|\sin\theta_{n-2}\sin\theta_{n-1}|$ が成り立つので，うまく実数 θ_{n-3} を選べば $u_{n-2}=(\cos\theta_{n-3})(\sin\theta_{n-2})(\sin\theta_{n-1})$ が成り立つ。すると $\sum_{k=1}^{n-3}(u_k)^2=(\sin^2\theta_{n-2})(\sin^2\theta_{n-1})-(u_{n-2})^2=(\sin^2\theta_{n-2})(\sin^2\theta_{n-1})(1-\cos^2\theta_{n-3})=(\sin^2\theta_{n-3})(\sin^2\theta_{n-2})(\sin^2\theta_{n-1})$ が成り立つ。
この手続きを繰り返していくと，うまく $\theta_{n-4},\ldots,\theta_2,\theta_1$ を選べば

$$
u_2=(\cos\theta_1)(\sin\theta_2)(\sin\theta_3)\cdots(\sin\theta_{n-1})
$$

とかけることがわかる。すると，次が成り立つ。

$$
\begin{aligned}
(u_1)^2 &= (\sin^2\theta_2)(\sin^2\theta_3)\cdots(\sin^2\theta_{n-1}) - (u_2)^2 \\
&= (\sin^2\theta_2)(\sin^2\theta_3)\cdots(\sin^2\theta_{n-1})(1-\cos^2\theta_1) \\
&= (\sin^2\theta_1)(\sin^2\theta_2)(\sin^2\theta_3)\cdots(\sin^2\theta_{n-1})
\end{aligned}
$$

したがって，必要なら θ_1 を $-\theta_1$ に取り換えることで (**この取り替えが行われても u_2 の値は変わらない**) $u_1 = \sin\theta_1\sin\theta_2\sin\theta_3\cdots\sin\theta_{n-1}$ とかける。

問 B.58(➔「付録　行列代数」p. 32) の解答例

(1) $\mathbf{1}_n = (\underbrace{1,\cdots,1}_{n})^\top$ は Q の固有値 1 の固有ベクトルである。Q を基本変形すれば $\mathrm{rank}(Q)=1$ であることは容易に確かめられる。

(2) 問 B.49 (**➔「付録　行列代数」p. 25**) の直交行列 R_n のパラメータ $\theta_1,\theta_2,\cdots,\theta_{n-1}$ を

$$
(R_n \text{ の第 } n \text{ 列}) = \begin{pmatrix} (\sin\theta_1)(\sin\theta_2)\cdots(\sin\theta_{n-1}) \\ (\cos\theta_1)(\sin\theta_2)\cdots(\sin\theta_{n-1}) \\ (\cos\theta_2)(\sin\theta_3)\cdots(\sin\theta_{n-1}) \\ \vdots \\ (\cos\theta_{n-2})(\sin\theta_{n-2}) \\ \cos\theta_{n-1} \end{pmatrix} = \frac{1}{\sqrt{n}}\begin{pmatrix} 1 \\ 1 \\ 1 \\ \vdots \\ 1 \\ 1 \end{pmatrix} \quad (= \mathbf{u}_n \text{ とおく})
$$

となるようにとり，R_n の列ベクトルを $R_n = (\mathbf{u}_1,\cdots,\mathbf{u}_{n-1},\mathbf{u}_n)$ と表すことにする。(1) の解答から $Q\mathbf{u}_n = \mathbf{u}_n$ であり，$k=1,\cdots,n-1$ に対しては

$$
Q\mathbf{u}_k = \begin{pmatrix} \frac{1}{n}((\mathbf{u}_k)_1 + (\mathbf{u}_k)_2 + \cdots + (\mathbf{u}_k)_n) \\ \frac{1}{n}((\mathbf{u}_k)_1 + (\mathbf{u}_k)_2 + \cdots + (\mathbf{u}_k)_n) \\ \vdots \\ \frac{1}{n}((\mathbf{u}_k)_1 + (\mathbf{u}_k)_2 + \cdots + (\mathbf{u}_k)_n) \end{pmatrix} = \frac{1}{\sqrt{n}}\begin{pmatrix} \langle \mathbf{u}_k, \mathbf{u}_n\rangle \\ \langle \mathbf{u}_k, \mathbf{u}_n\rangle \\ \vdots \\ \langle \mathbf{u}_k, \mathbf{u}_n\rangle \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix} = \mathbf{0}
$$

であるから，この R_n が求める直交行列の一例となっている。

(3) は自明であろう。