演習問題解答 （続・わかりやすいパターン認識）

第 1 章

【演習問題 1.1】

検査薬 A，B で続けて 2 回陽性となることを「陽^2」で表すと，ベイズの定理より

$$P(\text{感} \mid \text{陽}^2) = \frac{P(\text{陽}^2 \mid \text{感})}{P(\text{非})P(\text{陽}^2 \mid \text{非}) + P(\text{感})P(\text{陽}^2 \mid \text{感})} \cdot P(\text{感}) \qquad \text{(S.1.1)}$$

が成り立つ．ここで，両検査は独立であることに注意すると

$$P(\text{陽}^2 \mid \text{感}) = 0.98 \times 0.97 \qquad \text{(S.1.2)}$$

$$P(\text{陽}^2 \mid \text{非}) = 0.01 \times 0.04 \qquad \text{(S.1.3)}$$

であるので，これらを式 (S.1.1) に代入し

$$P(\text{感} \mid \text{陽}^2) = \frac{0.98 \times 0.97 \times 0.001}{0.999 \times 0.01 \times 0.04 + 0.001 \times 0.98 \times 0.97} \qquad \text{(S.1.4)}$$

$$= 0.704 \qquad \text{(S.1.5)}$$

を得る．この結果は式 (1.37) と一致する．

【演習問題 1.2】

この n 種の検査を受けてすべて陽性となったことを「陽^n」で表すと，これらの検査は互いに独立であるので

$$P(\text{陽}^n \mid \text{感}) = P(\text{陽} \mid \text{感})^n \qquad \text{(S.1.6)}$$

$$P(\text{陽}^n \mid \text{非}) = P(\text{陽} \mid \text{非})^n \qquad \text{(S.1.7)}$$

は明らかである．したがって，ある人がこの n 種の検査を受けてすべて陽性となったとき，この人がこの病気に感染している確率は，ベイズの定理より

$$P(\text{感} \mid \text{陽}^n) = \frac{P(\text{陽}^n \mid \text{感})}{P(\text{非})P(\text{陽}^n \mid \text{非}) + P(\text{感})P(\text{陽}^n \mid \text{感})} \cdot P(\text{感}) \tag{S.1.8}$$

$$= \frac{P(\text{陽} \mid \text{感})^n}{P(\text{非})P(\text{陽} \mid \text{非})^n + P(\text{感})P(\text{陽} \mid \text{感})^n} \cdot P(\text{感}) \tag{S.1.9}$$

と求められる．

ここで，式 (S.1.9) の分母，分子を $P(\text{陽} \mid \text{感})^n$ で割ると

$$P(\text{感} \mid \text{陽}^n) = \frac{P(\text{感})}{P(\text{非}) \cdot \big(P(\text{陽} \mid \text{非})/P(\text{陽} \mid \text{感})\big)^n + P(\text{感})} \tag{S.1.10}$$

を得る．**表 1.3** によれば，

$$\left(\frac{P(\text{陽} \mid \text{非})}{P(\text{陽} \mid \text{感})}\right)^n = \left(\frac{0.01}{0.98}\right)^n \tag{S.1.11}$$

であるので，n が大きくなると，この項は限りなく 0 に近づく．その結果，式 (S.1.10) は，罹患率 $P(\text{感})$ の値に関わらず 1 に近づくことがわかる．

【演習問題 1.3】

ベイズの定理の式 (1.45) に用いるのは以下の 3 式である．

$$P(o_B|\omega_A) = 1/2 \tag{S.1.12}$$

$$P(o_B|\omega_C) = 1/2 \tag{S.1.13}$$

$$P(o_B|\omega_B) = 0 \tag{S.1.14}$$

以上を式 (1.45) に代入すると

$$P(\omega_A|o_B)$$

$$= \frac{(1/3) \cdot (1/2)}{(1/3) \cdot (1/2) + (1/3) \cdot 0 + (1/3) \cdot (1/2)} \tag{S.1.15}$$

$$= 1/2 \tag{S.1.16}$$

が得られる．

【演習問題 1.4】

問 (1)

これまでと同じ記法を用い，扉 D_k が当たりであることを $S = \omega_k$，司会者が扉 D_k を開けることを $X = o_k$ で表すことにする．

ゲストが選んだ扉 D_1 が当たりである確率は，この時点で p_1 すなわち $P(\omega_1)$ である．以下，司会者のとる行動を場合分けして考える．

もし，ゲストの選んだ扉 D_1 が当たりの場合，司会者は $(n-1)$ 枚の扉 D_2〜D_n の何れかを無作為に選んで開けるはずである．したがって，扉 D_1 が当たりという条件で司会者が扉 D_2 を開ける確率 $P(o_2|\omega_1)$ は

$$P(o_2|\omega_1) = \frac{1}{n-1} \tag{S.1.17}$$

となる．

もし，扉 D_2 が当たりの場合，司会者がその扉を開けることはないので，

$$P(o_2|\omega_2) = 0 \tag{S.1.18}$$

である．

もし，扉 D_3〜D_n のいずれかが当たりのとき，例えば D_k $(3 \leq k \leq n)$ が当たりのとき，司会者は D_1 と D_k を除く $(n-2)$ 枚の扉の何れかを無作為に選んで開けるはずである．したがって，扉 D_2 を開ける確率は

$$P(o_2|\omega_i) = \frac{1}{n-2} \qquad (i = 3, 4, \ldots, n) \tag{S.1.19}$$

となる．これらより，司会者が扉 D_2 を開ける確率 $P(o_2)$ は

$$\begin{aligned}
&P(o_2) \\
&= \sum_{i=1}^{n} P(\omega_i)P(o_2|\omega_i) &\text{(S.1.20)}\\
&= P(\omega_1)P(o_2|\omega_1) + P(\omega_2)P(o_2|\omega_2) + \sum_{i=3}^{n} P(\omega_i)P(o_2|\omega_i) &\text{(S.1.21)}\\
&= \frac{1}{n-1}P(\omega_1) + \frac{1}{n-2}\sum_{i=3}^{n} P(\omega_i) &\text{(S.1.22)}
\end{aligned}$$

となる．ベイズの定理は，

$$P(\omega_1|o_2) = \frac{P(\omega_1)P(o_2|\omega_1)}{P(o_2)} \tag{S.1.23}$$

であり，$P(\omega_i) = p_i$ であるので，これらを代入し，司会者が外れの扉 D_2 を開けてみせたとき，ゲストの選んだ扉 D_1 が当たりである確率 $P(\omega_1|o_2)$ は

$$P(\omega_1|o_2) = \frac{(n-2)\,p_1}{(n-2)\,p_1 + (n-1)\sum_{i=3}^{n}\,p_i} \tag{S.1.24}$$

と求められる．

上式から明らかなように，一般に $P(\omega_1|o_2) \neq P(\omega_1)$ である．ここで

$$p_i = \frac{1}{n} \qquad (1 = 1, 2, \ldots, n) \tag{S.1.25}$$

とすると，

$$P(\omega_1|o_2) = \frac{(n-2)/n}{(n-2)/n + (n-1)(n-2)/n} \tag{S.1.26}$$

$$= \frac{n-2}{n-2+(n-1)(n-2)} \tag{S.1.27}$$

$$= \frac{1}{n} \tag{S.1.28}$$

が確かめられる．すなわち，当たりである確率（事前確率）が n 枚の扉で等しいときは，司会者が外れの扉を開けて見せても，ゲストの選んだ扉が当たりである確率は $1/n$ のまま変わらない．しかし，事前確率が等しいことは確率が変わらないための十分条件ではあるが，必要条件ではない．

このことは $n = 3$ として，「三つの扉問題」で確かめることができる．このとき，式 (S.1.24) は

$$P(\omega_1|o_2) = \frac{p_1}{p_1 + 2p_3} \tag{S.1.29}$$

となる．したがって，司会者が外れの扉を開けて見せても，ゲストの選んだ扉が当たりである確率が変わらないためには

$$p_1 = \frac{p_1}{p_1 + 2p_3} \tag{S.1.30}$$

が成り立つ必要があり，これより

$$p_1 + 2p_3 = 1 \tag{S.1.31}$$

となる．この解としては，$p_1 = 1/2$, $p_2 = p_3 = 1/4$ もその解のひとつでり，上で述べた $p_1 = p_2 = p_3 = 1/3$ に限らないことがわかる．

問 (2)

司会者が扉 D_k を開けずに $(n-2)$ 枚の外れの扉を開けて見せることを $X = \bar{o_k}$ で表すことにする．上と同様，場合分けを行う．

もし，ゲストの選んだ扉 D_1 が当たりの場合，司会者は $(n-1)$ 枚の扉 D_2〜D_n の何れかから，開けない扉を無作為に一つ選び，残り $(n-2)$ 枚の扉を開けるはずである．したがって，扉 D_1 が当たりという条件で，司会者が扉 D_n を残して他の $(n-2)$ 枚の扉を開ける確率 $P(\bar{o_n}|\omega_1)$ は，下式で表される．

$$P(\bar{o_n}|\omega_1) = \frac{1}{n-1} \tag{S.1.32}$$

もし，扉 D_2〜D_n のいずれかの扉が当たりの場合，例えば D_k $(2 \leq k \leq n)$ が当たりの場合，司会者は扉 D_1 と D_k 以外の $(n-2)$ 枚の扉を開けるしか選択肢はない．すなわち

$$P(\bar{o_n}|\omega_n) = 1 \tag{S.1.33}$$

$$P(\bar{o_n}|\omega_i) = 0 \qquad (i \neq n) \tag{S.1.34}$$

である．これらより，司会者が扉 D_n のみ残し，外れの扉として $(n-2)$ 枚の扉 D_2〜D_{n-1} を開けてみせる確率 $P(\bar{o_n})$ は

$$P(\bar{o_n}) = \sum_{i=1}^{n} P(\omega_i)P(\bar{o_n}|\omega_i) \tag{S.1.35}$$

$$= P(\omega_1)P(\bar{o_n}|\omega_1) + P(\omega_n)P(\bar{o_n}|\omega_n) \tag{S.1.36}$$

$$= \frac{1}{n-1}P(\omega_1) + P(\omega_n) \tag{S.1.37}$$

となる．ベイズの定理は，

$$P(\omega_1|\bar{o_n}) = \frac{P(\omega_1)P(\bar{o_n}|\omega_1)}{P(\bar{o_n})} \tag{S.1.38}$$

であり，上記結果を代入すると，司会者が扉 D_n のみ残し，外れの扉として $(n-2)$ 枚の扉 D_2〜D_{n-1} を開けてみせたとき，ゲストの選んだ扉 D_1 が当たりである確率 $P(\omega_1|\bar{o_n})$ は

$$P(\omega_1|\bar{o_n}) = \frac{p_1/(n-1)}{p_1/(n-1)+p_n} \tag{S.1.39}$$

$$= \frac{p_1}{p_1+(n-1)\,p_n} \tag{S.1.40}$$

と求められる．ここで式 (S.1.25) が成り立つ場合は,

$$P(\omega_1|\bar{o_n}) = \frac{1}{n} \tag{S.1.41}$$

となる．また，上記で $n=3$ とすれば「三つの扉問題」となる．

第 2 章

【演習問題 2.1】

式 (2.26) を以下に再掲する．

$$P_n(r) = {}_n\mathrm{C}_r\,\theta^r(1-\theta)^{n-r} \tag{S.2.1}$$

ここで，r は $0 \leq r \leq n$ なる整数であるので，$P_n(r)$ の差分をとって

$$P_n(r+1) - P_n(r)$$

$$= P_n(r)\left(\frac{P_n(r+1)}{P_n(r)} - 1\right) \tag{S.2.2}$$

$$= P_n(r)\left(\frac{{}_n\mathrm{C}_{r+1}\theta^{r+1}(1-\theta)^{n-r-1}}{{}_n\mathrm{C}_r\theta^r(1-\theta)^{n-r}} - 1\right) \tag{S.2.3}$$

$$= P_n(r)\left(\frac{n-r}{r+1}\cdot\frac{\theta}{1-\theta} - 1\right) \tag{S.2.4}$$

を得る．上式は r の単調減少関数であり，符号が正から負に変化するのは r が

$$r > (n+1)\theta - 1 \tag{S.2.5}$$

を満たす最小の整数のときであり，このとき $P_n(r)$ が最大となる．

ここで，$n=10$, $\theta=0.8$ とすると，式 (S.2.5) より，

$$r > (10+1)\times 0.8 - 1 = 7.8 \tag{S.2.6}$$

が得られ，$r=8$ のとき $P_n(r)$ が最大となることがわかる．

【演習問題 2.2】

コインを n 回投げたときの観察結果 $\mathbf{x}^{(n)}$ に関しては，ベイズの定理

$$
\begin{aligned}
P(\omega_i|\mathbf{x}^{(n)}) &= \frac{P(\mathbf{x}^{(n)}|\omega_i)}{P(\mathbf{x}^{(n)})}\cdot P(\omega_i) && \text{(S.2.7)}\\
&= \frac{P(\mathbf{x}^{(n)}|\omega_i)}{\sum_{j=1}^{3} P(\omega_j)\,P(\mathbf{x}^{(n)}|\omega_j)}\cdot P(\omega_i) \quad (i=1,2,3) && \text{(S.2.8)}
\end{aligned}
$$

が成り立つ．各試行の独立性

$$
P(\mathbf{x}^{(n)}|\omega_i) = P(x_n|\omega_i)\cdot P(\mathbf{x}^{(n-1)}|\omega_i) \qquad (i=1,2,3) \tag{S.2.9}
$$

が成り立つので

$$
\begin{aligned}
&P(\omega_i|\mathbf{x}^{(n)})\\
&= \frac{P(x_n|\omega_i)\,P(\mathbf{x}^{(n-1)}|\omega_i)\,P(\omega_i)}{\sum_{j=1}^{3} P(\omega_j)\,P(x_n|\omega_j)\,P(\mathbf{x}^{(n-1)}|\omega_j)} && \text{(S.2.10)}\\
&= \frac{P(x_n|\omega_i)}{\sum_{j=1}^{3} P(\omega_j|\mathbf{x}^{(n-1)})\,P(x_n|\omega_j)}\cdot P(\omega_i|\mathbf{x}^{(n-1)}) \quad (i=1,2,3) && \text{(S.2.11)}
\end{aligned}
$$

が得られる．ここで，式 (S.2.10) から式 (S.2.11) への導出には，式 (S.2.10) の分母，分子を $P(\mathbf{x}^{(n-1)})$ で割り，さらにベイズの定理

$$
P(\omega_i|\mathbf{x}^{(n-1)}) = \frac{P(\mathbf{x}^{(n-1)}|\omega_i)}{P(\mathbf{x}^{(n-1)})}\cdot P(\omega_i) \tag{S.2.12}
$$

を適用した．

第 3 章

【演習問題 3.1】

問 (1)

取り出したコインが $\omega_1,\omega_2,\omega_3$ である確率をそれぞれ $P(\omega_1)$，$P(\omega_2)$，$P(\omega_3)$ とすると，

$$
P(\omega_1)=\frac{1}{2},\quad P(\omega_2)=\frac{1}{3},\quad P(\omega_3)=\frac{1}{6} \tag{S.3.1}
$$

である．コインを投げて表，裏が出ることをそれぞれ H，T で表すと

$$P(H|\omega_1) = \frac{1}{4}, \quad P(H|\omega_2) = \frac{1}{2}, \quad P(H|\omega_3) = \frac{2}{3} \tag{S.3.2}$$

である．上式より

$$P(T|\omega_1) = \frac{3}{4}, \quad P(T|\omega_2) = \frac{1}{2}, \quad P(T|\omega_3) = \frac{1}{3} \tag{S.3.3}$$

となる．コインを投げて表の出る確率 $P(H)$ は

$$P(H) = \sum_{i=1}^{3} P(\omega_i)P(H|\omega_i) = \frac{1}{2}\cdot\frac{1}{4} + \frac{1}{3}\cdot\frac{1}{2} + \frac{1}{6}\cdot\frac{2}{3} = \frac{29}{72} \tag{S.3.4}$$

と求められるので，コインを投げて裏の出る確率 $P(T)$ は下式で表される．

$$P(T) = 1 - P(H) = \frac{43}{72} \tag{S.3.5}$$

　まず，コインを投げて表が出た場合を考える．ベイズの定理より

$$P(\omega_1|H) = \frac{P(H|\omega_1)P(\omega_1)}{P(H)} = \left(\frac{1}{4}\cdot\frac{1}{2}\right)\bigg/\frac{29}{72} = \frac{9}{29} \tag{S.3.6}$$

が得られ，同様にして

$$P(\omega_2|H) = \frac{12}{29}, \qquad P(\omega_3|H) = \frac{8}{29} \tag{S.3.7}$$

が得られる．これらの計算より

$$\operatorname*{argmax}_{\omega_i} \{P(\omega_i|H)\} = \omega_2 \tag{S.3.8}$$

であるので，ベイズ決定則に従えば，表が出た場合は ω_2 と判定される．

　次に，コインを投げて裏が出た場合を考える．上と同様にベイズの定理より

$$P(\omega_1|T) = \frac{27}{43}, \qquad P(\omega_2|T) = \frac{12}{43}, \qquad P(\omega_3|T) = \frac{4}{43} \tag{S.3.9}$$

が得られる．これらの計算より

$$\operatorname*{argmax}_{\omega_i} \{P(\omega_i|T)\} = \omega_1 \tag{S.3.10}$$

であるので，ベイズ決定則に従えば，裏が出た場合は ω_1 と判定される．

問 (2)

表が出て ω_2 と判定するときの誤り確率 e_H は

$$e_H = 1 - \frac{12}{29} = \frac{17}{29} \tag{S.3.11}$$

である．また，裏が出て ω_1 と判定するときの誤り確率 e_T は

$$e_T = 1 - \frac{27}{43} = \frac{16}{43} \tag{S.3.12}$$

である．これらより，求めるべきベイズ誤り確率 e_B は下式で表される．

$$\begin{align} e_B &= P(H) \cdot e_H + P(T) \cdot e_T = \frac{29}{72} \cdot \frac{17}{29} + \frac{43}{72} \cdot \frac{16}{43} \tag{S.3.13} \\ &= \frac{11}{24} \tag{S.3.14} \end{align}$$

このベイズ誤り確率は次のようにして求めることもできる．

$$\begin{align} e_B &= (\text{取り出したコインが } \omega_1 \text{ で，かつ表が出る確率}) \\ &\quad + (\text{取り出したコインが } \omega_2 \text{ で，かつ裏が出る確率}) \\ &\quad + (\text{取り出したコインが } \omega_3 \text{ である確率}) \tag{S.3.15} \\ &= P(\omega_1)P(H|\omega_1) + P(\omega_2)P(T|\omega_2) + P(\omega_3) \tag{S.3.16} \\ &= \frac{1}{2} \cdot \frac{1}{4} + \frac{1}{3} \cdot \frac{1}{2} + \frac{1}{6} \tag{S.3.17} \\ &= \frac{11}{24} \tag{S.3.18} \end{align}$$

【演習問題 3.2】

取り出したコインが ω_1 であったとする．式 (2.25) の $\omega_i = \omega_1$ とし，右辺の分母，分子を $\pi_1\, {\theta_1}^r (1-\theta_1)^{n-r}$ で割ると

$$\begin{align} & P(\omega_1|\mathbf{x}^{(n)}) \\ &= \frac{\pi_1}{\pi_1 + \pi_2 \left(\dfrac{\theta_2}{\theta_1}\right)^r \left(\dfrac{1-\theta_2}{1-\theta_1}\right)^{n-r} + \pi_3 \left(\dfrac{\theta_3}{\theta_1}\right)^r \left(\dfrac{1-\theta_3}{1-\theta_1}\right)^{n-r}} \tag{S.3.19} \\ &= \frac{\pi_1}{\pi_1 + \pi_2 f_2 + \pi_3 f_3} \tag{S.3.20} \end{align}$$

が得られる．ただし

$$f_2 \stackrel{\mathrm{def}}{=} \left(\frac{\theta_2}{\theta_1}\right)^r \left(\frac{1-\theta_2}{1-\theta_1}\right)^{n-r} \tag{S.3.21}$$

$$f_3 \stackrel{\mathrm{def}}{=} \left(\frac{\theta_3}{\theta_1}\right)^r \left(\frac{1-\theta_3}{1-\theta_1}\right)^{n-r} \tag{S.3.22}$$

と定義した．ここで，n が十分大きいときは近似的に $r = n\theta_1$ が成り立つので

$$f_2 = \left(\frac{\theta_2}{\theta_1}\right)^{n\theta_1} \left(\frac{1-\theta_2}{1-\theta_1}\right)^{n(1-\theta_1)} \tag{S.3.23}$$

$$= \left[\left(\frac{\theta_2}{\theta_1}\right)^{\theta_1} \left(\frac{1-\theta_2}{1-\theta_1}\right)^{1-\theta_1}\right]^n \tag{S.3.24}$$

となる．ここで

$$g_2 \stackrel{\mathrm{def}}{=} \left(\frac{\theta_2}{\theta_1}\right)^{\theta_1} \left(\frac{1-\theta_2}{1-\theta_1}\right)^{1-\theta_1} \tag{S.3.25}$$

とすると

$$\begin{aligned}&\log g_2\\ &= \theta_1(\log\theta_2 - \log\theta_1) + (1-\theta_1)\left(\log(1-\theta_2) - \log(1-\theta_1)\right)\end{aligned} \tag{S.3.26}$$

となる．上式において θ_1 を固定し θ_2 で微分すると

$$\frac{d\log g_2}{d\theta_2} = \frac{\theta_1 - \theta_2}{\theta_2(1-\theta_2)} \tag{S.3.27}$$

が得られ，$\theta_2(1-\theta_2) > 0$ であるので $\log g_2$ は $\theta_2 = \theta_1$ で最大となることがわかる．このとき g_2 も最大となり，このときの最大値は式 (S.3.25) より $g_2 = 1$ と求められる．ここで $\theta_2 \neq \theta_1$ であるので

$$g_2 = \left(\frac{\theta_2}{\theta_1}\right)^{\theta_1} \left(\frac{1-\theta_2}{1-\theta_1}\right)^{1-\theta_1} < 1 \tag{S.3.28}$$

が成り立つ．したがって

$$\lim_{n\to\infty} f_2 = \lim_{n\to\infty} {g_2}^n = 0 \tag{S.3.29}$$

となる．同様にして

$$\lim_{n\to\infty} f_3 = 0 \tag{S.3.30}$$

となるので，式 (S.3.20) より

$$\lim_{n\to\infty} P(\omega_1|\mathbf{x}^{(n)}) = \frac{\pi_1}{\pi_1} = 1 \tag{S.3.31}$$

が得られる．以上の議論は，取り出したコインが ω_2，ω_3 であったときの $P(\omega_2|\mathbf{x}^{(n)})$，$P(\omega_3|\mathbf{x}^{(n)})$ に対してもそれぞれ成り立つ．したがって，n が十分大きい場合には，事後確率は 1 に近づき，事前確率 π_i の影響は受けないことがわかる．

第 4 章

【演習問題 4.1】

式 (4.68) において

$$n_1 = n_2 = \cdots = n_m = 0$$

とすることにより

$$\int_D d\boldsymbol{\theta} = \frac{\Gamma(1)\cdots\Gamma(1)}{\Gamma(m)} = \frac{1}{(m-1)!} \tag{S.4.1}$$

が得られ，これと式 (4.66)，(4.61) より，式 (4.66) の一様分布 $p(\boldsymbol{\theta})$ は

$$p(\boldsymbol{\theta}) = (m-1)! \tag{S.4.2}$$

と求めることができる．

【演習問題 4.2】

いま，m 個の変数[*12] $\theta_1, \theta_2, \ldots, \theta_m$ が

$$\sum_{k=1}^{m} \theta_k = 1, \qquad 0 \le \theta_k \le 1 \tag{S.4.3}$$

[*12] もともとはパラメータであるが，ここでは変数として扱う

を満たす領域 D_m にあるとき

$$I_m \stackrel{\text{def}}{=} \int_{D_m} {\theta_1}^{\alpha_1 - 1} \cdots {\theta_m}^{\alpha_m - 1}\, d\boldsymbol{\theta} = \frac{\Gamma(\alpha_1)\cdots\Gamma(\alpha_m)}{\Gamma(\alpha_1+,\ldots,+\alpha_m)} \qquad \text{(S.4.4)}$$

となることを数学的帰納法により証明する．その際，式 (4.22) で紹介した次の関係式を用いる．

$$\int_0^1 u^{\alpha-1}\,(1-u)^{\beta-1}\,du = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)} \qquad (\alpha > 0,\ \beta > 0) \qquad \text{(S.4.5)}$$

まず，$m = 2$ のとき，式 (S.4.4) の積分 I_2 は

$$\begin{aligned} I_2 &= \int_{D_2} {\theta_1}^{\alpha_1-1} \cdot {\theta_2}^{\alpha_2-1}\, d\theta_1 d\theta_2 \\ &= \int_0^1 {\theta_1}^{\alpha_1-1} \cdot (1-\theta_1)^{\alpha_2-1}\, d\theta_1 && \text{(S.4.6)} \\ &= \frac{\Gamma(\alpha_1)\Gamma(\alpha_2)}{\Gamma(\alpha_1+\alpha_2)} && \text{(S.4.7)} \end{aligned}$$

となり，確かに式 (S.4.4) が成り立っている．ただし，式 (S.4.6) から式 (S.4.7) への変形には式 (S.4.5) を用いた．

次に $(m-1)$ 変数に対して式 (S.4.4) が成り立つと仮定する．すなわち

$$\sum_{k=1}^{m-1} \theta_k = 1, \qquad 0 \le \theta_k \le 1 \qquad \text{(S.4.8)}$$

を満たす領域 D_{m-1} に対して，

$$\begin{aligned} I_{m-1} &= \int_{D_{m-1}} {\theta_1}^{\alpha_1-1} \cdots {\theta_{m-1}}^{\alpha_{m-1}-1}\, d\theta_1 \cdots d\theta_{m-1} \\ &= \frac{\Gamma(\alpha_1)\cdots\Gamma(\alpha_{m-1})}{\Gamma(\alpha_1+,\ldots,+\alpha_{m-1})} && \text{(S.4.9)} \end{aligned}$$

が成り立つと仮定する．式 (S.4.8) より

$$\theta_{m-1} = 1 - \sum_{k=1}^{m-2} \theta_k \qquad \text{(S.4.10)}$$

であるので，式 (S.4.9) は

$$\int_{S_{m-2}} {\theta_1}^{\alpha_1-1}\cdots{\theta_{m-2}}^{\alpha_{m-2}-1}\left(1-\sum_{k=1}^{m-2}\theta_k\right)^{\alpha_{m-1}-1} d\theta_1\cdots d\theta_{m-2}$$
$$=\frac{\Gamma(\alpha_1)\cdots\Gamma(\alpha_{m-1})}{\Gamma(\alpha_1+,\ldots,+\alpha_{m-1})} \tag{S.4.11}$$

と書ける．ただし，S_{m-2} は

$$0\leq\sum_{k=1}^{m-2}\theta_k\leq 1,\qquad 0\leq\theta_k\leq 1 \tag{S.4.12}$$

を満たす領域である．

同様にして m 変数のときは式 (S.4.3) より

$$\theta_m=1-\sum_{k=1}^{m-1}\theta_k \tag{S.4.13}$$

であるので，式 (S.4.4) の積分は

$$I_m=\int_{S_{m-1}} {\theta_1}^{\alpha_1-1}\cdots{\theta_{m-1}}^{\alpha_{m-1}-1}$$
$$\cdot\left(1-\sum_{k=1}^{m-1}\theta_k\right)^{\alpha_m-1} d\theta_1\cdots d\theta_{m-1} \tag{S.4.14}$$

となる．ここで次式のような $\theta_{m-1}\to u$ の変数変換を行う．

$$\theta_{m-1}=u\left(1-\sum_{k=1}^{m-2}\theta_k\right) \tag{S.4.15}$$

明らかに u は 0 から 1 の範囲にある．上式の変数変換により

$$\begin{aligned}1-\sum_{k=1}^{m-1}\theta_k&=1-\sum_{k=1}^{m-2}\theta_k-\theta_{m-1}\\&=1-\sum_{k=1}^{m-2}-u\left(1-\sum_{k=1}^{m-2}\theta_k\right)\\&=(1-u)\left(1-\sum_{k=1}^{m-2}\theta_k\right)\end{aligned} \tag{S.4.16}$$

となり

$$\frac{d\theta_{m-1}}{du} = 1 - \sum_{k=1}^{m-2} \theta_k \tag{S.4.17}$$

であるので，これらを用いると式 (S.4.14) は

$$\begin{aligned} I_m &= \int_{S_{m-2}} {\theta_1}^{\alpha_1 - 1} \cdots {\theta_{m-2}}^{\alpha_{m-2}-1} \left(1 - \sum_{k=1}^{m-2} \theta_k\right)^{\alpha_{m-1}+\alpha_m-1} d\theta_1 \cdots d\theta_{m-2} \\ &\quad \cdot \int_0^1 u^{\alpha_{m-1}-1} \cdot (1-u)^{\alpha_m - 1}\, du \qquad \text{(S.4.18)} \\ &= I_a \cdot I_b \qquad \text{(S.4.19)} \end{aligned}$$

となる．ただし

$$\begin{aligned} I_a &\stackrel{\text{def}}{=} \int_{S_{m-2}} {\theta_1}^{\alpha_1 - 1} \cdots {\theta_{m-2}}^{\alpha_{m-2}-1} \\ &\qquad \cdot \left(1 - \sum_{k=1}^{m-2} \theta_k\right)^{\alpha_{m-1}+\alpha_m-1} d\theta_1 \cdots d\theta_{m-2} \qquad \text{(S.4.20)} \\ I_b &\stackrel{\text{def}}{=} \int_0^1 u^{\alpha_{m-1}-1} \cdot (1-u)^{\alpha_m - 1}\, du \qquad \text{(S.4.21)} \end{aligned}$$

である．式 (S.4.20) は，式 (S.4.11) の α_{m-1} を $\alpha_{m-1} + \alpha_m$ に置き換えることにより得られることに注意すると

$$I_a = \frac{\Gamma(\alpha_1)\cdots\Gamma(\alpha_{m-2})\Gamma(\alpha_{m-1}+\alpha_m)}{\Gamma(\alpha_1+,\ldots,+\alpha_{m-2}+\alpha_{m-1}+\alpha_m)} \tag{S.4.22}$$

となる．また，式 (S.4.21) は式 (S.4.5) より

$$I_b = \frac{\Gamma(\alpha_{m-1})\Gamma(\alpha_m)}{\Gamma(\alpha_{m-1}+\alpha_m)} \tag{S.4.23}$$

である．これらを式 (S.4.19) に代入することにより

$$\begin{aligned} I_m &= \frac{\Gamma(\alpha_1)\cdots\Gamma(\alpha_{m-2})\Gamma(\alpha_{m-1}+\alpha_m)\Gamma(\alpha_{m-1})\Gamma(\alpha_m)}{\Gamma(\alpha_1+,\ldots,+\alpha_m)\Gamma(\alpha_{m-1}+\alpha_m)} \\ &= \frac{\Gamma(\alpha_1)\cdots\Gamma(\alpha_m)}{\Gamma(\alpha_1+,\ldots,+\alpha_m)} \qquad \text{(S.4.24)} \end{aligned}$$

となり，$m \geq 2$ なるすべての m に対して式 (S.4.4) が成り立つがことが証明された.

【演習問題 4.3】

ベクトル表記した期待値 $\mathrm{E}[\boldsymbol{\theta}]$ の k 番目の要素 $\mathrm{E}[\theta_k]$ は

$$
\begin{aligned}
\mathrm{E}[\theta_k] &= \int_D \theta_k \cdot \mathrm{Dir}(\alpha_1, \ldots, \alpha_m)\, d\boldsymbol{\theta} && \text{(S.4.25)}\\
&= \int_D \theta_k \cdot \frac{\Gamma(\alpha)}{\prod_{j=1}^m \Gamma(\alpha_j)} \prod_{j=1}^m {\theta_j}^{\alpha_j - 1}\, d\boldsymbol{\theta} && \text{(S.4.26)}\\
&= \frac{\Gamma(\alpha)}{\prod_{j=1}^m \Gamma(\alpha_j)} \int_D {\theta_1}^{\alpha_1 - 1} \cdots {\theta_k}^{\alpha_k} \cdots {\theta_m}^{\alpha_m - 1} d\boldsymbol{\theta} && \text{(S.4.27)}\\
&= \frac{\Gamma(\alpha)}{\prod_{j=1}^m \Gamma(\alpha_j)} \cdot \frac{\Gamma(\alpha_1) \cdots \Gamma(\alpha_k + 1) \cdots \Gamma(\alpha_m)}{\Gamma(\alpha + 1)} && \text{(S.4.28)}\\
&= \frac{\Gamma(\alpha)\, \Gamma(\alpha_k + 1)}{\Gamma(\alpha + 1)\, \Gamma(\alpha_k)} && \text{(S.4.29)}\\
&= \frac{(\alpha - 1)!\, \alpha_k!}{\alpha!\, (\alpha_k - 1)!} && \text{(S.4.30)}\\
&= \frac{\alpha_k}{\alpha} \qquad (k = 1, 2, \ldots, m) && \text{(S.4.31)}
\end{aligned}
$$

と求められ，式 (4.76) が得られる．なお，式 (S.4.27) から式 (S.4.28) への変形には式 (4.68) を用いた.

分散 $\mathrm{V}[\boldsymbol{\theta}]$ の k 番目の要素 $\mathrm{V}[\theta_k]$ は

$$
\begin{aligned}
\mathrm{V}[\theta_k] &= \mathrm{E}\left[\left(\theta_k - \mathrm{E}[\theta_k]\right)^2\right]\\
&= \mathrm{E}\left[\theta_k^2 - 2\theta_k \cdot E[\theta_k] + E[\theta_k]^2\right]\\
&= \mathrm{E}[\theta_k^2] - \mathrm{E}[\theta_k]^2 && \text{(S.4.32)}
\end{aligned}
$$

として求められる．上式の $\mathrm{E}[\theta_k^2]$ は，$\mathrm{E}[\theta_k]$ を求めたときと同様にして

$$\begin{aligned}
\mathrm{E}[\theta_k^2] &= \int_D \theta_k^2 \cdot \mathrm{Dir}(\alpha_1, \ldots, \alpha_m)\, d\boldsymbol{\theta} \\
&= \int_D \theta_k^2 \cdot \frac{\Gamma(\alpha)}{\prod_{j=1}^m \Gamma(\alpha_j)} \prod_{j=1}^m {\theta_j}^{\alpha_j - 1}\, d\boldsymbol{\theta} \\
&= \frac{\Gamma(\alpha)}{\prod_{j=1}^m \Gamma(\alpha_j)} \int_D {\theta_1}^{\alpha_1-1} \cdots {\theta_k}^{\alpha_k+1} \cdots {\theta_m}^{\alpha_m-1} d\boldsymbol{\theta} \\
&= \frac{\Gamma(\alpha)}{\prod_{j=1}^m \Gamma(\alpha_j)} \cdot \frac{\Gamma(\alpha_1)\cdots\Gamma(\alpha_k+2)\cdots\Gamma(\alpha_m)}{\Gamma(\alpha+2)} \\
&= \frac{\Gamma(\alpha)\,\Gamma(\alpha_k+2)}{\Gamma(\alpha+2)\,\Gamma(\alpha_k)} \\
&= \frac{(\alpha-1)!\,(\alpha_k+1)!}{(\alpha+1)!\,(\alpha_k-1)!} \\
&= \frac{(\alpha_k+1)\alpha_k}{(\alpha+1)\alpha} \qquad \text{(S.4.33)}
\end{aligned}$$

と求めることができる．上式と式 (S.4.31) を式 (S.4.32) に代入することにより，

$$\begin{aligned}
\mathrm{V}[\theta_k] &= \frac{(\alpha_k+1)\alpha_k}{(\alpha+1)\alpha} - \left(\frac{\alpha_k}{\alpha}\right)^2 \\
&= \frac{\alpha_k(\alpha-\alpha_k)}{\alpha^2(\alpha+1)} \qquad \text{(S.4.34)}
\end{aligned}$$

が得られる．

モード $\mathrm{M}[\boldsymbol{\theta}]$ の k 番目の要素 $\mathrm{M}[\theta_k]$ は，ディリクレ分布の対数をとり

$$\mathrm{M}[\theta_k] = \underset{\theta_k}{\mathrm{argmax}} \{\log \mathrm{Dir}(\alpha_1, \ldots, \alpha_m)\} \qquad \text{(S.4.35)}$$

$$= \underset{\theta_k}{\mathrm{argmax}} \left\{ \sum_{j=1}^m (\alpha_j - 1) \log \theta_j \right\} \qquad \text{(S.4.36)}$$

として求められる．上式を解くには

$$\sum_{j=1}^m \theta_j = 1 \qquad \text{(S.4.37)}$$

の条件下で

$$\sum_{j=1}^{m}(\alpha_j - 1)\log\theta_j \tag{S.4.38}$$

を最大にする θ_k を求めればよい．ここで，定理 5.1 が適用でき，求めるべき $\mathrm{M}[\theta_k]$ は

$$\mathrm{M}[\theta_k] = \frac{\alpha_k - 1}{\displaystyle\sum_{j=1}^{m}(\alpha_j - 1)} \tag{S.4.39}$$

$$= \frac{\alpha_k - 1}{\alpha - m} \qquad (k = 1, 2, \ldots, m) \tag{S.4.40}$$

となり，式 (4.78) が成り立つことが確かめられる．ここで，式 (4.71) を用いた．さらに式 (4.72)，(4.75) を用いると，上式は

$$\mathrm{M}[\theta_k] = \frac{n_k}{n} \qquad (k = 1, 2, \ldots, m) \tag{S.4.41}$$

となることが確かめられる．

ベータ分布の期待値，分散，モードは，上記において $m \to 2$，$\alpha \to \alpha + \beta$，$\alpha_k \to \alpha$ と置き換えることにより導出できる．

第 5 章

【演習問題 5.1】

証明には付録 A.1 の結果を用いる．まず，この問題が凸計画問題であることを示す．そのためには，実行可能領域が凸集合で，目的関数が凸関数であることを示せばよい．

まず，実行可能領域が凸集合であることを示す．拘束条件の式 (5.22) を満たす点の集合，すなわち実行可能領域を $S \subset \mathbf{R}^n$ とする．いま，S 内の任意の 2 点

$$\left.\begin{aligned} \mathbf{x}_1 &= (x_{11}, x_{12}, \ldots, x_{1n})^t \ \in S \\ \mathbf{x}_2 &= (x_{21}, x_{22}, \ldots, x_{2n})^t \ \in S \end{aligned}\right\} \tag{S.5.1}$$

を考える．これらは拘束条件 (5.22) を満たすので

$$\sum_{i=1}^{n} x_{1i} = 1, \qquad \sum_{i=1}^{n} x_{2i} = 1 \tag{S.5.2}$$

が成り立つ．ここで，$0 \leq \lambda \leq 1$ を満たす λ を用いて

$$\begin{aligned} \mathbf{x} &= (x_1, x_2, \ldots, x_n)^t && \text{(S.5.3)} \\ &\stackrel{\text{def}}{=} \lambda \mathbf{x}_1 + (1-\lambda)\mathbf{x}_2 && \text{(S.5.4)} \end{aligned}$$

なる $\mathbf{x}$ を定義する．式 (S.5.4) の両辺の第 i 成分を比較することにより

$$x_i = \lambda x_{1i} + (1-\lambda) x_{2i} \qquad \text{(S.5.5)}$$

となるので

$$\begin{aligned} \sum_{i=1}^{n} x_i &= \lambda \sum_{i=1}^{n} x_{1i} + (1-\lambda) \sum_{i=1}^{n} x_{2i} && \text{(S.5.6)} \\ &= \lambda \cdot 1 + (1-\lambda) \cdot 1 && \text{(S.5.7)} \\ &= 1 && \text{(S.5.8)} \end{aligned}$$

が得られ，$\mathbf{x}$ も拘束条件 (5.22) を満たすことがわかる．したがって，$\mathbf{x} \in S$ であり，S は凸集合である．

次に関数 $-f(x_1, x_2, \ldots, x_n)$ が狭義凸関数であることを示す．式 (5.23) において

$$f_i(x_1, x_2, \ldots, x_n) = \log x_i \qquad \text{(S.5.9)}$$

と置くと

$$f(x_1, x_2, \ldots, x_n) = \sum_{i=1}^{n} w_i \, f_i(x_1, x_2, \ldots, x_n) \qquad \text{(S.5.10)}$$

と書ける．関数 $-\log x_i$ が狭義凸関数であることと，定理 A.2 より，関数 $-f(x_1, x_2, \ldots, x_n)$ が狭義凸関数であることは明らかである．したがって，本問題は凸計画問題であり，かつ関数 $-f_i(x_1, x_2, \ldots, x_n)$ が狭義凸関数であるので，定理 A.4 より，式 (5.31) は大域的最適解であり，唯一の最大点である．

【演習問題 5.2】

式 (5.1) で定まる実行可能領域が凸集合であることは，**演習問題 5.1** の解答で示したとおりである．式 (5.51) に式 (5.8) を代入し，対数尤度 $\log P(\mathbf{x})$ を π_i の

関数として表すと

$$f(\pi_1,\ldots,\pi_c) = \log P(\mathbf{x}) = \sum_{k=1}^{m} r_k \log\left(\sum_{l=1}^{c} \pi_l \cdot \theta_{lk}\right) \tag{S.5.11}$$

と書ける．以下では，$-f(\pi_1,\ldots,\pi_c)$ が凸関数であることを示す．

関数 $-f(\pi_1,\ldots,\pi_c)$ のヘッセ行列を $\mathbf{H}$ とし，その (i,j) 成分を h_{ij} とすると

$$\begin{aligned} h_{ij} &= -\frac{\partial^2 f}{\partial \pi_i \partial \pi_j} \\ &= \sum_{k=1}^{m} r_k \cdot \frac{\theta_{ik}\cdot\theta_{jk}}{\left(\displaystyle\sum_{l=1}^{c} \pi_l \cdot \theta_{lk}\right)^2} \\ &= \sum_{k=1}^{m} a_k^2 \cdot \theta_{ik} \cdot \theta_{jk} \end{aligned} \tag{S.5.12}$$

となる．ただし

$$a_k = \frac{{r_k}^{1/2}}{\displaystyle\sum_{l=1}^{c} \pi_l \cdot \theta_{lk}} \qquad (>0) \tag{S.5.13}$$

と置いた．ここでベクトル $\mathbf{u}_k$ を

$$\mathbf{u}_k \stackrel{\text{def}}{=} a_k\,(\theta_{1k},\ldots,\theta_{ck})^t \qquad (k=1,\ldots,m) \tag{S.5.14}$$

と定義すると，$-f(\pi_1,\ldots,\pi_c)$ のヘッセ行列 $\mathbf{H}$ は

$$\mathbf{H} = \sum_{k=1}^{m} \mathbf{u}_k \cdot \mathbf{u}_k^t \tag{S.5.15}$$

で表される．ここで，$\mathbf{0}$ でない任意の c 次元列ベクトル $\mathbf{y}$ に対して以下が成り

立つ.

$$\mathbf{y}^t \mathbf{H} \mathbf{y} = \mathbf{y}^t \left(\sum_{k=1}^{m} \mathbf{u}_k \cdot \mathbf{u}_k^t \right) \mathbf{y} \tag{S.5.16}$$

$$= \sum_{k=1}^{m} \mathbf{y}^t \mathbf{u}_k \cdot \mathbf{u}_k^t \mathbf{y} \tag{S.5.17}$$

$$= \sum_{k=1}^{m} \left(\mathbf{u}_k^t \mathbf{y} \right)^2 > 0 \tag{S.5.18}$$

上式より行列 $\mathbf{H}$ は正定値であるので，定理 A.1 より，関数 $-f(\pi_1, \ldots, \pi_c)$ は凸関数であることがわかる．実行可能領域が凸集合で，目的関数 $-f(\pi_1, \ldots, \pi_c)$ が凸関数であるので，本問題は凸計画問題である．したがって，定理 A.4 より，繰り返し演算によって得られる解は大域的最適解である．

【演習問題 5.3】

奇数の目，偶数の目をそれぞれ v_1，v_2 とし，それらが観測された回数をそれぞれ r_1，r_2 とすると対数尤度の式 (5.51) は

$$\log P(\mathbf{x}) = r_1 \log P(v_1) + r_2 \log P(v_2) \tag{S.5.19}$$

$$= r_1 \log P(v_1) + r_2 \log (1 - P(v_1)) \tag{S.5.20}$$

と書ける．ここで $P(v_1)$ は式 (5.8) より

$$P(v_1) = \pi_1 \theta_{11} + \pi_2 \theta_{21} + \pi_3 \theta_{31} \tag{S.5.21}$$

である．式 (S.5.20) から明らかなように, $\log P(\mathbf{x})$ の等高線 $\log P(\mathbf{x}) = const$ は $P(v_1) = const$ と等価であるので，v_c を定数とし，(S.5.21) において $P(v_1) = v_c$ と置くことにより，等高線は

$$\pi_1 \theta_{11} + \pi_2 \theta_{21} + \pi_3 \theta_{31} = v_c \tag{S.5.22}$$

を満たさなくてはならない．ここで，$\theta_{11}, \theta_{21}, \theta_{31}$ は既知の定数として扱っている．一方，パラメータ π_1, π_2, π_3 に関しては

$$\pi_1 + \pi_2 + \pi_3 = 1 \tag{S.5.23}$$

が成り立つ．式 (S.5.22)，(S.5.23) は，何れも (π_1, π_2, π_3) 空間中の平面を表し，両者を満たす (π_1, π_2, π_3) は二つの平面の交線上にある．

以上より，$\log P(\mathbf{x})$ の等高線は直線となることがわかる．

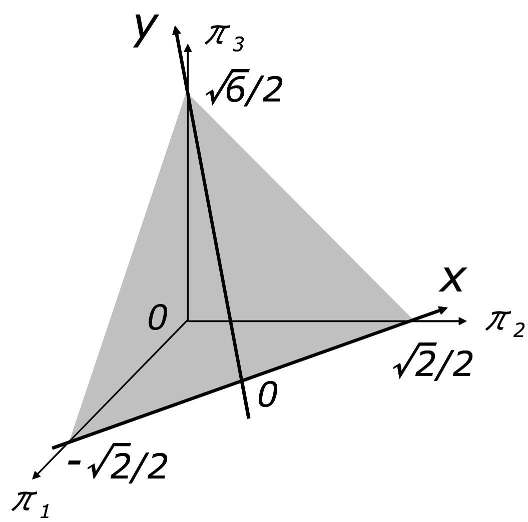

図 A.8: x-y 座標系

ここで，**図 5.4** で取りあげた (π_1, π_2, π_3) の存在範囲を示す 2 次元領域に対し，**図 A.8** の如く x-y 座標系を導入する．簡単な計算により，以下の変換式を導くことができる．

$$\left.\begin{aligned} \pi_1 &= \frac{1}{2} - \frac{\sqrt{2}}{2}x - \frac{\sqrt{6}}{6}y \\ \pi_2 &= \frac{1}{2} + \frac{\sqrt{2}}{2}x - \frac{\sqrt{6}}{6}y \\ \pi_3 &= \frac{\sqrt{6}}{3}y \end{aligned}\right\} \tag{S.5.24}$$

上式を式 (S.5.22) に代入し，整理すると

$$\frac{\sqrt{2}}{2}(\theta_{11}-\theta_{21})x+\frac{\sqrt{6}}{6}(\theta_{11}+\theta_{21}-2\,\theta_{31})y \\ +v_c-\frac{1}{2}(\theta_{11}+\theta_{21})=0 \quad \text{(S.5.25)}$$

を得る．上式で v_c を変化させると，同じ傾きの直線が複数得られることがわかる．実験では，$\theta_1=0.8$，$\theta_2=0.6$，$\theta_3=0.3$ としたので，これらを上式に代入すると

$$0.141\,x+0.327\,y+v_c-0.700=0 \quad \text{(S.5.26)}$$

を得る．対数尤度を最大にするパラメータ π_i は式 (5.53) を満たすことから

$$v_c=\frac{r_1}{n} \quad \text{(S.5.27)}$$

である．実験では，$n=10,000$，$r_1=4,746$ であるので，この結果を用いると，式 (S.5.26) より

$$0.141\,x+0.327\,y-0.225=0 \quad \text{(S.5.28)}$$

となり，この結果は**図 5.5** 中の太線で示されているとおりである．

第 6 章

【演習問題 6.1】

行列 $\mathbf{X}$ を行列 $\mathbf{UV}$ で近似したときの二乗誤差 E は，両行列の対応する要素の差の二乗和であるので

$$\begin{aligned} E &= \sum_{i,j}\left(x_{ij}-\sum_k u_{ik}v_{kj}\right)^2 \\ &= \sum_{i,j}\left({x_{ij}}^2-2x_{ij}\sum_k u_{ik}v_{kj}+\left(\sum_k u_{ik}v_{kj}\right)^2\right) \end{aligned} \quad \text{(S.6.1)}$$

であり，上式を u_{ik}, v_{kj} に関して最小にするには

$$J(\mathbf{U},\mathbf{V})=\sum_{i,j}\left(2x_{ij}\sum_k u_{ik}v_{kj}-\left(\sum_k u_{ik}v_{kj}\right)^2\right) \quad \text{(S.6.2)}$$

を最大にすればよい．

第 6 章で再三述べたように，EM アルゴリズムは対数関数に限らず，イェンゼンの不等式の当てはまる凸関数であれば適用可能である．上式の第 2 項は，2 次関数 x^2 の形をしており，凸関数であるので EM アルゴリズムが適用できる．

式 (6.75) と同様に

$$h^0_{ijk} = \frac{u^0_{ik} v^0_{kj}}{\sum_{l=1}^{K} u^0_{il} v^0_{lj}} \tag{S.6.3}$$

とおくと，イェンゼンの不等式により

$$\left(\sum_k u_{ik} v_{kj} \right)^2 \leq \sum_k h^0_{ijk} \left(\frac{u_{ik} v_{kj}}{h^0_{ijk}} \right)^2 = \sum_k \frac{(u_{ik} v_{kj})^2}{h^0_{ijk}} \tag{S.6.4}$$

が得られる．したがって式 (S.6.2) は

$$J(\mathbf{U}, \mathbf{V}) \geq \sum_{i,j} \left(2x_{ij} \sum_k u_{ik} v_{kj} - \sum_k \frac{(u_{ik} v_{kj})^2}{h^0_{ijk}} \right) \tag{S.6.5}$$

となり，上式の右辺は目的関数 $J(\mathbf{U}, \mathbf{V})$ の下限値を示している．

これまでと同様に，この下限値を u_{ik}，v_{kj} に関して最大化し，その時の u_{ik}，v_{kj} の値を u^0_{ik}，v^0_{kj} としてさらに u_{ik}，v_{kj} に関する最大化を行う．この処理を繰り返すことにより，$J(\mathbf{U}, \mathbf{V})$ の極大値が得られる．

そこで，まず式 (S.6.5) の右辺を u_{ik} で偏微分して零とおくと

$$\sum_j \left(2x_{ij} v_{kj} - \frac{2u_{ik} {v_{kj}}^2}{h^0_{ijk}} \right) = 0 \tag{S.6.6}$$

より

$$u_{ik} \cdot \sum_j \frac{{v_{kj}}^2}{h^0_{ijk}} = \sum_j x_{ij} v_{kj} \tag{S.6.7}$$

が得られ，上式に式 (S.6.3) を代入し，更新ルールとして整理すると，最終的に

$$u_{ik} \leftarrow u_{ik} \cdot \frac{\sum_j x_{ij} v_{kj}}{\sum_{j,l} u_{il} v_{kj} v_{lj}} \tag{S.6.8}$$

が得られる．

同様にして，式 (S.6.5) の右辺を v_{kj} で偏微分して零と置き，整理することにより，更新ルール

$$v_{kj} \leftarrow v_{kj} \cdot \frac{\sum_i x_{ij} u_{ik}}{\sum_{i,l} v_{lj} u_{ik} u_{il}} \tag{S.6.9}$$

が得られる．

【演習問題 6.2】

式 (6.96) より

$$G(\boldsymbol{\theta}^0, \boldsymbol{\phi}) = \sum_{i=1}^{c} \phi_i \log \frac{f_i(\boldsymbol{\theta}^0)}{\phi_i} \tag{S.6.10}$$

が得られる．上式を式 (6.88) の条件下で最大にするため，ラグランジュの未定乗数法を適用する．すなわち，λ を定数として，

$$L = G(\boldsymbol{\theta}^0, \boldsymbol{\phi}) - \lambda \left(\sum_{i=1}^{c} \phi_i - 1 \right) \tag{S.6.11}$$

とおき，L を ϕ_i で偏微分して 0 と置くことにより

$$\frac{\partial L}{\partial \phi_i} = \log \frac{f_i(\boldsymbol{\theta}^0)}{\phi_i} - 1 - \lambda = 0 \tag{S.6.12}$$

となり，これより

$$f_i(\boldsymbol{\theta}^0) = \phi_i e^{\lambda+1} \tag{S.6.13}$$

を得る．ここで，$G(\boldsymbol{\theta}^0, \boldsymbol{\phi})$ を最大にする ϕ_i を ϕ'_i で表すと，上式の両辺に $\sum_{i=1}^{c}$ を施して式 (6.88) を用いることにより，

$$\phi'_i = \frac{f_i(\boldsymbol{\theta}^0)}{\sum_{j=1}^{c} f_j(\boldsymbol{\theta}^0)} \qquad (i = 1, 2, \ldots, c) \tag{S.6.14}$$

と求められる．上記 ϕ'_i $(i = 1, \ldots, c)$ をまとめて $\boldsymbol{\phi}'$ で表すと，式 (S.6.10) より $G(\boldsymbol{\theta}^0, \boldsymbol{\phi})$ の最大値は

$$
\begin{aligned}
G(\boldsymbol{\theta}^0, \boldsymbol{\phi}') &= \sum_{i=1}^{c} \frac{f_i(\boldsymbol{\theta}^0)}{\sum_{j=1}^{c} f_j(\boldsymbol{\theta}^0)} \log \frac{\sum_{j=1}^{c} f_j(\boldsymbol{\theta}^0)}{f_i(\boldsymbol{\theta}^0)} f_i(\boldsymbol{\theta}^0) \\
&= \frac{\sum_{i=1}^{c} f_i(\boldsymbol{\theta}^0)}{\sum_{j=1}^{c} f_j(\boldsymbol{\theta}^0)} \log \sum_{j=1}^{c} f_j(\boldsymbol{\theta}^0) \\
&= \log \sum_{j=1}^{c} f_j(\boldsymbol{\theta}^0) \\
&= J(\boldsymbol{\theta}^0)
\end{aligned}
\tag{S.6.15}
$$

となり，条件 2 を満たすことがわかる．

【演習問題 6.3】

関数 $G(\boldsymbol{\theta}, \boldsymbol{\phi}')$ と $J(\boldsymbol{\theta})$ が，$\boldsymbol{\theta} = \boldsymbol{\theta}^0$ において同じ値，同じ勾配を持つことを示せばよい．

式 (6.95) より，両関数は $\boldsymbol{\theta} = \boldsymbol{\theta}^0$ において同じ値を持つことは明らかである．式 (6.91) より

$$
G(\boldsymbol{\theta}, \boldsymbol{\phi}') = \sum_{i=1}^{c} \phi'_i \cdot F\left(\frac{f_i(\boldsymbol{\theta})}{\phi'_i}\right) \tag{S.6.16}
$$

であり，$\boldsymbol{\theta} = \boldsymbol{\theta}^0$ における $G(\boldsymbol{\theta}, \boldsymbol{\phi}')$ の勾配は，$G(\boldsymbol{\theta}, \boldsymbol{\phi}')$ を $\boldsymbol{\theta}$ で微分し，$\boldsymbol{\theta} = \boldsymbol{\theta}^0$

と置くことにより

$$
\begin{aligned}
&\frac{\partial}{\partial\boldsymbol{\theta}}G(\boldsymbol{\theta},\boldsymbol{\phi}')\bigg|_{\boldsymbol{\theta}=\boldsymbol{\theta}^0} \\
&= \sum_{i=1}^{c}\phi_i'\cdot F'\left(\frac{f_i(\boldsymbol{\theta})}{\phi_i'}\right)\cdot\frac{1}{\phi_i'}\cdot\frac{\partial}{\partial\boldsymbol{\theta}}f_i(\boldsymbol{\theta})\bigg|_{\boldsymbol{\theta}=\boldsymbol{\theta}^0} && \text{(S.6.17)}\\
&= \sum_{i=1}^{c}F'\left(\sum_{j=1}^{c}f_j(\boldsymbol{\theta})\right)\cdot\frac{\partial}{\partial\boldsymbol{\theta}}f_i(\boldsymbol{\theta})\bigg|_{\boldsymbol{\theta}=\boldsymbol{\theta}^0} && \text{(S.6.18)}\\
&= F'\left(\sum_{j=1}^{c}f_j(\boldsymbol{\theta})\right)\cdot\sum_{i=1}^{c}\frac{\partial}{\partial\boldsymbol{\theta}}f_i(\boldsymbol{\theta})\bigg|_{\boldsymbol{\theta}=\boldsymbol{\theta}^0} && \text{(S.6.19)}\\
&= \frac{\partial}{\partial\boldsymbol{\theta}}J(\boldsymbol{\theta})\bigg|_{\boldsymbol{\theta}=\boldsymbol{\theta}^0} && \text{(S.6.20)}
\end{aligned}
$$

となり，両関数の $\boldsymbol{\theta} = \boldsymbol{\theta}^0$ における勾配は等しいことがわかる．ここで，式 (S.6.17) から式 (S.6.18) への変形には，$\boldsymbol{\phi}'$ が式 (6.94) を満たすことを用いた．

以上より，両関数は $\boldsymbol{\theta} = \boldsymbol{\theta}^0$ において接することが確かめられる．

第 7 章

【演習問題 7.1】

これまで c 種のサイコロを含む箱を 1 種類だけ用意していたのを，サイコロの種類と同じ c 種類（箱 1, 箱 2, ..., 箱 c）用意する．そして，サイコロ $\omega_1,\omega_2,\ldots,\omega_c$ の含有率は，それぞれの箱で異なっており，箱 i に含まれるサイコロ ω_j の含有率を $P_i(\omega_j)$ とする．

ここで，ある時点でサイコロ ω_i を取り出した場合には，次に投げるサイコロは箱 i から取り出すことにすると，次に投げるサイコロとして ω_j が選ばれる確率は $P_i(\omega_j)$ である．したがって，サイコロの含有率 $P_i(\omega_j)$ を

$$P_i(\omega_j) = a_{ij} \tag{S.7.1}$$

と設定することにより，式 (7.2) の条件を満たす行列 $\mathbf{A}$ を実現できる．

式 (7.13) の遷移行列 $\mathbf{A}$ を例にとると，以下のようになる．すなわち，3 つの箱（箱 1, 箱 2, 箱 3）があって，箱 1 には，サイコロ ω_1，ω_2，ω_3 が 10%, 40%, 50%,

箱 2 には，20%, 10%, 70%，箱 3 には，30%, 10%, 60%，それぞれ含まれている．すると式 (S.7.1) より

$$\left.\begin{array}{lll} P_1(\omega_1) = a_{11} = 0.1, & P_1(\omega_2) = a_{12} = 0.4, & P_1(\omega_3) = a_{13} = 0.5 \\ P_2(\omega_1) = a_{21} = 0.2, & P_2(\omega_2) = a_{22} = 0.1, & P_2(\omega_3) = a_{23} = 0.7 \\ P_3(\omega_1) = a_{31} = 0.3, & P_3(\omega_2) = a_{32} = 0.1, & P_3(\omega_3) = a_{33} = 0.6 \end{array}\right\} \tag{S.7.2}$$

となる．この様子が**図 A.9** に示されている．

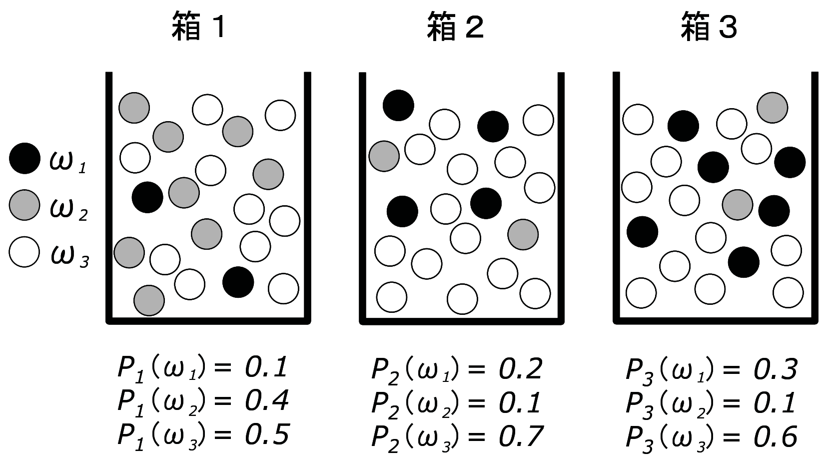

図 A.9: 含有率の異なる 3 種の箱

次に，ある確率に従い，最初のサイコロ s_1 を選び，そのサイコロを投げて出た目を確認し，観測結果 x_1 を得る．最初に取り出したサイコロ s_1 が ω_3 であったとすると（$s_1 = \omega_3$），次に投げるサイコロ s_2 を箱 3 から無作為に取り出して投げ，観測結果 x_2 を得る．このとき，箱 3 よりコイン ω_1，ω_2，ω_3 を取り出す確率はそれぞれ，0.3, 0.1, 0.6 である．もしこのサイコロが ω_1 であったとすると（$s_2 = \omega_1$），次に投げるサイコロ s_3 は箱 1 から取り出す．以下同様に，本操作を，サイコロ s_n を投げて観測結果 x_n を得るまで繰り返す．

このような方法で，遷移確率 $\mathbf{A}$ を持つマルコフ過程を実現できる．

第8章

【演習問題 8.1】

式 (8.9) において $t = 1$ とすることにより

$$\beta_1(i) = P(x_2 x_3 \cdots x_n | s_1 = \omega_i) \tag{S.8.1}$$

を得る．

$$\begin{aligned} P(x_1 x_2 \cdots x_n | s_1 = \omega_i) &= b(\omega_i, x_1)\, P(x_2 x_3 \cdots x_n | s_1 = \omega_i) && \text{(S.8.2)} \\ &= b(\omega_i, x_1)\, \beta_1(i) && \text{(S.8.3)} \end{aligned}$$

であるから

$$\begin{aligned} P(\mathbf{x}) &= \sum_{i=1}^{c} P(s_1 = \omega_i)\, P(x_1 x_2 \cdots x_n | s_1 = \omega_i) && \text{(S.8.4)} \\ &= \sum_{i=1}^{c} \rho_i\, b(\omega_i, x_1)\, \beta_1(i) && \text{(S.8.5)} \end{aligned}$$

と求めることができる．

【演習問題 8.2】

マルコフ性を持たない場合，サイコロ ω_j を取り出す確率は，過去の事象に影響されず，またどの時点 t でも等しいので，式 (7.2) の a_{ij} は

$$\begin{aligned} a_{ij} &= P(s_{t+1} = \omega_j, | s_t = \omega_i) && \text{(S.8.6)} \\ &= P(s_{t+1} = \omega_j) && \text{(S.8.7)} \\ &= P(\omega_j) && \text{(S.8.8)} \end{aligned}$$

となる．

また，マルコフ性を持たない場合，式 (8.18) の $\gamma_t(i)$ は

$$\begin{aligned} \gamma_t(i) &= P(s_t = \omega_i | \mathbf{x}) && \text{(S.8.9)} \\ &= P(\omega_i | x_t) && \text{(S.8.10)} \end{aligned}$$

と表せる．何故なら，t 番目に取り出したサイコロが ω_i である確率は観測結果 x_t にのみ依存し，他の観測結果 $x_\tau(\tau \neq t)$ には依存しないからである．

同様に式 (8.37) の $\xi_t(i,j)$ は

$$\xi_t(i,j) = P(s_t = \omega_i,\, s_{t+1} = \omega_j|\mathbf{x}) \tag{S.8.11}$$

$$= P(\omega_i|x_t)\,P(\omega_j|x_{t+1}) \tag{S.8.12}$$

$$= \gamma_t(i)\,\gamma_{t+1}(j) \tag{S.8.13}$$

と表される．

これらを式 (8.67) に代入することにより

$$\hat{P}(\omega_j) = \frac{\displaystyle\sum_{t=1}^{n-1} P(\omega_i|x_t)\,P(\omega_j|x_{t+1})}{\displaystyle\sum_{t=1}^{n-1} P(\omega_i|x_t)} \tag{S.8.14}$$

を得る．式 (S.8.14) を

$$\hat{P}(\omega_j)\sum_{t=1}^{n-1} P(\omega_i|x_t) = \sum_{t=1}^{n-1} P(\omega_i|x_t)\,P(\omega_j|x_{t+1}) \tag{S.8.15}$$

と変形し，両辺に $\sum_{i=1}^{c}$ を施すと

$$\hat{P}(\omega_j)\sum_{t=1}^{n-1}\sum_{i=1}^{c} P(\omega_i|x_t) = \sum_{t=1}^{n-1} P(\omega_j|x_{t+1})\sum_{i=1}^{c} P(\omega_i|x_t) \tag{S.8.16}$$

が得らる．式 (5.15) で示したように

$$\sum_{i=1}^{c} P(\omega_i|x_t) = 1 \qquad (t = 1, 2, \ldots, n) \tag{S.8.17}$$

であるので，式 (S.8.16) は

$$(n-1)\,\hat{P}(\omega_j) = \sum_{t=1}^{n-1} P(\omega_j|x_{t+1}) \tag{S.8.18}$$

となる．これより，最終的に

$$\hat{P}(\omega_j) = \frac{1}{(n-1)}\sum_{t=1}^{n-1} P(\omega_j|x_{t+1}) \tag{S.8.19}$$

が得られる．上式は，本質的に式 (5.91) と等しい．

次に b_{jk} がどのように書けるかを示す．式 (7.5) の b_{jk} は

$$b_{jk} = P(x_t = v_k | s_t = \omega_j) \tag{S.8.20}$$
$$= P(v_k | \omega_j) \tag{S.8.21}$$

と表される．式 (S.8.10) と式 (S.8.21) を式 (8.86) に代入することにより

$$\hat{P}(v_k|\omega_j) = \frac{\displaystyle\sum_{t=1}^{n} \delta(x_t, v_k)\, P(\omega_j|x_t)}{\displaystyle\sum_{t=1}^{n} P(\omega_j|x_t)} \tag{S.8.22}$$

を得る．ここで，式 (S.8.22) の分子は，式 (5.82)，(5.83) で述べたように，観測結果が v_k となったもののうち，選んだサイコロが ω_j であった回数の期待値を表している．すなわち，n 回の観測のうち $x_t = v_k$ となった回数を r_k とすると

$$\sum_{t=1}^{n} \delta(x_t, v_k)\, P(\omega_j|x_t) = r_k\, P(\omega_j|v_k) \tag{S.8.23}$$

である．また，式 (S.8.22) の分母は，式 (5.78)，(5.79) で述べたように，n 回のうち，サイコロ ω_j を選んだ回数の期待値である．すなわち

$$\sum_{t=1}^{n} P(\omega_j|x_t) = \sum_{l=1}^{m} r_l\, P(\omega_j|v_l) \tag{S.8.24}$$

である．式 (S.8.23)，(S.8.24) より，式 (S.8.22) は

$$\hat{P}(v_k|\omega_j) = \frac{r_k\, P(\omega_j|v_k)}{\displaystyle\sum_{l=1}^{m} r_l\, P(\omega_j|v_l)} \tag{S.8.25}$$

と書き直される．上式は式 (5.93) と同じであることがわかる．

以上より，バウム・ウェルチ　アルゴリズムにおいてマルコフ性を取り除くと，パラメータ推定式は式 (5.91)，(5.93) と一致することが確かめられる．

【演習問題 8.3】

(1) 式 (8.7) の証明

式 (8.1) より，下式が成り立つ．

$$P(x_1 \cdots x_n) = \sum_{s_1 \cdots s_n} P(x_1 \cdots x_n,\ s_1 \cdots s_n) \tag{S.8.26}$$

ここで $P(x_1 \cdots x_n,\ s_1 \cdots s_n)$ は，式 (7.25) で示したように

$$\begin{aligned} P(x_1 \cdots x_n,\ s_1 \cdots s_n) &= \prod_{t=1}^{n} a(s_{t-1}, s_t)\, b(s_t, x_t) &\text{(S.8.27)}\\ &= \prod_{t=1}^{n} P(s_t|s_{t-1}) P(x_t|s_t) &\text{(S.8.28)} \end{aligned}$$

である．ただし

$$a(s_0, s_1) = P(s_1|s_0) = P(s_1) \tag{S.8.29}$$

とする．

以上より

$$\begin{aligned} &P(x_1 \cdots x_n,\ s_t = \omega_i)\\ &= \sum_{\substack{s_1 \cdots s_{t-1}\\ s_{t+1} \cdots s_n}} P(x_1 \cdots x_{t-1},\ s_1 \cdots s_{t-1}) P(s_t = \omega_i|s_{t-1}) P(x_t|s_t = \omega_i)\\ &\qquad \cdot P(s_{t+1}|s_t = \omega_i) P(x_{t+1}|s_{t+1}) P(x_{t+2} \cdots x_n,\ s_{t+2} \cdots s_n) &\text{(S.8.30)}\\ &= \sum_{s_1 \cdots s_{t-1}} P(x_1 \cdots x_{t-1},\ s_1 \cdots s_{t-1}) P(s_t = \omega_i|s_{t-1}) P(x_t|s_t = \omega_i)\\ &\cdot \sum_{s_{t+1} \cdots s_n} P(s_{t+1}|s_t = \omega_i) P(x_{t+1}|s_{t+1}) P(x_{t+2} \cdots x_n,\ s_{t+2} \cdots s_n) &\text{(S.8.31)} \end{aligned}$$

となる．式 (S.8.31) の第 1 項は

$$\begin{aligned} &\sum_{s_1 \cdots s_{t-1}} P(x_1 \cdots x_{t-1},\ s_1 \cdots s_{t-1}) P(s_t = \omega_i|s_{t-1}) P(x_t|s_t = \omega_i)\\ &\qquad = P(x_1 \cdots x_t,\ s_t = \omega_i) &\text{(S.8.32)}\\ &\qquad = \alpha_t(i) &\text{(S.8.33)} \end{aligned}$$

であり，第 2 項は

$$\sum_{s_{t+1}\cdots s_n} P(s_{t+1}|s_t=\omega_i)P(x_{t+1}|s_{t+1})P(x_{t+2}\cdots x_n,\ s_{t+2}\cdots s_n)$$

$$= P(x_{t+1}\cdots x_n|s_t=\omega_i) \tag{S.8.34}$$

$$= \beta_t(i) \tag{S.8.35}$$

である．したがって，式 (S.8.31) より

$$P(x_1\cdots x_n,\ s_t=\omega_i) = \alpha_t(i)\cdot\beta_t(i) \tag{S.8.36}$$

が得られ，式 (8.7) が証明された．

(2) 式 (8.10) の証明

式 (S.8.33) より

$$\alpha_t(j) = P(x_1\cdots x_t,\ s_t=\omega_j) \tag{S.8.37}$$

$$= \sum_{s_1\cdots s_{t-1}} P(x_1\cdots x_{t-1},\ s_1\cdots s_{t-1}) \cdot P(s_t=\omega_j|s_{t-1})P(x_t|s_t=\omega_j) \tag{S.8.38}$$

が得られる．また，式 (S.8.28) より

$$P(x_1\cdots x_{t-1}, s_1\cdots s_{t-1}) = \prod_{k=1}^{t-1} P(s_k|s_{k-1})P(x_k|s_k) \tag{S.8.39}$$

となるので，式 (S.8.38) に代入すると下式が得られる．

$$\alpha_t(j) = \sum_{s_1\cdots s_{t-1}}\prod_{k=1}^{t-1} P(s_k|s_{k-1})P(x_k|s_k)P(s_t=\omega_j|s_{t-1})P(x_t|s_t=\omega_j)$$

$$= \sum_{s_1\cdots s_{t-1}}\prod_{k=1}^{t-2} P(s_k|s_{k-1})P(x_k|s_k) \cdot P(s_{t-1}|s_{t-2})P(x_{t-1}|s_{t-1})P(s_t=\omega_j|s_{t-1})P(x_t|s_t=\omega_j) \tag{S.8.40}$$

ここで式 (S.8.39) より

$$\prod_{k=1}^{t-2} P(s_k|s_{k-1})P(x_k|s_k) = P(x_1\cdots x_{t-2},\ s_1\cdots s_{t-2}) \tag{S.8.41}$$

であることを用い，さらに和の演算を

$$\sum_{s_1 \cdots s_{t-1}} = \sum_{s_1 \cdots s_{t-2}} \sum_{s_{t-1}}$$

と分解すると，式 (S.8.40) は

$$\begin{aligned}
&\alpha_t(j) \\
&= \sum_{s_1 \cdots s_{t-2}} \sum_{s_{t-1}} P(x_1 \cdots x_{t-2},\, s_1 \cdots s_{t-2}) \\
&\quad \cdot P(s_{t-1}|s_{t-2}) P(x_{t-1}|s_{t-1}) P(s_t = \omega_j|s_{t-1}) P(x_t|s_t = \omega_j) \qquad \text{(S.8.42)} \\
&= \sum_{i=1}^{c} \left[\sum_{s_1 \cdots s_{t-2}} P(x_1 \cdots x_{t-2},\, s_1 \cdots s_{t-2}) P(s_{t-1} = \omega_i|s_{t-2}) \right. \\
&\quad \left. \cdot P(x_{t-1}|s_{t-1} = \omega_i) \right] \cdot P(s_t = \omega_j|s_{t-1} = \omega_i) P(x_t|s_t = \omega_j) \qquad \text{(S.8.43)}
\end{aligned}$$

と書ける．式 (S.8.42) から式 (S.8.43) への変形は，演算 $\sum_{s_{t-1}}$ が s_{t-1} を ω_1 から ω_c まで変化させて加算する処理であることを用いた．

式 (S.8.43) の [·] 内の式は，式 (S.8.38) より

$$\begin{aligned}
&\sum_{s_1 \cdots s_{t-2}} P(x_1 \cdots x_{t-2},\, s_1 \cdots s_{t-2}) P(s_{t-1} = \omega_i|s_{t-2}) P(x_{t-1}|s_{t-1} = \omega_i) \\
&\qquad = \alpha_{t-1}(i) \qquad \text{(S.8.44)}
\end{aligned}$$

である．したがって，式 (S.8.43) は

$$\begin{aligned}
\alpha_t(j) &= \sum_{i=1}^{c} \alpha_{t-1}(i)\, P(s_t = \omega_j|s_{t-1} = \omega_i) P(x_t|s_t = \omega_j) \\
&= \left[\sum_{i=1}^{c} \alpha_{t-1}(i)\, a_{ij} \right] b(\omega_j, x_t) \qquad \text{(S.8.45)}
\end{aligned}$$

となり，式 (8.10) が証明された．

(3) 式 (8.16) の証明

式 (S.8.35) より

$$\beta_t(i) = P(x_{t+1} \cdots x_n | s_t = \omega_i) \tag{S.8.46}$$

$$= \sum_{s_{t+1} \cdots s_n} P(s_{t+1} | s_t = \omega_i) P(x_{t+1} | s_{t+1}) \cdot P(x_{t+2} \cdots x_n,\ s_{t+2} \cdots s_n) \tag{S.8.47}$$

が得られる．また，式 (S.8.28) より

$$P(x_{t+2} \cdots x_n,\ s_{t+2} \cdots s_n) = \prod_{k=t+2}^{n} P(s_k | s_{k-1}) P(x_k | s_k) \tag{S.8.48}$$

となるので，式 (S.8.47) に代入すると下式が得られる．

$$\begin{aligned}\beta_t(i) &= \sum_{s_{t+1} \cdots s_n} P(s_{t+1} | s_t = \omega_i) P(x_{t+1} | s_{t+1}) \cdot \prod_{k=t+2}^{n} P(s_k | s_{k-1}) P(x_k | s_k) \\ &= \sum_{s_{t+1} \cdots s_n} P(s_{t+1} | s_t = \omega_i) P(x_{t+1} | s_{t+1}) P(s_{t+2} | s_{t+1}) P(x_{t+2} | s_{t+2}) \\ &\qquad \cdot \prod_{k=t+3}^{n} P(s_k | s_{k-1}) P(x_k | s_k)\end{aligned} \tag{S.8.49}$$

ここで式 (S.8.48) より

$$\prod_{k=t+3}^{n} P(s_k | s_{k-1}) P(x_k | s_k) = P(x_{t+3} \cdots x_n,\ s_{t+2} \cdots s_n) \tag{S.8.50}$$

であることを用い，さらに和の演算を

$$\sum_{s_{t+1} \cdots s_n} = \sum_{s_{t+1}} \sum_{s_{t+2} \cdots s_n} \tag{S.8.51}$$

と分解すると，式 (S.8.49) は

$$\beta_t(i) = \sum_{s_{t+1}} \sum_{s_{t+2}\cdots s_n} P(s_{t+1}|s_t=\omega_i)P(x_{t+1}|s_{t+1})P(s_{t+2}|s_{t+1})P(x_{t+2}|s_{t+2}) \cdot P(x_{t+3}\cdots x_n,\ s_{t+2}\cdots s_n) \tag{S.8.52}$$

$$= \sum_{j=1}^{c} P(s_{t+1}=\omega_j|s_t=\omega_i)P(x_{t+1}|s_{t+1}=\omega_j) \cdot \sum_{s_{t+2}\cdots s_n} P(s_{t+2}|s_{t+1}=\omega_j)P(x_{t+2}|s_{t+2})P(x_{t+3}\cdots x_n,\ s_{t+2}\cdots s_n) \tag{S.8.53}$$

と書ける．式 (S.8.52) から式 (S.8.53) への変形では，演算 $\sum_{s_{t+1}}$ が s_{t+1} を ω_1 から ω_c まで変化させて加算する処理であることを用いた．

式 (S.8.53) 第 2 項の和演算は，式 (S.8.47) より

$$\sum_{s_{t+2}\cdots s_n} P(s_{t+2}|s_{t+1}=\omega_j)P(x_{t+2}|s_{t+2})P(x_{t+3}\cdots x_n,\ s_{t+3}\cdots s_n) = \beta_{t+1}(j) \tag{S.8.54}$$

である．したがって，式 (S.8.53) は

$$\beta_t(i) = \sum_{j=1}^{c} P(s_{t+1}=\omega_j|s_t=\omega_i)P(x_{t+1}|s_{t+1}=\omega_j)\beta_{t+1}(j) = \sum_{j=1}^{c} a_{ij}\, b(\omega_j, x_{t+1})\, \beta_{t+1}(j) \tag{S.8.55}$$

となり，式 (8.16) が証明された．

(4) 式 (8.24) の証明

式 (7.12) より

$$P(x_1,\cdots,x_t,\ s_1,\cdots,s_{t-1},\ s_t=\omega_j) = \prod_{k=1}^{t-1} P(x_k|s_k)\cdot P(x_t|s_t=\omega_j) \tag{S.8.56}$$

が成り立つので

$$\psi_t(j) = \max_{s_1 \cdots s_{t-1}} P(x_1, \cdots, x_t,\ s_1, \cdots, s_{t-1},\ s_t = \omega_j) \tag{S.8.57}$$

$$= \max_{s_1 \cdots s_{t-1}} \prod_{k=1}^{t-1} P(x_k|s_k) \cdot P(x_t|s_t = \omega_j) \tag{S.8.58}$$

と書ける．ここで，最大化の演算を

$$\max_{s_1 \cdots s_{t-1}} \{\,\cdot\,\} = \max_{s_{t-1}} \left\{ \max_{s_1 \cdots s_{t-2}} \{\,\cdot\,\} \right\}$$

と分解すると，式 (S.8.58) は

$$\begin{aligned}\psi_t(j) &\\ = \max_{s_{t-1}} &\left\{ \max_{s_1 \cdots s_{t-2}} \prod_{k=1}^{t-2} P(x_k|s_k) \cdot P(x_{t-1}|s_{t-1}) P(s_t = \omega_j|s_{t-1}) \right\} \\ &\cdot P(x_t|s_t = \omega_j) \qquad \text{(S.8.59)} \\ = \max_{i} &\left\{ \max_{s_1 \cdots s_{t-2}} \prod_{k=1}^{t-2} P(x_k|s_k) \cdot P(x_{t-1}|s_{t-1} = \omega_i) P(s_t = \omega_j|s_{t-1} = \omega_i) \right\} \\ &\cdot P(x_t|s_t = \omega_j) \qquad \text{(S.8.60)}\end{aligned}$$

と書ける．式 (S.8.59) から式 (S.8.60) への変形は，演算 $\max_{s_{t-1}}$ が s_{t-1} を ω_1 から ω_c まで変化させて最大値を求める処理であることを用いた．

式 (S.8.60) において

$$P(s_t = \omega_j|s_{t-1} = \omega_i) = a_{ij} \tag{S.8.61}$$

$$P(x_t|s_t = \omega_j) = b(\omega_j, x_t) \tag{S.8.62}$$

であり，式 (S.8.58) より

$$\max_{s_1 \cdots s_{t-2}} \prod_{k=1}^{t-2} P(x_k|s_k) \cdot P(x_{t-1}|s_{t-1} = \omega_i) = \psi_{t-1}(i) \tag{S.8.63}$$

であるので，式 (S.8.60) は

$$\psi_t(j) = \max_{i} \{\psi_{t-1}(i)\, a_{ij}\}\, b(\omega_j, x_t) \tag{S.8.64}$$

となり，式 (8.24) が証明された.

(5) 式 (8.41) の証明

確率 $P(\mathbf{x},\ s_t=\omega_i,\ s_{t+1}=\omega_j)$ は

$$\begin{aligned}
&P(\mathbf{x},\ s_t=\omega_i,\ s_{t+1}=\omega_j)\\
&=\sum_{\substack{s_1\cdots s_{t-1}\\ s_{t+2}\cdots s_n}} P(x_1\cdots x_{t-1},\ s_1\cdots s_{t-1})P(s_t=\omega_i|s_{t-1})P(x_t|s_t=\omega_i)\\
&\quad\cdot P(s_{t+1}=\omega_j|s_t=\omega_i)P(x_{t+1}|s_{t+1}=\omega_j)\\
&\quad\cdot P(s_{t+2}|s_{t+1}=\omega_j)P(x_{t+2}|s_{t+2})P(x_{t+3}\cdots x_n,\ s_{t+3}\cdots s_n) \qquad \text{(S.8.65)}\\
&=P(s_{t+1}=\omega_j|s_t=\omega_i)P(x_{t+1}|s_{t+1}=\omega_j)\\
&\quad\cdot\sum_{s_1\cdots s_{t-1}} P(x_1\cdots x_{t-1},\ s_1\cdots s_{t-1})P(s_t=\omega_i|s_{t-1})P(x_t|s_t=\omega_i)\\
&\quad\cdot\sum_{s_{t+2}\cdots s_n} P(s_{t+2}|s_{t+1}=\omega_j)P(x_{t+2}|s_{t+2})P(x_{t+3}\cdots x_n,\ s_{t+3}\cdots s_n)
\end{aligned} \qquad \text{(S.8.66)}$$

と書ける．上式において，第 1 項は

$$P(s_{t+1}=\omega_j|s_t=\omega_i)P(x_{t+1}|s_{t+1}=\omega_j)=a_{ij}\,b(\omega_j,x_{t+1}) \qquad \text{(S.8.67)}$$

である．また第 2 項は，式 (S.8.33) より

$$\begin{aligned}
\sum_{s_1\cdots s_{t-1}} &P(x_1\cdots x_{t-1},\ s_1\cdots s_{t-1})P(s_t=\omega_i|s_{t-1})P(x_t|s_t=\omega_i)\\
&=\alpha_t(i)
\end{aligned} \qquad \text{(S.8.68)}$$

であり，第 3 項は，式 (S.8.35) より

$$\begin{aligned}
\sum_{s_{t+2}\cdots s_n} &P(s_{t+2}|s_{t+1}=\omega_j)P(x_{t+2}|s_{t+2})P(x_{t+3}\cdots x_n,\ s_{t+3}\cdots s_n)\\
&=\beta_{t+1}(j)
\end{aligned} \qquad \text{(S.8.69)}$$

である．したがって，式 (S.8.66) は

$$P(\mathbf{x},\ s_t=\omega_i,\ s_{t+1}=\omega_j)=\alpha_t(i)\,a_{ij}\,b(\omega_j,x_{t+1})\,\beta_{t+1}(j) \qquad \text{(S.8.70)}$$

となり，式 (8.41) が証明された．

第9章

【演習問題 9.1】

式の導出にあたっては，付録 A.3 の公式を用いる．

パラメータ $\boldsymbol{\theta}_i$ を推定するための式 (9.47) における $\nabla_{\boldsymbol{\theta}_i}$ は，偏微分 $\partial/\partial\boldsymbol{\mu}_i$，$\partial/\partial\boldsymbol{\Sigma}_i$ を表している．したがって，式 (9.47) は

$$\sum_{t=1}^{n} P(\omega_i|\mathbf{x}_t;\boldsymbol{\theta})\frac{\partial}{\partial\boldsymbol{\mu}_i}\log p(\mathbf{x}_t|\omega_i;\boldsymbol{\theta}_i) = \mathbf{0} \tag{S.9.1}$$

$$\sum_{t=1}^{n} P(\omega_i|\mathbf{x}_t;\boldsymbol{\theta})\frac{\partial}{\partial\boldsymbol{\Sigma}_i}\log p(\mathbf{x}_t|\omega_i;\boldsymbol{\theta}_i) = \mathbf{0} \tag{S.9.2}$$

となる．式 (9.44) より

$$\begin{aligned}&\log p(\mathbf{x}_t|\omega_i;\boldsymbol{\theta}_i)\\ &\quad= -\frac{d}{2}\log(2\pi) - \frac{1}{2}\log|\boldsymbol{\Sigma}_i| - \frac{1}{2}(\mathbf{x}_t-\boldsymbol{\mu}_i)^t{\boldsymbol{\Sigma}_i}^{-1}(\mathbf{x}_t-\boldsymbol{\mu}_i)\end{aligned} \tag{S.9.3}$$

であるので，式 (A.3.5) を用いると

$$\frac{\partial}{\partial\boldsymbol{\mu}_i}\log p(\mathbf{x}_t|\omega_i;\boldsymbol{\theta}_i) = {\boldsymbol{\Sigma}_i}^{-1}(\mathbf{x}_t-\boldsymbol{\mu}_i) \tag{S.9.4}$$

が得られる．上式を式 (S.9.1) に代入し，左から $\boldsymbol{\Sigma}_i$ を乗ずると

$$\sum_{t=1}^{n} P(\omega_i|\mathbf{x}_t;\boldsymbol{\theta})(\mathbf{x}_t-\boldsymbol{\mu}_i) = \mathbf{0} \tag{S.9.5}$$

が得られ，これより

$$\hat{\boldsymbol{\mu}}_i = \frac{\displaystyle\sum_{t=1}^{n} P(\omega_i|\mathbf{x}_t;\boldsymbol{\theta})\,\mathbf{x}_t}{\displaystyle\sum_{t=1}^{n} P(\omega_i|\mathbf{x}_t;\boldsymbol{\theta})} \tag{S.9.6}$$

となり，式 (9.48) が得られる．

一方，式 (A.3.1)，(A.3.6)，(A.3.7) を用いると

$$\begin{aligned}
&\frac{\partial}{\partial \boldsymbol{\Sigma}_i} \log p(\mathbf{x}_t|\omega_i;\boldsymbol{\theta}_i) \\
&= -\frac{1}{2}\frac{\partial}{\partial \boldsymbol{\Sigma}_i}\log|\boldsymbol{\Sigma}_i| - \frac{1}{2}\frac{\partial}{\partial \boldsymbol{\Sigma}_i}\left[(\mathbf{x}_t-\boldsymbol{\mu}_i)^t{\boldsymbol{\Sigma}_i}^{-1}(\mathbf{x}_t-\boldsymbol{\mu}_i)\right] \qquad \text{(S.9.7)}\\
&= -\frac{1}{2}{\boldsymbol{\Sigma}_i}^{-1} - \frac{1}{2}\frac{\partial}{\partial \boldsymbol{\Sigma}_i}\left[\mathrm{tr}\Big({\boldsymbol{\Sigma}_i}^{-1}(\mathbf{x}_t-\boldsymbol{\mu}_i)(\mathbf{x}_t-\boldsymbol{\mu}_i)^t\Big)\right] \qquad \text{(S.9.8)}\\
&= -\frac{1}{2}{\boldsymbol{\Sigma}_i}^{-1} + \frac{1}{2}{\boldsymbol{\Sigma}_i}^{-1}(\mathbf{x}_t-\boldsymbol{\mu}_i)(\mathbf{x}_t-\boldsymbol{\mu}_i)^t{\boldsymbol{\Sigma}_i}^{-1} \qquad \text{(S.9.9)}\\
&= \frac{1}{2}{\boldsymbol{\Sigma}_i}^{-1}\Big((\mathbf{x}_t-\boldsymbol{\mu}_i)(\mathbf{x}_t-\boldsymbol{\mu}_i)^t - \boldsymbol{\Sigma}_i\Big){\boldsymbol{\Sigma}_i}^{-1} \qquad \text{(S.9.10)}
\end{aligned}$$

と計算できる．本結果を式 (S.9.2) に代入し，左右から $\boldsymbol{\Sigma}_i$ を乗ずることにより

$$\sum_{t=1}^{n} P(\omega_i|\mathbf{x}_t;\boldsymbol{\theta})\Big((\mathbf{x}_t-\boldsymbol{\mu}_i)(\mathbf{x}_t-\boldsymbol{\mu}_i)^t - \boldsymbol{\Sigma}_i\Big) = \mathbf{0} \qquad \text{(S.9.11)}$$

が得られる．これより，最終的に

$$\hat{\boldsymbol{\Sigma}}_i = \frac{\displaystyle\sum_{t=1}^{n} P(\omega_i|\mathbf{x}_t;\boldsymbol{\theta})\,(\mathbf{x}_t-\hat{\boldsymbol{\mu}}_i)(\mathbf{x}_t-\hat{\boldsymbol{\mu}}_i)^t}{\displaystyle\sum_{t=1}^{n} P(\omega_i|\mathbf{x}_t;\boldsymbol{\theta})} \qquad \text{(S.9.12)}$$

となり，式 (9.49) が得られる．

第１０章

【演習問題 10.1】

式 (10.37) の条件下で式 (10.41) を最大にする π_i を求めるには，ラグランジュの未定乗数法を適用する．すなわち，定数 λ を用いて

$$L = \log p(\mathbf{x};\boldsymbol{\theta}) - \lambda\left(\sum_{i=1}^{n}\pi_i - 1\right) \qquad \text{(S.10.1)}$$

と置き，$\partial L/\partial \pi_i = 0$ となる π_i を求める．式 (10.41) より

$$\frac{\partial L}{\partial \pi_i} = \sum_{k=1}^{n}\frac{f_{ik}}{\displaystyle\sum_{j=1}^{n}\pi_j\cdot f_{jk}} - \lambda \qquad \text{(S.10.2)}$$

であるから，上式を 0 と置くことにより

$$\lambda = \sum_{k=1}^{n} \frac{f_{ik}}{\sum_{j=1}^{n} \pi_j \cdot f_{jk}} \tag{S.10.3}$$

が得られる．上式の両辺に π_i を乗じることにより

$$\lambda\,\pi_i = \sum_{k=1}^{n} \frac{\pi_i \cdot f_{ik}}{\sum_{j=1}^{n} \pi_j \cdot f_{jk}} \tag{S.10.4}$$

が得られ，さらに両辺に $\sum_{i=1}^{n}$ を施すと

$$\lambda \sum_{i=1}^{n} \pi_i = \sum_{k=1}^{n} \frac{\sum_{i=1}^{n} \pi_i \cdot f_{ik}}{\sum_{j=1}^{n} \pi_j \cdot f_{jk}} \tag{S.10.5}$$

が得られ，これより

$$\lambda = n \tag{S.10.6}$$

が得られる．この結果を式 (S.10.4) に代入すると

$$\pi_i = \frac{1}{n} \sum_{k=1}^{n} \frac{\pi_i \cdot f_{ik}}{\sum_{j=1}^{n} \pi_j \cdot f_{jk}} \tag{S.10.7}$$

となり，繰り返し演算式 (10.45) が得られる．

【演習問題 10.2】

式 (10.37) で定まる実行可能領域が凸集合であることは，**演習問題 5.1** の解答で示したとおりである．

式 (10.41) を次のように書き換える．

$$\log p(\mathbf{x};\boldsymbol{\theta}) = \sum_{k=1}^{n} l_k(\mathbf{x};\boldsymbol{\theta}) \tag{S.10.8}$$

ただし

$$l_k(\mathbf{x};\boldsymbol{\theta}) \stackrel{\text{def}}{=} \log\left(\sum_{i=1}^{n} \pi_i \cdot f_{ik}\right) \tag{S.10.9}$$

である．以下では，$-l_k(\mathbf{x};\boldsymbol{\theta})$ が凸関数であることを示す．

関数 $-l_k(\mathbf{x};\boldsymbol{\theta})$ のヘッセ行列を $\mathbf{H}$ とし，その (i,j) 成分を h_{ij} とすると

$$\begin{aligned} h_{ij} &= -\frac{\partial^2 l_k(\mathbf{x};\boldsymbol{\theta})}{\partial \pi_i \partial \pi_j} && \text{(S.10.10)} \\ &= \frac{f_{ik} \cdot f_{jk}}{\left(\displaystyle\sum_{l=1}^{n} \pi_l \cdot f_{lk}\right)^2} && \text{(S.10.11)} \\ &= u_{ik} \cdot u_{jk} \qquad (i,j = 1,2,\ldots,n) && \text{(S.10.12)} \end{aligned}$$

が得られる．上式で

$$u_{ik} \stackrel{\text{def}}{=} \frac{f_{ik}}{\displaystyle\sum_{l=1}^{n} \pi_l \cdot f_{lk}} > 0 \tag{S.10.13}$$

である．さらに次のような n 次元列ベクトル $\mathbf{u}_k$ を導入する．

$$\mathbf{u}_k \stackrel{\text{def}}{=} (u_{1k}, u_{2k}, \ldots, u_{nk})^t \qquad (k = 1,2,\ldots,n) \tag{S.10.14}$$

するとヘッセ行列 $\mathbf{H}$ は次式で表される．

$$\mathbf{H} = \mathbf{u}_k \cdot \mathbf{u}_k^t \tag{S.10.15}$$

ここで，$\mathbf{0}$ でない任意の n 次元列ベクトル $\mathbf{y}$ に対して以下が成り立つ．

$$\begin{aligned} \mathbf{y}^t\mathbf{H}\mathbf{y} &= \mathbf{y}^t\mathbf{u}_k \cdot \mathbf{u}_k^t\mathbf{y} && \text{(S.10.16)} \\ &= (\mathbf{u}_k^t\mathbf{y})^t \cdot (\mathbf{u}_k^t\mathbf{y}) && \text{(S.10.17)} \\ &= (\mathbf{u}_k^t\mathbf{y})^2 \quad > 0 && \text{(S.10.18)} \end{aligned}$$

上式より行列 $\mathbf{H}$ は正定値であるので，定理 A.1(2) より，関数 $-l_k(\mathbf{x};\boldsymbol{\theta})$ は狭義凸関数であることがわかる．したがって，式 (S.10.8) と定理 A.2 より，関数 $-\log p(\mathbf{x};\boldsymbol{\theta})$ は凸関数（狭義凸関数）となる．

以上より，本問題は凸計画問題であることがわかる．

第１１章

【演習問題 11.1】

問 (1)

分割対象となる n 人の客を一列に並べたとする．ここで「テーブル 1」と記したカードを n_1 枚,「テーブル 2」と記したカードを n_2 枚，. . . ,「テーブル c」と記したカードを n_c 枚用意し，それらを n 人の客の前に並べれば，客を c 個のテーブルに分割したことになる．したがって，N_0 はこのカードの並べ方の数に等しい．ゆえに N_0 は

$$N_0 = \frac{n!}{n_1!\, n_2! \,\cdots\, n_c!} \tag{S.11.1}$$

として求められる．

問 (2)

もし $n_1, n_2, \ldots, n_c$ の c 個の数値がすべて異なるなら

$$N_1 = N_0 \tag{S.11.2}$$

である．しかし，この中に c_1 個同じ数値が含まれていたら，N_0 は同じ分割方法を $c_1!$ 重に重複して数えた結果となっているので，$c_1!$ で除算しなくてはならない．したがって，$n_1, n_2, \ldots, n_c$ の c 個の数値が，c_1 個，c_2 個，. . .，c_r 個の同じ数値を含む場合[*13]

$$N_1 = \frac{1}{c_1!\, c_2! \,\cdots\, c_r!} \cdot N_0 \tag{S.11.3}$$

となる．ただし

$$\sum_{i=1}^{r} c_i = c \tag{S.11.4}$$

[*13] 例えば，$c = 7, n = 20$ で，$(n_1, n_2, n_3, n_4, n_5, n_6, n_7) = (4, 3, 1, 3, 5, 3, 1)$ の場合は，$r = 4$ で，$(c_1, c_2, c_3, c_4) = (1, 1, 2, 3)$ となる．

である．ここで，$n_1, n_2, \ldots, n_c$ の数値がすべて異なるなら

$$c_1 = c_2 = \cdots = c_r = 1 \tag{S.11.5}$$

であるので，式 (S.11.3) は式 (S.11.2) と一致する．

問 (3)

テーブルの区別ができる場合は，N_1 通りの分割方法のそれぞれに対して，テーブルの割り当て方を $c!$ 通り考えることができる．したがって

$$N_2 = c! \cdot N_1 \tag{S.11.6}$$

$$= \frac{c!}{c_1!\, c_2!\, \cdots\, c_r!} \cdot N_0 \tag{S.11.7}$$

となる．

【演習問題 11.2】

事後分布 $P(\boldsymbol{\pi}|s_1, ..., s_{n-1})$ を用いて，$\boldsymbol{\pi}$ を積分消去すると

$$\begin{aligned}
P(s_n = \omega_i|s_1 \cdots s_{n-1}) &= \int P(s_n = \omega_i|s_1 \cdots s_{n-1}, \boldsymbol{\pi})p(\boldsymbol{\pi}|s_1 \cdots s_{n-1})\, d\boldsymbol{\pi} \\
&= \int P(s_n = \omega_i|\boldsymbol{\pi})p(\boldsymbol{\pi}|s_1 \cdots s_{n-1})\, d\boldsymbol{\pi} \\
&= \int \pi_i \cdot p(\boldsymbol{\pi}|s_1 \cdots s_{n-1})\, d\boldsymbol{\pi} \\
&\equiv E[\pi_i|s_1, ..., s_{n-1}] \qquad \text{(S.11.8)}
\end{aligned}$$

が得られる．上で $E[\pi_i|s_1, ..., s_{n-1}]$ は，$\boldsymbol{\pi}$ の事後分布による π_i の条件付き期待値を表している．ここで，事後分布 $p(\boldsymbol{\pi}|s_1 \cdots s_{n-1})$ に対しては，ベイズの定理より

$$p(\boldsymbol{\pi}|s_1 \cdots s_{n-1}) \propto P(s_1 \cdots s_{n-1}|\boldsymbol{\pi}) \cdot p(\boldsymbol{\pi}) \tag{S.11.9}$$

が成り立つ．一方，式 (11.23)，(11.25) で示したように

$$P(s_1 \cdots s_{n-1}|\boldsymbol{\pi}) = \prod_{i=1}^{c} {\pi_i}^{n_i} \tag{S.11.10}$$

$$p(\boldsymbol{\pi}) = \frac{\Gamma(\alpha)}{\Gamma(\alpha/c)^c} \cdot \prod_{i=1}^{c} {\pi_i}^{\alpha/c-1} \tag{S.11.11}$$

であるので，これらを式 (S.11.9) に代入することにより

$$p(\boldsymbol{\pi}|s_1\cdots s_{n-1}) = \frac{1}{Z}\prod_{i=1}^{c}{\pi_i}^{n_i+\alpha/c-1} \tag{S.11.12}$$

が得られる．ただし Z は定数である．ここで

$$\int p(\boldsymbol{\pi}|s_1\cdots s_{n-1})d\boldsymbol{\pi} = 1 \tag{S.11.13}$$

と式 (4.68) より

$$Z = \frac{\prod_{i=1}^{c}\Gamma(n_i+\alpha/c)}{\Gamma(n-1+\alpha)} \tag{S.11.14}$$

と求められる．これを式 (S.11.12) に代入することにより

$$\begin{aligned} p(\boldsymbol{\pi}|s_1\cdots s_{n-1}) &= \frac{\Gamma(n-1+\alpha)}{\prod_{i=1}^{c}\Gamma(n_i+\alpha/c)}\cdot\prod_{i=1}^{c}{\pi_i}^{n_i+\alpha/c-1} \\ &= \mathrm{Dir}(n_1+\alpha/c,\cdots,n_c+\alpha/c) \end{aligned} \tag{S.11.15}$$

となり，$p(\boldsymbol{\pi}|s_1\cdots s_{n-1})$ はディリクレ分布であることがわかる．式 (S.11.8) で示したように，$P(s_n=\omega_i|s_1\cdots s_{n-1})$ はディリクレ分布における π_i の条件付き期待値であるので，式 (4.76) より

$$P(s_n=\omega_i|s_1\cdots s_{n-1}) = E[\pi_i|s_1\cdots s_{n-1}] \tag{S.11.16}$$

$$= \frac{n_i+\alpha/c}{\sum_{j=1}^{c}(n_j+\alpha/c)} \tag{S.11.17}$$

$$= \frac{n_i+\alpha/c}{n-1+\alpha} \tag{S.11.18}$$

が得られる．

第１２章

【演習問題 12.1】

観測データは，平均ベクトル $\boldsymbol{\mu}$，精度行列 $\boldsymbol{\Lambda}$ の正規分布に従うので，k 番目のデータ $\mathbf{x}_k$ の分布関数 $p(\mathbf{x}_k|\boldsymbol{\mu})$ は式 (12.26) より

$$\begin{aligned} p(\mathbf{x}_k|\boldsymbol{\mu}) &= \mathcal{N}(\mathbf{x}_k;\boldsymbol{\mu},\boldsymbol{\Lambda}^{-1}) \\ &= \frac{|\boldsymbol{\Lambda}|^{1/2}}{(2\pi)^{d/2}}\exp\left[-\frac{1}{2}(\mathbf{x}_k-\boldsymbol{\mu})^t\boldsymbol{\Lambda}(\mathbf{x}_k-\boldsymbol{\mu})\right] \end{aligned} \tag{S.12.1}$$

と表される[*14]．

観測データ $\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_n$ をまとめて $\mathbf{x}^{(n)}$ と記すと[*15]，$\mathbf{x}^{(n)}$ の分布関数 $p(\mathbf{x}^{(n)}|\boldsymbol{\mu})$ は

$$p(\mathbf{x}^{(n)}|\boldsymbol{\mu}) = \prod_{k=1}^{n} p(\mathbf{x}_k|\boldsymbol{\mu}) \tag{S.12.2}$$

$$= C_1 \cdot \exp\left[-\frac{1}{2}\sum_{k=1}^{n}(\mathbf{x}_k - \boldsymbol{\mu})^t \boldsymbol{\Lambda}(\mathbf{x}_k - \boldsymbol{\mu})\right] \tag{S.12.3}$$

となる．上式で C_1 は $\boldsymbol{\mu}$ に依存しない定数である．以下使用される定数 C_2，C_3，C_4 も同様である．

ここで，$\boldsymbol{\mu}$ の事前分布 $p(\boldsymbol{\mu})$ として以下の正規分布を仮定する．

$$\begin{aligned} p(\boldsymbol{\mu}) &= \mathcal{N}(\boldsymbol{\mu}; \boldsymbol{\mu}_0, \Lambda_0{}^{-1}) \\ &= \frac{|\boldsymbol{\Lambda}_0|^{1/2}}{(2\pi)^{d/2}} \exp\left[-\frac{1}{2}(\boldsymbol{\mu} - \boldsymbol{\mu}_0)^t \boldsymbol{\Lambda}_0 (\boldsymbol{\mu} - \boldsymbol{\mu}_0)\right] \end{aligned} \tag{S.12.4}$$

上式で $\boldsymbol{\mu}_0$，$\boldsymbol{\Lambda}_0$ は，$\boldsymbol{\mu}$ のそれぞれ平均ベクトル，精度行列である．ベイズの定理より

$$p(\boldsymbol{\mu}|\mathbf{x}^{(n)}) = \frac{p(\mathbf{x}^{(n)}|\boldsymbol{\mu}) \cdot p(\boldsymbol{\mu})}{p(\mathbf{x}^{(n)})} \tag{S.12.5}$$

$$= C_2 \cdot \exp[\,q\,] \tag{S.12.6}$$

を得る．上式の q は式 (S.12.3)，(S.12.4) より

$$q = -\frac{1}{2}\sum_{k=1}^{n}(\mathbf{x}_k - \boldsymbol{\mu})^t \boldsymbol{\Lambda}(\mathbf{x}_k - \boldsymbol{\mu}) - \frac{1}{2}(\boldsymbol{\mu} - \boldsymbol{\mu}_0)^t \boldsymbol{\Lambda}_0 (\boldsymbol{\mu} - \boldsymbol{\mu}_0) \tag{S.12.7}$$

$$= -\frac{1}{2}\boldsymbol{\mu}^t\left(\boldsymbol{\Lambda}_0 + n\boldsymbol{\Lambda}\right)\boldsymbol{\mu} + \boldsymbol{\mu}^t\left(\boldsymbol{\Lambda}_0\boldsymbol{\mu}_0 + n\boldsymbol{\Lambda}\boldsymbol{\mu}_a\right) + C_3 \tag{S.12.8}$$

となる．ここで，$\boldsymbol{\Lambda}$ は対称行列であるので，$\boldsymbol{\mu}^t\boldsymbol{\Lambda}\mathbf{x}_k = \mathbf{x}_k^t\boldsymbol{\Lambda}\boldsymbol{\mu}$ となることを用いた．行列 $\boldsymbol{\Lambda}_0$ についても同様である．

[*14] ベイズ推定を適用しようとしているので，$p(\mathbf{x}_k; \boldsymbol{\mu})$ ではなく，$p(\mathbf{x}_k|\boldsymbol{\mu})$ と表記した．

[*15] ただし，観測の順序は意味を持たない．

上式の $\boldsymbol{\mu}_a$ は

$$\boldsymbol{\mu}_a = \frac{1}{n}\sum_{k=1}^{n}\mathbf{x}_k \tag{S.12.9}$$

であり，式 (9.51) で示したように，$\boldsymbol{\mu}$ の最尤推定解である．ここで，式 (S.12.8) において

$$\boldsymbol{\mu}_b = (\boldsymbol{\Lambda}_0 + n\boldsymbol{\Lambda})^{-1}(\boldsymbol{\Lambda}_0\boldsymbol{\mu}_0 + n\boldsymbol{\Lambda}\boldsymbol{\mu}_a) \tag{S.12.10}$$

$$\boldsymbol{\Lambda}_b = \boldsymbol{\Lambda}_0 + n\boldsymbol{\Lambda} \tag{S.12.11}$$

とおくと，q は

$$q = -\frac{1}{2}\boldsymbol{\mu}^t\boldsymbol{\Lambda}_b\boldsymbol{\mu} + \boldsymbol{\mu}^t\boldsymbol{\Lambda}_b\boldsymbol{\mu}_b + C_3 \tag{S.12.12}$$

$$= -\frac{1}{2}(\boldsymbol{\mu} - \boldsymbol{\mu}_b)^t\boldsymbol{\Lambda}_b(\boldsymbol{\mu} - \boldsymbol{\mu}_b) + C_4 \tag{S.12.13}$$

と書ける．したがって，$\boldsymbol{\mu}$ の事後分布を表す式 (S.12.6) は

$$p(\boldsymbol{\mu}|\mathbf{x}^{(n)}) = \frac{|\boldsymbol{\Lambda}_b|^{1/2}}{(2\pi)^{d/2}}\exp\left[-\frac{1}{2}(\boldsymbol{\mu} - \boldsymbol{\mu}_b)^t\boldsymbol{\Lambda}_b(\boldsymbol{\mu} - \boldsymbol{\mu}_b)\right] \tag{S.12.14}$$

$$= \mathcal{N}(\boldsymbol{\mu}; \boldsymbol{\mu}_b, \Lambda_b{}^{-1}) \tag{S.12.15}$$

となることがわかる．すなわち，$p(\boldsymbol{\mu}|\mathbf{x}^{(n)})$ は，平均ベクトル $\boldsymbol{\mu}_b$，精度行列 $\boldsymbol{\Lambda}_b$ の d 次元正規分布となり，$\boldsymbol{\mu}$ の共役事前分布は正規分布であることがわかる．

式 (S.12.10)，(S.12.11) において $n = 0$ とすると

$$\boldsymbol{\mu}_b = \boldsymbol{\mu}_0 \tag{S.12.16}$$

$$\boldsymbol{\Lambda}_b = \boldsymbol{\Lambda}_0 \tag{S.12.17}$$

が得られ，$\boldsymbol{\mu}$ の事後分布は，式 (S.12.4) の事前分布と一致し，妥当な結果が得られる．また，$n \to \infty$ とすると

$$\boldsymbol{\mu}_b = \left(\frac{1}{n}\boldsymbol{\Lambda}_0 + \boldsymbol{\Lambda}\right)^{-1}\left(\frac{1}{n}\boldsymbol{\Lambda}_0\boldsymbol{\mu}_0 + \boldsymbol{\Lambda}\boldsymbol{\mu}_a\right) \to \boldsymbol{\mu}_a \tag{S.12.18}$$

$$\boldsymbol{\Lambda}_b{}^{-1} \to \mathbf{0} \tag{S.12.19}$$

が得られ（$\mathbf{0}$ はゼロ行列），観測データ数が大きくなると，$\boldsymbol{\mu}$ の事後分布は，最尤推定解 $\boldsymbol{\mu}_a$ で大きな値をとる急峻な分布となる．すなわち，観測データ数が無限大の極限では，ベイズ推定の解は最尤推定解と一致することがわかる．

以上のことから，式 (S.12.10) は，n の値に依存して，$\boldsymbol{\mu}_0$ と $\boldsymbol{\mu}_a$ の中間のベクトル $\boldsymbol{\mu}_b$ を生成する機能を有していることがわかる．

ちなみに，$\boldsymbol{\Lambda}_0 \to \mathbf{0}$ とすると，式 (S.12.4) から明らかなように $p(\boldsymbol{\mu})$ は一様分布となり，式 (S.12.10)，(S.12.11) より下式が得られる．

$$\boldsymbol{\mu}_b = \boldsymbol{\mu}_a \tag{S.12.20}$$

$$\boldsymbol{\Lambda}_b = n\boldsymbol{\Lambda} \tag{S.12.21}$$

これまで述べた内容を $d = 1$ の場合で確認すると，より直観的な理解が得られる．この場合，精度行列 $\boldsymbol{\Lambda}$ の代わりに，分散の逆数である精度 $\lambda\,(= 1/\sigma^2)$ を用いる．式 (S.12.11)，(S.12.10) において $d = 1$ とすると，

$$\lambda_b = \lambda_0 + n\lambda \tag{S.12.22}$$

$$\mu_b = \frac{\lambda_0}{\lambda_0 + n\lambda} \cdot \mu_0 + \frac{n\lambda}{\lambda_0 + n\lambda} \cdot \mu_a \tag{S.12.23}$$

が得られる．上式で，(μ, λ) は観測データの確率分布である正規分布の平均と精度，(μ_0, λ_0) は μ の事前分布である正規分布の平均と精度，(μ_b, λ_b) は μ の事後分布である正規分布の平均と精度である．また，μ_a は μ の最尤推定解である．

【演習問題 12.2】

先の**演習問題 12.1** と同様にして，

$$\begin{aligned} p(\mathbf{x}_k|\boldsymbol{\Lambda}) &= \mathcal{N}(\mathbf{x}_k; \boldsymbol{\mu}, \boldsymbol{\Lambda}^{-1}) \\ &= \frac{|\boldsymbol{\Lambda}|^{1/2}}{(2\pi)^{d/2}} \exp\left[-\frac{1}{2}(\mathbf{x}_k - \boldsymbol{\mu})^t \boldsymbol{\Lambda}(\mathbf{x}_k - \boldsymbol{\mu})\right] \end{aligned} \tag{S.12.24}$$

であり，$\mathbf{x}^{(n)}$ の分布関数 $p(\mathbf{x}^{(n)}|\boldsymbol{\Lambda})$ は

$$p(\mathbf{x}^{(n)}|\boldsymbol{\Lambda}) = \prod_{k=1}^{n} p(\mathbf{x}_k|\boldsymbol{\Lambda}) \tag{S.12.25}$$

$$= C_1 \cdot |\boldsymbol{\Lambda}|^{n/2} \cdot \exp\left[-\frac{1}{2}\sum_{k=1}^{n}(\mathbf{x}_k - \boldsymbol{\mu})^t \boldsymbol{\Lambda}(\mathbf{x}_k - \boldsymbol{\mu})\right] \tag{S.12.26}$$

となる．上式で C_1 は $\mathbf{\Lambda}$ に依存しない定数である．以下使用される定数 C_2，C_3，C_4 も同様である．上式は，$\mathbf{\Lambda}$ を C_1 に含められない点が式 (S.12.3) とは異なる．

ここで，$\mathbf{\Lambda}$ の事前分布 $p(\mathbf{\Lambda})$ として，パラメータ ν，$\mathbf{S}$ を持つウィシャート分布 $\mathcal{W}(\mathbf{\Lambda};\nu,\mathbf{S})$ を仮定すると，式 (12.30) の分母は $\mathbf{\Lambda}$ に無関係な定数と見なせるので

$$
\begin{aligned}
p(\mathbf{\Lambda}) &= \mathcal{W}(\mathbf{\Lambda};\nu,\mathbf{S}) && \text{(S.12.27)}\\
&= C_2\cdot|\mathbf{\Lambda}|^{(\nu-d-1)/2}\cdot\exp\left[-\frac{1}{2}\mathrm{tr}\left(\mathbf{S}^{-1}\mathbf{\Lambda}\right)\right] && \text{(S.12.28)}
\end{aligned}
$$

と書ける．式 (S.12.26)，(S.12.28) をベイズの定理

$$
p(\mathbf{\Lambda}|\mathbf{x}^{(n)}) = \frac{p(\mathbf{x}^{(n)}|\mathbf{\Lambda})\cdot p(\mathbf{\Lambda})}{p(\mathbf{x}^{(n)})} \qquad \text{(S.12.29)}
$$

に代入すると

$$
p(\mathbf{\Lambda}|\mathbf{x}^{(n)}) = C_3\cdot|\mathbf{\Lambda}|^{(\nu_b-d-1)/2}\cdot\exp[\,q\,] \qquad \text{(S.12.30)}
$$

となる．ここで

$$
\begin{aligned}
\nu_b &= \nu+n && \text{(S.12.31)}\\
q &= -\frac{1}{2}\mathrm{tr}\left(\mathbf{S}^{-1}\mathbf{\Lambda}\right)-\frac{1}{2}\sum_{k=1}^{n}(\mathbf{x}_k-\boldsymbol{\mu})^t\mathbf{\Lambda}(\mathbf{x}_k-\boldsymbol{\mu}) && \text{(S.12.32)}
\end{aligned}
$$

である．付録 A.3 の式 (A.3.2) より

$$
(\mathbf{x}_k-\boldsymbol{\mu})^t\mathbf{\Lambda}(\mathbf{x}_k-\boldsymbol{\mu}) = \mathrm{tr}\left((\mathbf{x}_k-\boldsymbol{\mu})(\mathbf{x}_k-\boldsymbol{\mu})^t\mathbf{\Lambda}\right) \qquad \text{(S.12.33)}
$$

となることを用いると

$$
\begin{aligned}
q &= -\frac{1}{2}\cdot\mathrm{tr}\left(\mathbf{S}^{-1}\mathbf{\Lambda}\right)-\frac{1}{2}\sum_{k=1}^{n}\mathrm{tr}\left((\mathbf{x}_k-\boldsymbol{\mu})(\mathbf{x}_k-\boldsymbol{\mu})^t\mathbf{\Lambda}\right) && \text{(S.12.34)}\\
&= -\frac{1}{2}\cdot\mathrm{tr}\left[\left(\mathbf{S}^{-1}+\sum_{k=1}^{n}(\mathbf{x}_k-\boldsymbol{\mu})(\mathbf{x}_k-\boldsymbol{\mu})^t\right)\mathbf{\Lambda}\right] && \text{(S.12.35)}\\
&= -\frac{1}{2}\cdot\mathrm{tr}\left({\mathbf{S}_b}^{-1}\mathbf{\Lambda}\right) && \text{(S.12.36)}
\end{aligned}
$$

が得られる．ただし

$$\mathbf{S}_b{}^{-1} = \mathbf{S}^{-1} + \sum_{k=1}^{n}(\mathbf{x}_k - \boldsymbol{\mu})(\mathbf{x}_k - \boldsymbol{\mu})^t \tag{S.12.37}$$

$$= \mathbf{S}^{-1} + n \cdot \mathbf{S}_a{}^{-1} \tag{S.12.38}$$

である．ここで

$$\mathbf{S}_a{}^{-1} = \frac{1}{n}\sum_{k=1}^{n}(\mathbf{x}_k - \boldsymbol{\mu})(\mathbf{x}_k - \boldsymbol{\mu})^t \tag{S.12.39}$$

であり，$\mathbf{S}_a$ は観測データより最尤推定で求めた精度行列である．以上の結果を用いると，式 (S.12.30) より

$$p(\boldsymbol{\Lambda}|\mathbf{x}^{(n)}) = C_4 \cdot |\boldsymbol{\Lambda}|^{(\nu_b - d - 1)/2} \cdot \exp\left[-\frac{1}{2}\mathrm{tr}\left(\mathbf{S}_b{}^{-1}\boldsymbol{\Lambda}\right)\right] \tag{S.12.40}$$

$$= \mathcal{W}(\boldsymbol{\Lambda}; \nu_b, \mathbf{S}_b) \tag{S.12.41}$$

が得られ，$\boldsymbol{\Lambda}$ の事後分布は，パラメータ ν_b，$\mathbf{S}_b$ を持つウィシャート分布となることがわかる．したがって，精度行列の共役事前分布はウィシャート分布であることが確かめられた．

式 (S.12.31)，(S.12.38) より，パラメータ ν，$\mathbf{S}$ に関して以下のことがわかる．

両式において $n = 0$ とすると，それぞれ $\nu_b = \nu$，$\mathbf{S}_b = \mathbf{S}$ となり，式 (S.12.41) は式 (S.12.27) と一致する．すなわち，データを観測する前は，$\boldsymbol{\Lambda}$ の事後分布は事前分布に等しいという妥当な結果となる．

また，n 個の観測データが得られると，ν_b は n だけ増大し，$\mathbf{S}_b{}^{-1}$ は $n \cdot \mathbf{S}_a{}^{-1}$ だけ増大する．したがって，パラメータ ν は，ν 個のデータが事前に観測されていることに相当し，パラメータ $\mathbf{S}$ は，精度行列が $\nu \cdot \mathbf{S}$（すなわち，共分散行列が $\frac{1}{\nu} \cdot \mathbf{S}^{-1}$）であるような，$\nu$ 個のデータが事前に観測されていることに相当する．

【演習問題 12.3】

演習問題 12.1，**演習問題 12.2** と同様にして，

$$p(\mathbf{x}_k|\boldsymbol{\mu}, \boldsymbol{\Lambda}) = \mathcal{N}(\mathbf{x}_k; \boldsymbol{\mu}, \boldsymbol{\Lambda}^{-1}) \tag{S.12.42}$$

$$= \frac{|\boldsymbol{\Lambda}|^{1/2}}{(2\pi)^{d/2}} \exp\left[-\frac{1}{2}(\mathbf{x}_k - \boldsymbol{\mu})^t \boldsymbol{\Lambda}(\mathbf{x}_k - \boldsymbol{\mu})\right] \tag{S.12.43}$$

であり，$\mathbf{x}^{(n)}$ の分布関数 $p(\mathbf{x}^{(n)}|\boldsymbol{\mu}, \boldsymbol{\Lambda})$ は

$$\begin{aligned}
&p(\mathbf{x}^{(n)}|\boldsymbol{\mu}, \boldsymbol{\Lambda}) \\
&\quad = \prod_{k=1}^{n} p(\mathbf{x}_k|\boldsymbol{\mu}, \boldsymbol{\Lambda}) && \text{(S.12.44)} \\
&\quad = C_1 \cdot |\boldsymbol{\Lambda}|^{n/2} \cdot \exp\left[-\frac{1}{2}\sum_{k=1}^{n}(\mathbf{x}_k - \boldsymbol{\mu})^t \boldsymbol{\Lambda}(\mathbf{x}_k - \boldsymbol{\mu})\right] && \text{(S.12.45)}
\end{aligned}$$

となる．上式で C_1 は $\boldsymbol{\mu}$，$\boldsymbol{\Lambda}$ に依存しない定数である．以後使用される定数 C_2，C_3，C_4 も同様である．

ここで，$\boldsymbol{\mu}$，$\boldsymbol{\Lambda}$ の事前分布 $p(\boldsymbol{\mu}, \boldsymbol{\Lambda})$ として正規－ウィシャート分布を仮定し

$$\begin{aligned}
p(\boldsymbol{\mu}, \boldsymbol{\Lambda}) &= \mathcal{N}\left(\boldsymbol{\mu}; \boldsymbol{\mu}_0, (\beta\boldsymbol{\Lambda})^{-1}\right) \cdot \mathcal{W}(\boldsymbol{\Lambda}; \nu, \mathbf{S}) && \text{(S.12.46)} \\
&= C_2 \cdot |\beta\boldsymbol{\Lambda}|^{1/2} \exp\left[-\frac{1}{2}(\boldsymbol{\mu} - \boldsymbol{\mu}_0)^t \beta\boldsymbol{\Lambda}(\boldsymbol{\mu} - \boldsymbol{\mu}_0)\right] \\
&\qquad \cdot |\boldsymbol{\Lambda}|^{(\nu-d-1)/2} \cdot \exp\left[-\frac{1}{2}\mathrm{tr}\left(\mathbf{S}^{-1}\boldsymbol{\Lambda}\right)\right] && \text{(S.12.47)}
\end{aligned}$$

とする．以上をベイズの定理に代入することにより

$$\begin{aligned}
p(\boldsymbol{\mu}, \boldsymbol{\Lambda}|\mathbf{x}^{(n)}) &= \frac{p(\mathbf{x}^{(n)}|\boldsymbol{\mu}, \boldsymbol{\Lambda}) \cdot p(\boldsymbol{\mu}, \boldsymbol{\Lambda})}{p(\mathbf{x}^{(n)})} && \text{(S.12.48)} \\
&= C_3 \cdot |\boldsymbol{\Lambda}|^{(\nu+n-d)/2} \cdot \exp[\,q\,] && \text{(S.12.49)}
\end{aligned}$$

を得る．上式の q は，式 (S.12.45)，(S.12.47) より

$$\begin{aligned}
q &= -\frac{1}{2}\mathrm{tr}\left(\mathbf{S}^{-1}\boldsymbol{\Lambda}\right) \\
&\qquad -\frac{1}{2}(\boldsymbol{\mu} - \boldsymbol{\mu}_0)^t \beta\boldsymbol{\Lambda}(\boldsymbol{\mu} - \boldsymbol{\mu}_0) - \frac{1}{2}\sum_{k=1}^{n}(\mathbf{x}_k - \boldsymbol{\mu})^t \boldsymbol{\Lambda}(\mathbf{x}_k - \boldsymbol{\mu}) \\
&= -\frac{1}{2}\mathrm{tr}\left(\mathbf{S}^{-1}\boldsymbol{\Lambda}\right) - \frac{1}{2}(n+\beta)\boldsymbol{\mu}^t\boldsymbol{\Lambda}\boldsymbol{\mu} + \boldsymbol{\mu}^t\boldsymbol{\Lambda}\left(\sum_{k=1}^{n}\mathbf{x}_k + \beta\boldsymbol{\mu}_0\right) \\
&\qquad -\frac{1}{2}\sum_{k=1}^{n}{\mathbf{x}_k}^t\boldsymbol{\Lambda}\mathbf{x}_k - \frac{1}{2}\beta\boldsymbol{\mu}_0^t\boldsymbol{\Lambda}\boldsymbol{\mu}_0 && \text{(S.12.50)}
\end{aligned}$$

である．ここで，以下の如く $\boldsymbol{\mu}_b$，$\boldsymbol{\Lambda}_b$ を定義する．

$$\boldsymbol{\mu}_b = \frac{\sum_{k=1}^{n} \mathbf{x}_k + \beta \boldsymbol{\mu}_0}{n+\beta} \tag{S.12.51}$$

$$\boldsymbol{\Lambda}_b = (n+\beta)\boldsymbol{\Lambda} \tag{S.12.52}$$

これにより式 (S.12.50) は

$$\begin{aligned}
q = &-\frac{1}{2}\mathrm{tr}\left(\mathbf{S}^{-1}\boldsymbol{\Lambda}\right) - \frac{1}{2}\boldsymbol{\mu}^t\boldsymbol{\Lambda}_b\boldsymbol{\mu} + \boldsymbol{\mu}^t\boldsymbol{\Lambda}_b\boldsymbol{\mu}_b \\
&-\frac{1}{2}\sum_{k=1}^{n}\mathbf{x}_k{}^t\boldsymbol{\Lambda}\mathbf{x}_k - \frac{1}{2}\beta\,\boldsymbol{\mu}_0^t\boldsymbol{\Lambda}\boldsymbol{\mu}_0 \qquad \text{(S.12.53)}\\
= &-\frac{1}{2}\mathrm{tr}\left(\mathbf{S}^{-1}\boldsymbol{\Lambda}\right) - \frac{1}{2}(\boldsymbol{\mu}-\boldsymbol{\mu}_b)^t\boldsymbol{\Lambda}_b(\boldsymbol{\mu}-\boldsymbol{\mu}_b) + \frac{1}{2}\boldsymbol{\mu}_b{}^t\boldsymbol{\Lambda}_b\boldsymbol{\mu}_b \\
&-\frac{1}{2}\sum_{k=1}^{n}\mathbf{x}_k{}^t\boldsymbol{\Lambda}\mathbf{x}_k - \frac{1}{2}\beta\,\boldsymbol{\mu}_0^t\boldsymbol{\Lambda}\boldsymbol{\mu}_0 \qquad \text{(S.12.54)}\\
= &-\frac{1}{2}\mathrm{tr}\left(\mathbf{S}^{-1}\boldsymbol{\Lambda}\right) - \frac{1}{2}(\boldsymbol{\mu}-\boldsymbol{\mu}_b)^t\boldsymbol{\Lambda}_b(\boldsymbol{\mu}-\boldsymbol{\mu}_b) + \frac{1}{2}\mathrm{tr}\left((n+\beta)\,\boldsymbol{\mu}_b\boldsymbol{\mu}_b^t\boldsymbol{\Lambda}\right) \\
&-\frac{1}{2}\mathrm{tr}\left(\sum_{k=1}^{n}\mathbf{x}_k\mathbf{x}_k^t\boldsymbol{\Lambda}\right) - \frac{1}{2}\beta\,\mathrm{tr}\left(\boldsymbol{\mu}_0\boldsymbol{\mu}_0^t\boldsymbol{\Lambda}\right) \qquad \text{(S.12.55)}\\
= &-\frac{1}{2}(\boldsymbol{\mu}-\boldsymbol{\mu}_b)^t\boldsymbol{\Lambda}_b(\boldsymbol{\mu}-\boldsymbol{\mu}_b) \\
&-\frac{1}{2}\mathrm{tr}\left[\left(\mathbf{S}^{-1} - (n+\beta)\boldsymbol{\mu}_b\boldsymbol{\mu}_b^t + \sum_{k=1}^{n}\mathbf{x}_k\mathbf{x}_k^t + \beta\,\boldsymbol{\mu}_0\boldsymbol{\mu}_0^t\right)\boldsymbol{\Lambda}\right] \qquad \text{(S.12.56)}\\
= &-\frac{1}{2}(\boldsymbol{\mu}-\boldsymbol{\mu}_b)^t\boldsymbol{\Lambda}_b(\boldsymbol{\mu}-\boldsymbol{\mu}_b) - \frac{1}{2}\mathrm{tr}\left(\mathbf{S}_b^{-1}\boldsymbol{\Lambda}\right) \qquad \text{(S.12.57)}
\end{aligned}$$

と変形できる．ただし，

$$\mathbf{S}_b^{-1} = \mathbf{S}^{-1} - (n+\beta)\boldsymbol{\mu}_b\boldsymbol{\mu}_b^t + \sum_{k=1}^{n}\mathbf{x}_k\mathbf{x}_k^t + \beta\,\boldsymbol{\mu}_0\boldsymbol{\mu}_0^t \tag{S.12.58}$$

である．ここで，式 (S.12.54) から式 (S.12.55) への変形には，付録 A.3 の式 (A.3.2) を用いた．なお

$$\bar{\mathbf{x}} = \frac{1}{n}\sum_{k=1}^{n}\mathbf{x}_k \tag{S.12.59}$$

と置くと，簡単な計算により，式 (S.12.58) は

$$\mathbf{S}_b^{-1} = \mathbf{S}^{-1} + \sum_{k=1}^{n} (\mathbf{x}_k - \bar{\mathbf{x}})(\mathbf{x}_k - \bar{\mathbf{x}})^t + \frac{n\beta}{n+\beta}(\bar{\mathbf{x}} - \boldsymbol{\mu}_0)(\bar{\mathbf{x}} - \boldsymbol{\mu}_0)^t \tag{S.12.60}$$

となることが確かめられる．式 (S.12.57) を式 (S.12.49) に代入すると

$$\begin{aligned}
&p(\boldsymbol{\mu}, \boldsymbol{\Lambda}|\mathbf{x}^{(n)}) \\
&= C_3 \cdot |\boldsymbol{\Lambda}|^{(\nu+n-d)/2} \cdot \exp\left[-\frac{1}{2}(\boldsymbol{\mu} - \boldsymbol{\mu}_b)^t \boldsymbol{\Lambda}_b (\boldsymbol{\mu} - \boldsymbol{\mu}_b)\right] \cdot \exp\left[-\frac{1}{2}\mathrm{tr}\left(\mathbf{S}_b^{-1}\boldsymbol{\Lambda}\right)\right] \\
&= C_3 \cdot |\boldsymbol{\Lambda}|^{(\nu+n-d)/2} \cdot (2\pi)^{d/2} |\boldsymbol{\Lambda}_b|^{-1/2} \cdot \mathcal{N}(\boldsymbol{\mu}; \boldsymbol{\mu}_b, \boldsymbol{\Lambda}_b^{-1}) \cdot \exp\left[-\frac{1}{2}\mathrm{tr}\left(\mathbf{S}_b^{-1}\boldsymbol{\Lambda}\right)\right] \\
&= C_4 \cdot |\boldsymbol{\Lambda}|^{(\nu+n-d-1)/2} \cdot \mathcal{N}\left(\boldsymbol{\mu}; \boldsymbol{\mu}_b, ((n+\beta)\boldsymbol{\Lambda})^{-1}\right) \cdot \exp\left[-\frac{1}{2}\mathrm{tr}\left(\mathbf{S}_b^{-1}\boldsymbol{\Lambda}\right)\right] \\
&= C_4 \cdot \mathcal{N}\left(\boldsymbol{\mu}; \boldsymbol{\mu}_b, (\beta_b\boldsymbol{\Lambda})^{-1}\right) \cdot |\boldsymbol{\Lambda}|^{(\nu_b-d-1)/2} \cdot \exp\left[-\frac{1}{2}\mathrm{tr}\left(\mathbf{S}_b^{-1}\boldsymbol{\Lambda}\right)\right] \\
&= \mathcal{N}\left(\boldsymbol{\mu}; \boldsymbol{\mu}_b, (\beta_b\boldsymbol{\Lambda})^{-1}\right) \cdot \mathcal{W}(\boldsymbol{\Lambda}; \nu_b, \mathbf{S}_b) \qquad \text{(S.12.61)}
\end{aligned}$$

が得られる．ただし，上式の ν_b，β_b は下式で表される．

$$\nu_b = \nu + n \tag{S.12.62}$$

$$\beta_b = \beta + n \tag{S.12.63}$$

式 (S.12.61) より，事後分布 $p(\boldsymbol{\mu}, \boldsymbol{\Lambda}|\mathbf{x}^{(n)})$ も正規－ウィシャート分布となり，$\boldsymbol{\mu}$，$\boldsymbol{\Lambda}$ の共役事前分布が正規－ウィシャート分布であることが確かめられた．

【演習問題 12.4】

題意より

$$\mathcal{N}(\mathbf{x}; \boldsymbol{\mu}, \boldsymbol{\Lambda}^{-1}) = \frac{|\boldsymbol{\Lambda}|^{1/2}}{(2\pi)^{d/2}} \exp\left[-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu})^t \boldsymbol{\Lambda} (\mathbf{x} - \boldsymbol{\mu})\right] \tag{S.12.64}$$

$$\mathcal{N}\left(\boldsymbol{\mu}; \boldsymbol{\mu}_0, (\beta\boldsymbol{\Lambda})^{-1}\right) = \frac{|\beta\boldsymbol{\Lambda}|^{1/2}}{(2\pi)^{d/2}} \exp\left[-\frac{1}{2}(\boldsymbol{\mu} - \boldsymbol{\mu}_0)^t \beta\boldsymbol{\Lambda} (\boldsymbol{\mu} - \boldsymbol{\mu}_0)\right] \tag{S.12.65}$$

である．以上より

$$\mathcal{N}(\mathbf{x};\boldsymbol{\mu},\boldsymbol{\Lambda}^{-1})\cdot\mathcal{N}\left(\boldsymbol{\mu};\boldsymbol{\mu}_0,(\beta\boldsymbol{\Lambda})^{-1}\right)=\frac{\beta^{d/2}|\boldsymbol{\Lambda}|}{(2\pi)^d}\cdot\exp\left[f_1\right] \quad \text{(S.12.66)}$$

と書ける．ただし

$$\begin{aligned} f_1 &= -\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^t\boldsymbol{\Lambda}(\mathbf{x}-\boldsymbol{\mu})-\frac{1}{2}(\boldsymbol{\mu}-\boldsymbol{\mu}_0)^t\beta\boldsymbol{\Lambda}(\boldsymbol{\mu}-\boldsymbol{\mu}_0) && \text{(S.12.67)}\\ &= -\frac{1}{2}\mathrm{tr}\left[\left((\mathbf{x}-\boldsymbol{\mu})(\mathbf{x}-\boldsymbol{\mu})^t+\beta(\boldsymbol{\mu}-\boldsymbol{\mu}_0)(\boldsymbol{\mu}-\boldsymbol{\mu}_0)^t\right)\boldsymbol{\Lambda}\right] && \text{(S.12.68)}\\ &= -\frac{1}{2}\mathrm{tr}\left(F_1\boldsymbol{\Lambda}\right) && \text{(S.12.69)} \end{aligned}$$

である．上で式 (S.12.67) から式 (S.12.68) への変形には式 (A.3.2) を用いた．式 (S.12.69) の F_1 は

$$\begin{aligned} F_1 &= (\mathbf{x}-\boldsymbol{\mu})(\mathbf{x}-\boldsymbol{\mu})^t+\beta(\boldsymbol{\mu}-\boldsymbol{\mu}_0)(\boldsymbol{\mu}-\boldsymbol{\mu}_0)^t && \text{(S.12.70)}\\ &= (1+\beta)\left[\boldsymbol{\mu}\boldsymbol{\mu}^t-\boldsymbol{\mu}\frac{(\mathbf{x}+\beta\boldsymbol{\mu}_0)^t}{1+\beta}-\frac{\mathbf{x}+\beta\boldsymbol{\mu}_0}{1+\beta}\boldsymbol{\mu}^t\right]\\ &\qquad +\beta\boldsymbol{\mu}_0\boldsymbol{\mu}_0^t+\mathbf{x}\mathbf{x}^t && \text{(S.12.71)} \end{aligned}$$

と書ける．ここで

$$\boldsymbol{\mu}_a=\frac{\mathbf{x}+\beta\boldsymbol{\mu}_0}{1+\beta} \quad \text{(S.12.72)}$$

と置くと

$$\begin{aligned} F_1 &= (1+\beta)(\boldsymbol{\mu}\boldsymbol{\mu}^t-\boldsymbol{\mu}\boldsymbol{\mu}_a^t-\boldsymbol{\mu}_a\boldsymbol{\mu}^t)+\beta\boldsymbol{\mu}_0\boldsymbol{\mu}_0^t+\mathbf{x}\mathbf{x}^t\\ &= (1+\beta)(\boldsymbol{\mu}-\boldsymbol{\mu}_a)(\boldsymbol{\mu}-\boldsymbol{\mu}_a)^t-(1+\beta)\boldsymbol{\mu}_a\boldsymbol{\mu}_a^t+\beta\boldsymbol{\mu}_0\boldsymbol{\mu}_0^t+\mathbf{x}\mathbf{x}^t\\ &= (1+\beta)(\boldsymbol{\mu}-\boldsymbol{\mu}_a)(\boldsymbol{\mu}-\boldsymbol{\mu}_a)^t+\frac{\beta}{1+\beta}(\mathbf{x}-\boldsymbol{\mu}_0)(\mathbf{x}-\boldsymbol{\mu}_0)^t && \text{(S.12.73)} \end{aligned}$$

が得られる．ここで，さらに

$$\begin{aligned} \boldsymbol{\Lambda}_a &= (1+\beta)\boldsymbol{\Lambda} && \text{(S.12.74)}\\ \mathbf{S}_a^{-1} &= \frac{\beta}{1+\beta}(\mathbf{x}-\boldsymbol{\mu}_0)(\mathbf{x}-\boldsymbol{\mu}_0)^t && \text{(S.12.75)} \end{aligned}$$

と置くと，式 (S.12.66) より

$$
\begin{aligned}
&\mathcal{N}(\mathbf{x};\boldsymbol{\mu},\boldsymbol{\Lambda}^{-1})\cdot\mathcal{N}\left(\boldsymbol{\mu};\boldsymbol{\mu}_0,(\beta\boldsymbol{\Lambda})^{-1}\right)\\
&=\frac{\beta^{d/2}|\boldsymbol{\Lambda}|}{(2\pi)^d}\cdot\exp\left[-\frac{1}{2}\mathrm{tr}\left(F_1\boldsymbol{\Lambda}\right)\right]\\
&=\frac{\beta^{d/2}|\boldsymbol{\Lambda}|}{(2\pi)^d}\cdot\exp\left[-\frac{1}{2}\mathrm{tr}\left((\boldsymbol{\mu}-\boldsymbol{\mu}_a)(\boldsymbol{\mu}-\boldsymbol{\mu}_a)^t\boldsymbol{\Lambda}_a\right)\right]\cdot\exp\left[-\frac{1}{2}\mathrm{tr}\left(\mathbf{S}_a^{-1}\boldsymbol{\Lambda}\right)\right]\\
&=\frac{\beta^{d/2}|\boldsymbol{\Lambda}|}{(2\pi)^d}\cdot\exp\left[-\frac{1}{2}(\boldsymbol{\mu}-\boldsymbol{\mu}_a)^t\boldsymbol{\Lambda}_a(\boldsymbol{\mu}-\boldsymbol{\mu}_a)\right]\cdot\exp\left[-\frac{1}{2}\mathrm{tr}\left(\mathbf{S}_a^{-1}\boldsymbol{\Lambda}\right)\right]\\
&=\left(\frac{\beta}{2(1+\beta)\pi}\right)^{d/2}|\boldsymbol{\Lambda}|^{1/2}\cdot\mathcal{N}\left(\boldsymbol{\mu};\boldsymbol{\mu}_a,\boldsymbol{\Lambda}_a^{-1}\right)\cdot\exp\left[-\frac{1}{2}\mathrm{tr}\left(\mathbf{S}_a^{-1}\boldsymbol{\Lambda}\right)\right]
\end{aligned}
\tag{S.12.76}
$$

が得られる．

【演習問題 12.5】

式 (12.35) 左辺の $p(\mathbf{x}_k|\boldsymbol{\theta}_{new})$，$G_0(\boldsymbol{\theta}_{new})$ は，それぞれ式 (12.27)，式 (12.32) で示したように

$$
\begin{aligned}
p(\mathbf{x}_k|\boldsymbol{\theta}_{new})&=\mathcal{N}(\mathbf{x}_k;\boldsymbol{\mu},\boldsymbol{\Lambda}^{-1})\\
G_0(\boldsymbol{\theta}_{new})&=\mathcal{N}\left(\boldsymbol{\mu};\boldsymbol{\mu}_0,(\beta\boldsymbol{\Lambda})^{-1}\right)\cdot\mathcal{W}(\boldsymbol{\Lambda};\nu,\mathbf{S})
\end{aligned}
$$

である．ここで

$$
\boldsymbol{\theta}_{new}=(\boldsymbol{\mu},\boldsymbol{\Lambda}^{-1}) \tag{S.12.77}
$$

である．したがって，**演習問題 12.4** の式 (S.12.76) を用いて

$$
\begin{aligned}
&p(\mathbf{x}_k|\boldsymbol{\theta}_{new})\cdot G_0(\boldsymbol{\theta}_{new})\\
&=\mathcal{N}(\mathbf{x}_k;\boldsymbol{\mu},\boldsymbol{\Lambda}^{-1})\cdot\mathcal{N}\left(\boldsymbol{\mu};\boldsymbol{\mu}_0,(\beta\boldsymbol{\Lambda})^{-1}\right)\cdot\mathcal{W}(\boldsymbol{\Lambda};\nu,\mathbf{S})\\
&=\left(\frac{\beta}{2(1+\beta)\pi}\right)^{d/2}|\boldsymbol{\Lambda}|^{1/2}\cdot\mathcal{N}\left(\boldsymbol{\mu};\boldsymbol{\mu}_a,\boldsymbol{\Lambda}_a^{-1}\right)\cdot\exp\left[-\frac{1}{2}\mathrm{tr}\left(\mathbf{S}_a^{-1}\boldsymbol{\Lambda}\right)\right]\\
&\qquad\cdot\mathcal{W}(\boldsymbol{\Lambda};\nu,\mathbf{S})
\end{aligned}
\tag{S.12.78}
$$

と書ける．ただし，

$$\boldsymbol{\mu}_a = \frac{\mathbf{x}_k + \beta\boldsymbol{\mu}_0}{1+\beta} \tag{S.12.79}$$

$$\boldsymbol{\Lambda}_a = (1+\beta)\boldsymbol{\Lambda} \tag{S.12.80}$$

$$\mathbf{S}_a^{-1} = \frac{\beta}{1+\beta}(\mathbf{x}_k - \boldsymbol{\mu}_0)(\mathbf{x}_k - \boldsymbol{\mu}_0)^t \tag{S.12.81}$$

である．上式のウィシャート分布関数 $\mathcal{W}(\boldsymbol{\Lambda};\nu,\mathbf{S})$ は，式 (12.30) より

$$\mathcal{W}(\boldsymbol{\Lambda};\nu,\mathbf{S}) = |\boldsymbol{\Lambda}|^{(\nu-d-1)/2} \cdot \exp\left[-\frac{1}{2}\mathrm{tr}\left(\mathbf{S}^{-1}\boldsymbol{\Lambda}\right)\right] \Big/ C_1 \tag{S.12.82}$$

と書ける．ただし，C_1 は $\boldsymbol{\Lambda}$ に依存しない正規化定数で，次式で表される．

$$C_1 = 2^{\nu d/2} \cdot \pi^{d(d-1)/4} \cdot |\mathbf{S}|^{\nu/2} \cdot \prod_{j=1}^{d} \Gamma\left(\frac{\nu+1-j}{2}\right) \tag{S.12.83}$$

以上より

$$\begin{aligned}
& p(\mathbf{x}_k|\boldsymbol{\theta}_{new}) \cdot G_0(\boldsymbol{\theta}_{new}) \\
&= \left(\frac{\beta}{2(1+\beta)\pi}\right)^{d/2} \cdot \mathcal{N}\left(\boldsymbol{\mu};\boldsymbol{\mu}_a,\boldsymbol{\Lambda}_a^{-1}\right) \\
&\qquad \cdot |\boldsymbol{\Lambda}|^{1/2} \cdot \exp\left[-\frac{1}{2}\mathrm{tr}\left(\mathbf{S}_a^{-1}\boldsymbol{\Lambda}\right)\right] \cdot \mathcal{W}(\boldsymbol{\Lambda};\nu,\mathbf{S}) \\
&= \left(\frac{\beta}{2(1+\beta)\pi}\right)^{d/2} \cdot \mathcal{N}\left(\boldsymbol{\mu};\boldsymbol{\mu}_a,\boldsymbol{\Lambda}_a^{-1}\right) \\
&\qquad \cdot |\boldsymbol{\Lambda}|^{(\nu-d)/2} \cdot \exp\left[-\frac{1}{2}\mathrm{tr}\left(\mathbf{S}^{-1}\boldsymbol{\Lambda} + \mathbf{S}_a{}^{-1}\boldsymbol{\Lambda}\right)\right] \Big/ C_1 \\
&= \left(\frac{\beta}{2(1+\beta)\pi}\right)^{d/2} \cdot \mathcal{N}\left(\boldsymbol{\mu};\boldsymbol{\mu}_a,\boldsymbol{\Lambda}_a^{-1}\right) \\
&\qquad \cdot |\boldsymbol{\Lambda}|^{(\nu_b-d-1)/2} \cdot \exp\left[-\frac{1}{2}\mathrm{tr}\left(\mathbf{S}_b^{-1}\boldsymbol{\Lambda}\right)\right] \Big/ C_1
\end{aligned} \tag{S.12.84}$$

となる．上式で

$$\nu_b = \nu + 1 \tag{S.12.85}$$

$$\mathbf{S}_b^{-1} = \mathbf{S}^{-1} + \mathbf{S}_a^{-1} \tag{S.12.86}$$

と置いた．ここで，ν_b，$\mathbf{S}_b$ をパラメータとするウィシャート分布関数 $\mathcal{W}(\mathbf{\Lambda}; \nu_b, \mathbf{S}_b)$ を導入すると

$$\begin{aligned} & p(\mathbf{x}_k | \boldsymbol{\theta}_{new}) \cdot G_0(\boldsymbol{\theta}_{new}) \\ &= \left(\frac{\beta}{2(1+\beta)\pi} \right)^{d/2} \cdot \frac{C_2}{C_1} \cdot \mathcal{N}\left(\boldsymbol{\mu}; \boldsymbol{\mu}_a, \mathbf{\Lambda}_a^{-1}\right) \cdot \mathcal{W}(\mathbf{\Lambda}; \nu_b, \mathbf{S}_b) \end{aligned} \tag{S.12.87}$$

と書ける．上式の C_2 は，ウィシャート分布関数 $\mathcal{W}(\mathbf{\Lambda}; \nu_b, \mathbf{S}_b)$ の正規化定数であり，式 (S.12.83) と同様にして

$$C_2 = 2^{\nu_b d/2} \cdot \pi^{d(d-1)/4} \cdot |\mathbf{S}_b|^{\nu_b/2} \cdot \prod_{j=1}^{d} \Gamma\left(\frac{\nu_b + 1 - j}{2} \right) \tag{S.12.88}$$

$$= 2^{(\nu+1)d/2} \cdot \pi^{d(d-1)/4} \cdot |\mathbf{S}_b|^{(\nu+1)/2} \cdot \prod_{j=1}^{d} \Gamma\left(\frac{\nu + 2 - j}{2} \right) \tag{S.12.89}$$

で定義される．ここで

$$\begin{aligned} & \int \mathcal{N}\left(\boldsymbol{\mu}; \boldsymbol{\mu}_a, \mathbf{\Lambda}_a^{-1}\right) \cdot \mathcal{W}(\mathbf{\Lambda}; \nu_b, \mathbf{S}_b)\, d\boldsymbol{\theta}_{new} \\ &= \iint \mathcal{N}\left(\boldsymbol{\mu}; \boldsymbol{\mu}_a, \left((1+\beta)\mathbf{\Lambda}\right)^{-1}\right) \cdot \mathcal{W}(\mathbf{\Lambda}; \nu_b, \mathbf{S}_b)\, d\boldsymbol{\mu} d\mathbf{\Lambda} \\ &= \int \mathcal{W}(\mathbf{\Lambda}; \nu_b, \mathbf{S}_b) \int \mathcal{N}\left(\boldsymbol{\mu}; \boldsymbol{\mu}_a, \left((1+\beta)\mathbf{\Lambda}\right)^{-1}\right) d\boldsymbol{\mu} d\mathbf{\Lambda} \\ &= 1 \end{aligned} \tag{S.12.90}$$

であることを用いると，式 (S.12.87) の積分は

$$
\begin{aligned}
&\int p(\mathbf{x}_k|\boldsymbol{\theta}_{new}) \cdot G_0(\boldsymbol{\theta}_{new})\, d\boldsymbol{\theta}_{new} \\
&= \left(\frac{\beta}{2(1+\beta)\pi}\right)^{d/2} \cdot \frac{C_2}{C_1} && \text{(S.12.91)} \\
&= \left(\frac{\beta}{(1+\beta)\pi}\right)^{d/2} \cdot \frac{|\mathbf{S}_b|^{(\nu+1)/2} \cdot \prod_{j=1}^{d} \Gamma\left(\frac{\nu+2-j}{2}\right)}{|\mathbf{S}|^{\nu/2} \cdot \prod_{j=1}^{d} \Gamma\left(\frac{\nu+1-j}{2}\right)} && \text{(S.12.92)} \\
&= \left(\frac{\beta}{(1+\beta)\pi}\right)^{d/2} \cdot \frac{|\mathbf{S}_b|^{(\nu+1)/2} \cdot \Gamma\left(\frac{\nu+1}{2}\right)}{|\mathbf{S}|^{\nu/2} \cdot \Gamma\left(\frac{\nu+1-d}{2}\right)} && \text{(S.12.93)}
\end{aligned}
$$

となる．式 (S.12.92) の分母，分子に含まれるガンマ関数は，大部分が約分で消去され，式 (S.12.93) が得られる．ここで $\mathbf{S}_b$ は，式 (S.12.81)，(S.12.86) より

$$
\begin{aligned}
\mathbf{S}_b^{-1} &= \mathbf{S}^{-1} + \mathbf{S}_a^{-1} \\
&= \mathbf{S}^{-1} + \frac{\beta}{1+\beta}(\mathbf{x}_k - \boldsymbol{\mu}_0)(\mathbf{x}_k - \boldsymbol{\mu}_0)^t && \text{(S.12.94)}
\end{aligned}
$$

の逆行列として求められる．

【演習問題 12.6】

式 (12.15) の分子を $f_3(\boldsymbol{\theta}_i)$ とおき

$$
f_3(\boldsymbol{\theta}_i) = G_0(\boldsymbol{\theta}_i) \cdot \prod_{\mathbf{x}_k \in \omega_i} p(\mathbf{x}_k|\boldsymbol{\theta}_i) \tag{S.12.95}
$$

とすると，式 (12.15) は

$$
p\left(\boldsymbol{\theta}_i|\{\mathbf{x}_k;\ \mathbf{x}_k \in \omega_i\}\right) = \frac{f_3(\boldsymbol{\theta}_i)}{\int f_3(\boldsymbol{\theta}_i)\, d\boldsymbol{\theta}_i} \tag{S.12.96}
$$

となる．式 (12.27)，(12.32) より

$$
\begin{aligned}
p(\mathbf{x}_k|\boldsymbol{\theta}_i) &= \mathcal{N}(\mathbf{x}_k; \boldsymbol{\mu}_i, \boldsymbol{\Lambda}_i^{-1}) && \text{(S.12.97)} \\
G_0(\boldsymbol{\theta}_i) &= \mathcal{N}\left(\boldsymbol{\mu}_i; \boldsymbol{\mu}_0, (\beta\boldsymbol{\Lambda}_i)^{-1}\right) \cdot \mathcal{W}(\boldsymbol{\Lambda}_i; \nu, \mathbf{S}) && \text{(S.12.98)}
\end{aligned}
$$

であるので

$$
\begin{aligned}
f_3(\boldsymbol{\theta}_i) &= \mathcal{N}\left(\boldsymbol{\mu}_i; \boldsymbol{\mu}_0, (\beta\boldsymbol{\Lambda}_i)^{-1}\right) \cdot \mathcal{W}(\boldsymbol{\Lambda}_i; \nu, \mathbf{S}) \cdot \prod_{\mathbf{x}_k \in \omega_i} \mathcal{N}(\mathbf{x}_k; \boldsymbol{\mu}_i, \boldsymbol{\Lambda}_i^{-1}) \\
&= \left(\frac{\beta}{(2\pi)^{n_i+1}}\right)^{d/2} \cdot |\boldsymbol{\Lambda}_i|^{(n_i+1)/2} \\
&\qquad \cdot \exp\left[f_4(\boldsymbol{\theta}_i)\right] \cdot \mathcal{W}(\boldsymbol{\Lambda}_i; \nu, \mathbf{S})
\end{aligned}
\tag{S.12.99}
$$

と書ける．ここで

$$
\begin{aligned}
f_4(\boldsymbol{\theta}_i) = &-\frac{1}{2}\sum_{\mathbf{x}_k \in \omega_i} (\mathbf{x}_k - \boldsymbol{\mu}_i)^t \boldsymbol{\Lambda}_i (\mathbf{x}_k - \boldsymbol{\mu}_i) \\
&-\frac{1}{2}(\boldsymbol{\mu}_i - \boldsymbol{\mu}_0)^t \beta \boldsymbol{\Lambda}_i (\boldsymbol{\mu}_i - \boldsymbol{\mu}_0)
\end{aligned}
\tag{S.12.100}
$$

である．式 (A.3.2) を用いて上式を書き換えると

$$
\begin{aligned}
&f_4(\boldsymbol{\theta}_i) \\
&= -\frac{1}{2}\mathrm{tr}\left[\left(\sum_{\mathbf{x}_k \in \omega_i} (\mathbf{x}_k - \boldsymbol{\mu}_i)(\mathbf{x}_k - \boldsymbol{\mu}_i)^t + \beta(\boldsymbol{\mu}_i - \boldsymbol{\mu}_0)(\boldsymbol{\mu}_i - \boldsymbol{\mu}_0)^t\right)\boldsymbol{\Lambda}_i\right] \\
&= -\frac{1}{2}\mathrm{tr}\Bigg[\Bigg(\sum_{\mathbf{x}_k \in \omega_i} \mathbf{x}_k \mathbf{x}_k^t + (n_i + \beta)\boldsymbol{\mu}_i \boldsymbol{\mu}_i^t + \beta \boldsymbol{\mu}_0 \boldsymbol{\mu}_0^t \\
&\qquad -\boldsymbol{\mu}_i\Big(\sum_{\mathbf{x}_k \in \omega_i} \mathbf{x}_k^t + \beta\boldsymbol{\mu}_0^t\Big) - \Big(\sum_{\mathbf{x}_k \in \omega_i} \mathbf{x}_k + \beta\boldsymbol{\mu}_0\Big)\boldsymbol{\mu}_i^t\Bigg)\boldsymbol{\Lambda}_i\Bigg]
\end{aligned}
\tag{S.12.101}
$$

となる．ここで

$$
\bar{\mathbf{x}} = \frac{1}{n_i}\sum_{\mathbf{x}_k \in \omega_i} \mathbf{x}_k \tag{S.12.102}
$$

$$
\boldsymbol{\mu}_c = \frac{n_i \bar{\mathbf{x}} + \beta\boldsymbol{\mu}_0}{n_i + \beta} \tag{S.12.103}
$$

$$
\boldsymbol{\Lambda}_c = (n_i + \beta)\boldsymbol{\Lambda}_i \tag{S.12.104}
$$

と置くと

$$
\begin{aligned}
f_4(\boldsymbol{\theta}_i) = &-\frac{1}{2}\mathrm{tr}\left[(\boldsymbol{\mu}_i\boldsymbol{\mu}_i^t - \boldsymbol{\mu}_i\boldsymbol{\mu}_c^t - \boldsymbol{\mu}_c\boldsymbol{\mu}_i^t)\boldsymbol{\Lambda}_c\right] \\
&-\frac{1}{2}\mathrm{tr}\left[\left(\sum_{\mathbf{x}_k\in\omega_i}\mathbf{x}_k\mathbf{x}_k^t + \beta\boldsymbol{\mu}_0\boldsymbol{\mu}_0^t\right)\boldsymbol{\Lambda}_i\right] \\
= &-\frac{1}{2}(\boldsymbol{\mu}_i - \boldsymbol{\mu}_c)^t\boldsymbol{\Lambda}_c(\boldsymbol{\mu}_i - \boldsymbol{\mu}_c) \\
&-\frac{1}{2}\mathrm{tr}\left[\left(\sum_{\mathbf{x}_k\in\omega_i}\mathbf{x}_x\mathbf{x}_k^t + \beta\boldsymbol{\mu}_0\boldsymbol{\mu}_0^t - (n_i+\beta)\boldsymbol{\mu}_c\boldsymbol{\mu}_c^t\right)\boldsymbol{\Lambda}_i\right] \\
= &-\frac{1}{2}(\boldsymbol{\mu}_i - \boldsymbol{\mu}_c)^t\boldsymbol{\Lambda}_c(\boldsymbol{\mu}_i - \boldsymbol{\mu}_c) - \frac{1}{2}\mathrm{tr}\left(\mathbf{S}_c^{-1}\boldsymbol{\Lambda}_i\right)
\end{aligned}
\tag{S.12.105}
$$

となる．ここで $\mathbf{S}_c^{-1}$ は

$$
\begin{aligned}
&\mathbf{S}_c^{-1} \\
&= \sum_{\mathbf{x}_k\in\omega_i}\mathbf{x}_x\mathbf{x}_k^t + \beta\boldsymbol{\mu}_0\boldsymbol{\mu}_0^t - (n_i+\beta)\boldsymbol{\mu}_c\boldsymbol{\mu}_c^t \\
&= \sum_{\mathbf{x}_k\in\omega_i}\mathbf{x}_k\mathbf{x}_k^t - n_i\bar{\mathbf{x}}\bar{\mathbf{x}}^t + \frac{n_i\beta}{n_i+\beta}(\bar{\mathbf{x}} - \boldsymbol{\mu}_0)(\bar{\mathbf{x}} - \boldsymbol{\mu}_0)^t \\
&= \sum_{\mathbf{x}_k\in\omega_i}(\mathbf{x}_k - \bar{\mathbf{x}})(\mathbf{x}_k - \bar{\mathbf{x}})^t + \frac{n_i\beta}{n_i+\beta}(\bar{\mathbf{x}} - \boldsymbol{\mu}_0)(\bar{\mathbf{x}} - \boldsymbol{\mu}_0)^t
\end{aligned}
\tag{S.12.106}
$$

である．式 (S.12.105) を式 (S.12.99) に代入し，式 (S.12.104) を用いると

$$
\begin{aligned}
f_3(\boldsymbol{\theta}_i) &= \left(\frac{\beta}{(2\pi)^{n_i+1}}\right)^{d/2}\cdot\frac{(2\pi)^{d/2}}{|\boldsymbol{\Lambda}_c|^{1/2}}\mathcal{N}(\boldsymbol{\mu}_i;\boldsymbol{\mu}_c,\boldsymbol{\Lambda}_c^{-1}) \\
&\quad\cdot|\boldsymbol{\Lambda}_i|^{(n_i+1)/2}\exp\left[-\frac{1}{2}\mathrm{tr}\left(\mathbf{S}_c^{-1}\boldsymbol{\Lambda}_i\right)\right]\cdot\mathcal{W}(\boldsymbol{\Lambda}_i;\nu,\mathbf{S}) \\
&= \left(\frac{\beta}{(n_i+\beta)\,(2\pi)^{n_i}}\right)^{d/2}\cdot\mathcal{N}(\boldsymbol{\mu}_i;\boldsymbol{\mu}_c,\boldsymbol{\Lambda}_c^{-1}) \\
&\quad\cdot|\boldsymbol{\Lambda}_i|^{n_i/2}\cdot\exp\left[-\frac{1}{2}\mathrm{tr}\left(\mathbf{S}_c^{-1}\boldsymbol{\Lambda}_i\right)\right]\cdot\mathcal{W}(\boldsymbol{\Lambda}_i;\nu,\mathbf{S})
\end{aligned}
\tag{S.12.107}
$$

が得られる．ここで定数 C_0 を

$$
C_0 = \left(\frac{\beta}{(n_i+\beta)\,(2\pi)^{n_i}}\right)^{d/2}
\tag{S.12.108}
$$

と置くと

$$\begin{aligned}
&f_3(\boldsymbol{\theta}_i)\\
&= C_0 \cdot \mathcal{N}(\boldsymbol{\mu}_i; \boldsymbol{\mu}_c, \boldsymbol{\Lambda}_c^{-1}) \cdot |\boldsymbol{\Lambda}_i|^{n_i/2} \cdot \exp\left[-\frac{1}{2}\mathrm{tr}\left(\mathbf{S}_c^{-1}\boldsymbol{\Lambda}_i\right)\right]\\
&\quad \cdot \frac{1}{C_3}|\boldsymbol{\Lambda}_i|^{(\nu-d-1)/2} \exp\left[-\frac{1}{2}\mathrm{tr}\left(\mathbf{S}^{-1}\boldsymbol{\Lambda}_i\right)\right]\\
&= \frac{C_0}{C_3} \cdot \mathcal{N}(\boldsymbol{\mu}_i; \boldsymbol{\mu}_c, \boldsymbol{\Lambda}_c^{-1})\\
&\quad \cdot |\boldsymbol{\Lambda}_i|^{(n_i+\nu-d-1)/2} \exp\left[-\frac{1}{2}\mathrm{tr}\Big(\left(\mathbf{S}^{-1} + \mathbf{S}_c^{-1}\right)\boldsymbol{\Lambda}_i\Big)\right] \qquad \text{(S.12.109)}
\end{aligned}$$

が得られる．上式で C_3 はウィシャート分布 $\mathcal{W}(\boldsymbol{\Lambda}_i; \nu, \mathbf{S})$ の正規化定数である．ここで

$$\nu_c = \nu + n_i \qquad \text{(S.12.110)}$$

$$\mathbf{S}_q^{-1} = \mathbf{S}^{-1} + \mathbf{S}_c^{-1} \qquad \text{(S.12.111)}$$

と置くと

$$\begin{aligned}
f_3(\boldsymbol{\theta}_i) &= \frac{C_0}{C_3} \cdot \mathcal{N}(\boldsymbol{\mu}_i; \boldsymbol{\mu}_c, \boldsymbol{\Lambda}_c^{-1}) \cdot |\boldsymbol{\Lambda}_i|^{(\nu_c-d-1)/2} \cdot \exp\left[-\frac{1}{2}\mathrm{tr}\left(\mathbf{S}_q^{-1}\boldsymbol{\Lambda}_i\right)\right]\\
&= \frac{C_0 \cdot C_4}{C_3} \cdot \mathcal{N}(\boldsymbol{\mu}_i; \boldsymbol{\mu}_c, \boldsymbol{\Lambda}_c^{-1}) \cdot \mathcal{W}(\boldsymbol{\Lambda}_i; \nu_c, \mathbf{S}_q) \qquad \text{(S.12.112)}
\end{aligned}$$

が得られる．上式で C_4 はウィシャート分布 $\mathcal{W}(\boldsymbol{\Lambda}_i; \nu_c, \mathbf{S}_q)$ の正規化定数である．この結果を式 (S.12.96) に代入し

$$\begin{aligned}
&\int \mathcal{N}(\boldsymbol{\mu}_i; \boldsymbol{\mu}_c, \boldsymbol{\Lambda}_c^{-1}) \cdot \mathcal{W}(\boldsymbol{\Lambda}_i; \nu_c, \mathbf{S}_q)\, d\boldsymbol{\theta}_i\\
&= \int \mathcal{W}(\boldsymbol{\Lambda}_i; \nu_c, \mathbf{S}_q) \left(\int \mathcal{N}(\boldsymbol{\mu}_i; \boldsymbol{\mu}_c, \boldsymbol{\Lambda}_c^{-1})\, d\boldsymbol{\mu}_i\right) d\boldsymbol{\Lambda}_i\\
&= 1 \qquad \text{(S.12.113)}
\end{aligned}$$

であることを用いることにより

$$p\left(\boldsymbol{\theta}_i | \{\mathbf{x}_k;\ \mathbf{x}_k \in \omega_i\}\right) = \mathcal{N}(\boldsymbol{\mu}_i; \boldsymbol{\mu}_c, \boldsymbol{\Lambda}_c^{-1}) \cdot \mathcal{W}(\boldsymbol{\Lambda}_i; \nu_c, \mathbf{S}_q) \qquad \text{(S.12.114)}$$

が得られる．

【演習問題 12.7】

求めるべき積分を I と置き，式 (12.44) の左辺を再掲すると

$$I = \int p(\mathbf{x}_k|\boldsymbol{\theta}_i) \cdot p(\boldsymbol{\theta}_i|\mathbf{x}_{-k})\, d\boldsymbol{\theta}_i \tag{S.12.115}$$

である．上式の $p(\boldsymbol{\theta}_i|\mathbf{x}_{-k})$ の算出には，**演習問題 12.6** の結果が利用できる．すなわち，式 (S.12.114) の左辺 $p\,(\boldsymbol{\theta}_i|\{\mathbf{x}_k;\ \mathbf{x}_k \in \omega_i\})$ を $p\,(\boldsymbol{\theta}_i|\{\mathbf{x}_j;\ \mathbf{x}_j \in \omega_i, j \neq k\})$ に置き換え，$\mathbf{S}_q$，$\bar{\mathbf{x}}$，$\boldsymbol{\mu}_c$ を求める際に用いた $\sum_{\mathbf{x}_k \in \omega_i} \mathbf{x}_k$ を $\sum_{\substack{\mathbf{x}_j \in \omega_i \\ j \neq k}} \mathbf{x}_j$ に置き換え，n_i を n'_i とすればよい．ただし，n'_i は $\mathbf{x}_k$ を除いたときのクラスタ ω_i のパターン数である．このように置き換えた $\mathbf{S}_q$，$\bar{\mathbf{x}}$，$\boldsymbol{\mu}_c$ を，それぞれ $\mathbf{S}'_q$，$\bar{\mathbf{x}}'$，$\boldsymbol{\mu}'_c$ と記すと，以下が得られる．

$$\begin{aligned}{\mathbf{S}'_q}^{-1} &= \mathbf{S}^{-1} + \sum_{\substack{\mathbf{x}_j \in \omega_i \\ j \neq k}} (\mathbf{x}_j - \bar{\mathbf{x}}')(\mathbf{x}_j - \bar{\mathbf{x}}')^t \\ &\qquad + \frac{n'_i \beta}{n'_i + \beta} (\bar{\mathbf{x}}' - \boldsymbol{\mu}_0)(\bar{\mathbf{x}}' - \boldsymbol{\mu}_0)^t \end{aligned} \tag{S.12.116}$$

$$\bar{\mathbf{x}}' = \frac{1}{n'_i} \sum_{\substack{\mathbf{x}_j \in \omega_i \\ j \neq k}} \mathbf{x}_j \tag{S.12.117}$$

$$\boldsymbol{\mu}'_c = \frac{n'_i \bar{\mathbf{x}}' + \beta \boldsymbol{\mu}_0}{n'_i + \beta} \tag{S.12.118}$$

式 (S.12.115) の被積分関数を $f_5(\boldsymbol{\theta}_i)$ と置くと，式 (S.12.114) より

$$\begin{aligned}&f_5(\boldsymbol{\theta}_i) \\ &= p(\mathbf{x}_k|\boldsymbol{\theta}_i) p(\boldsymbol{\theta}_i|\mathbf{x}_{-k}) \\ &= \mathcal{N}(\mathbf{x}_k; \boldsymbol{\mu}_i, \boldsymbol{\Lambda}_i^{-1}) \cdot \mathcal{N}(\boldsymbol{\mu}_i; \boldsymbol{\mu}'_c, \boldsymbol{\Lambda}_c^{-1}) \cdot \mathcal{W}(\boldsymbol{\Lambda}_i; \nu_c, \mathbf{S}'_q)\end{aligned} \tag{S.12.119}$$

が得られる．上式における二つの正規分布の積に対しては，**演習問題 12.4** の結果が使える．すなわち，式 (S.12.104) より，$\boldsymbol{\Lambda}_c = (n'_i + \beta)\boldsymbol{\Lambda}_i$ が成り立つので，

式 (S.12.76) の左辺を以下のごとく置き換えればよい.

$$\left.\begin{array}{lcl}\beta & \to & n_i' + \beta \\ \mathbf{x} & \to & \mathbf{x}_k \\ \boldsymbol{\mu} & \to & \boldsymbol{\mu}_i \\ \boldsymbol{\mu}_0 & \to & \boldsymbol{\mu}_c' \\ \boldsymbol{\Lambda} & \to & \boldsymbol{\Lambda}_i\end{array}\right\} \tag{S.12.120}$$

その結果

$$\begin{aligned}&\mathcal{N}(\mathbf{x}_k; \boldsymbol{\mu}_i, \boldsymbol{\Lambda}_i^{-1}) \cdot \mathcal{N}(\boldsymbol{\mu}_i; \boldsymbol{\mu}_c', \boldsymbol{\Lambda}_c^{-1}) \\ &= C_5 \cdot |\boldsymbol{\Lambda}_i|^{1/2} \cdot \mathcal{N}\left(\boldsymbol{\mu}_i; \boldsymbol{\mu}_e, \boldsymbol{\Lambda}_e^{-1}\right) \cdot \exp\left[-\frac{1}{2}\mathrm{tr}\left(\mathbf{S}_e^{-1}\boldsymbol{\Lambda}_i\right)\right]\end{aligned} \tag{S.12.121}$$

が得られる. ただし

$$C_5 = \left(\frac{n_i' + \beta}{2(1 + n_i' + \beta)\pi}\right)^{d/2} \tag{S.12.122}$$

$$\boldsymbol{\mu}_e = \frac{\mathbf{x}_k + (n_i' + \beta)\boldsymbol{\mu}_c'}{1 + n_i' + \beta} \tag{S.12.123}$$

$$\boldsymbol{\Lambda}_e = (1 + n_i' + \beta)\boldsymbol{\Lambda}_i \tag{S.12.124}$$

$$\mathbf{S}_e^{-1} = \frac{n_i' + \beta}{1 + n_i' + \beta}(\mathbf{x}_k - \boldsymbol{\mu}_c')(\mathbf{x}_k - \boldsymbol{\mu}_c')^t \tag{S.12.125}$$

である. 式 (S.12.123), (S.12.124), (S.12.125) は, 式 (S.12.72), (S.12.74), (S.12.75) にそれぞれ対応している．式 (S.12.121) の結果より，$f_5(\boldsymbol{\theta}_i)$ は

$$\begin{aligned} f_5(\boldsymbol{\theta}_i) &= C_5 \cdot \mathcal{N}\left(\boldsymbol{\mu}_i; \boldsymbol{\mu}_e, \boldsymbol{\Lambda}_e^{-1}\right) \\ &\quad \cdot |\boldsymbol{\Lambda}_i|^{1/2} \cdot \exp\left[-\frac{1}{2}\mathrm{tr}\left(\mathbf{S}_e^{-1}\boldsymbol{\Lambda}_i\right)\right] \cdot \mathcal{W}(\boldsymbol{\Lambda}_i; \nu_c, \mathbf{S}_q') \\ &= C_5 \cdot \mathcal{N}\left(\boldsymbol{\mu}_i; \boldsymbol{\mu}_e, \boldsymbol{\Lambda}_e^{-1}\right) \\ &\quad \cdot |\boldsymbol{\Lambda}_i|^{1/2} \cdot \exp\left[-\frac{1}{2}\mathrm{tr}\left(\mathbf{S}_e^{-1}\boldsymbol{\Lambda}_i\right)\right] \\ &\quad \cdot \frac{1}{C_6} \cdot |\boldsymbol{\Lambda}_i|^{(\nu_c-d-1)/2} \cdot \exp\left[-\frac{1}{2}\mathrm{tr}\left({\mathbf{S}_q'}^{-1}\boldsymbol{\Lambda}_i\right)\right] \\ &= \frac{C_5}{C_6} \cdot \mathcal{N}\left(\boldsymbol{\mu}_i; \boldsymbol{\mu}_e, \boldsymbol{\Lambda}_e^{-1}\right) \\ &\quad \cdot |\boldsymbol{\Lambda}_i|^{(\nu_c-d)/2} \cdot \exp\left[-\frac{1}{2}\mathrm{tr}\left(\left({\mathbf{S}_q'}^{-1} + \mathbf{S}_e^{-1}\right)\boldsymbol{\Lambda}_i\right)\right] \end{aligned} \tag{S.12.126}$$

と書ける．上式で C_6 はウィシャート分布 $\mathcal{W}(\boldsymbol{\Lambda}_i; \nu_c, \mathbf{S}_q')$ の正規化定数で

$$C_6 = 2^{\nu_c d/2} \cdot \pi^{d(d-1)/4} \cdot |\mathbf{S}_q'|^{\nu_c/2} \cdot \prod_{j=1}^{d} \Gamma\left(\frac{\nu_c + 1 - j}{2}\right) \tag{S.12.127}$$

である．ここで

$$\nu_r = \nu_c + 1 \tag{S.12.128}$$

$$\mathbf{S}_r^{-1} = {\mathbf{S}_q'}^{-1} + \mathbf{S}_e^{-1} \tag{S.12.129}$$

と置くと，$f_5(\boldsymbol{\theta}_i)$ は

$$f_5(\boldsymbol{\theta}_i) = C_5 \cdot \frac{C_7}{C_6} \cdot \mathcal{N}\left(\boldsymbol{\mu}_i; \boldsymbol{\mu}_e, \boldsymbol{\Lambda}_e^{-1}\right) \cdot \mathcal{W}(\boldsymbol{\Lambda}_i; \nu_r, \mathbf{S}_r) \tag{S.12.130}$$

となる．上式で C_7 はウィシャート分布 $\mathcal{W}(\boldsymbol{\Lambda}_i; \nu_r, \mathbf{S}_r)$ の正規化定数で

$$C_7 = 2^{\nu_r d/2} \cdot \pi^{d(d-1)/4} \cdot |\mathbf{S}_r|^{\nu_r/2} \cdot \prod_{j=1}^{d} \Gamma\left(\frac{\nu_r + 1 - j}{2}\right) \tag{S.12.131}$$

である．式 (S.12.127)，(S.12.131)，(S.12.110)，(S.12.128) を用いると

$$
\begin{aligned}
\frac{C_7}{C_6} &= 2^{(\nu_r-\nu_c)d/2} \cdot \frac{|\mathbf{S}_r|^{\nu_r/2} \cdot \prod_{j=1}^{d} \Gamma\left(\dfrac{\nu_r+1-j}{2}\right)}{|\mathbf{S}'_q|^{\nu_c/2} \cdot \prod_{j=1}^{d} \Gamma\left(\dfrac{\nu_c+1-j}{2}\right)} \qquad \text{(S.12.132)} \\
&= 2^{d/2} \cdot \frac{|\mathbf{S}_r|^{(\nu+n'_i+1)/2} \cdot \Gamma\left(\dfrac{\nu+n'_i+1}{2}\right)}{|\mathbf{S}'_q|^{(\nu+n'_i)/2} \cdot \Gamma\left(\dfrac{\nu+n'_i+1-d}{2}\right)} \qquad \text{(S.12.133)}
\end{aligned}
$$

である．上式は，式 (S.12.132) の分母，分子に含まれるガンマ関数を約分した結果である．

求めるべき積分 I は，式 (S.12.122)，(S.12.130)，(S.12.133) より

$$
\begin{aligned}
I &= \int p(\mathbf{x}_k|\boldsymbol{\theta}_i) \cdot p(\boldsymbol{\theta}_i|\mathbf{x}_{-k})\, d\boldsymbol{\theta}_i \\
&= \int f_5(\boldsymbol{\theta}_i)\, d\boldsymbol{\theta}_i \\
&= C_5 \cdot \frac{C_7}{C_6} \\
&= \left(\frac{n'_i+\beta}{(1+n'_i+\beta)\pi}\right)^{d/2} \cdot \frac{|\mathbf{S}_r|^{(\nu+n'_i+1)/2} \cdot \Gamma\left(\dfrac{\nu+n'_i+1}{2}\right)}{|\mathbf{S}'_q|^{(\nu+n'_i)/2} \cdot \Gamma\left(\dfrac{\nu+n'_i+1-d}{2}\right)}
\end{aligned}
\qquad \text{(S.12.134)}
$$

となり，式 (12.44) が得られる．ここで $\mathbf{S}'_q$ は，式 (S.12.106)，(S.12.111) より

$$
\begin{aligned}
{\mathbf{S}'_q}^{-1} &= \mathbf{S}^{-1} + \mathbf{S}_c^{-1} \\
&= \mathbf{S}^{-1} + \sum_{\substack{\mathbf{x}_j \in \omega_i \\ j \neq k}} (\mathbf{x}_j - \bar{\mathbf{x}}')(\mathbf{x}_j - \bar{\mathbf{x}}')^t \\
&\qquad + \frac{n'_i\beta}{n'_i+\beta}(\bar{\mathbf{x}}' - \boldsymbol{\mu}_0)(\bar{\mathbf{x}}' - \boldsymbol{\mu}_0)^t
\end{aligned}
\qquad \text{(S.12.135)}
$$

として求められる．また $\mathbf{S}_r$ は，式 (S.12.125)，(S.12.129) より

$$\begin{aligned}\mathbf{S}_r^{-1} &= \mathbf{S}_q'^{-1} + \mathbf{S}_e^{-1} \\ &= \mathbf{S}_q'^{-1} + \frac{n_i' + \beta}{1 + n_i' + \beta}(\mathbf{x}_k - \boldsymbol{\mu}_c')(\mathbf{x}_k - \boldsymbol{\mu}_c')^t \end{aligned} \tag{S.12.136}$$

として求められる．

【演習問題 12.8】

式 (12.49) の被積分関数は，**演習問題 12.6** の式 (S.12.95) で示した $f_3(\boldsymbol{\theta}_i)$ と一致する．したがって，式 (S.12.112)，(S.12.113) より

$$\begin{aligned}\int G_0(\boldsymbol{\theta}_i) \cdot \prod_{\mathbf{x}_k \in \omega_i} p(\mathbf{x}_k|\boldsymbol{\theta}_i) d\boldsymbol{\theta}_i &= \int f_3(\boldsymbol{\theta}_i) d\boldsymbol{\theta}_i \\ &= \frac{C_0 \cdot C_4}{C_3} \end{aligned} \tag{S.12.137}$$

が得られる．**演習問題 12.6** より，C_3，C_4 は，それぞれウィシャート分布 $\mathcal{W}(\boldsymbol{\Lambda}_i; \nu, \mathbf{S})$，$\mathcal{W}(\boldsymbol{\Lambda}_i; \nu_c, \mathbf{S}_q)$ の正規化定数であるので，

$$C_3 = 2^{\nu d/2} \cdot \pi^{d(d-1)/4} \cdot |\mathbf{S}|^{\nu/2} \cdot \prod_{j=1}^{d} \Gamma\left(\frac{\nu + 1 - j}{2}\right) \tag{S.12.138}$$

$$C_4 = 2^{\nu_c d/2} \cdot \pi^{d(d-1)/4} \cdot |\mathbf{S}_q|^{\nu_c/2} \cdot \prod_{j=1}^{d} \Gamma\left(\frac{\nu_c + 1 - j}{2}\right) \tag{S.12.139}$$

である．一方 C_0 は式 (S.12.108) で与えられ，ν_c は式 (S.12.110) で与えられて

いるので，これらを代入すると

$$
\begin{aligned}
&\int G_0(\boldsymbol{\theta}_i) \cdot \prod_{\mathbf{x}_k \in \omega_i} p(\mathbf{x}_k|\boldsymbol{\theta}_i) d\boldsymbol{\theta}_i \\
&= \frac{C_0 \cdot C_4}{C_3} \\
&= \left(\frac{\beta}{(n_i + \beta)\,\pi^{n_i}} \right)^{d/2} \\
&\qquad \cdot \frac{|\mathbf{S}_q|^{(\nu+n_i)/2} \cdot \prod_{j=1}^{d} \Gamma\left(\frac{\nu + n_i + 1 - j}{2} \right)}{|\mathbf{S}|^{\nu/2} \cdot \prod_{j=1}^{d} \Gamma\left(\frac{\nu + 1 - j}{2} \right)}
\end{aligned} \tag{S.12.140}
$$

が得られる．

第１３章

【演習問題 13.1】

$$
\begin{aligned}
&\prod_l \prod_{k' \neq k} P(R_{k'l}|\theta_{ij})^{\delta_{k'}(i) \cdot \delta_l(j)} \\
&= \prod_l \prod_{k' \neq k} \theta_{ij}^{R_{k'l}\delta_{k'}(i)\delta_l(j)} (1 - \theta_{ij})^{(1-R_{k'l})\delta_{k'}(i)\delta_l(j)} \\
&= \theta_{ij}^{\sum_l \sum_{k' \neq k} R_{k'l}\delta_{k'}(i)\delta_l(j)} (1 - \theta_{ij})^{\sum_l \sum_{k' \neq k} (1-R_{k'l})\delta_{k'}(i)\delta_l(j)}
\end{aligned} \tag{S.13.1}
$$

また，$p(\theta_{ij}) = \mathrm{Be}(\theta_{ij} : a, b) = \theta_{ij}^{a-1}(1-\theta_{ij})^{b-1}/B(a,b)$ に注意すると，式 (13.18) は次式のように計算できる．

$$
\begin{aligned}
&\prod_j \frac{\theta_{ij}^{\sum_l \sum_{k' \neq k} R_{k'l}\delta_{k'}(i)\delta_l(j)+a-1} (1 - \theta_{ij})^{\sum_l \sum_{k' \neq k} (1-R_{k'l})\delta_{k'}(i)\delta_l(j)+b-1}}{\int \theta_{ij}^{\sum_l \sum_{k' \neq k} R_{k'l}\delta_{k'}(i)\delta_l(j)+a-1} (1 - \theta_{ij})^{\sum_l \sum_{k' \neq k} (1-R_{k'l})\delta_{k'}(i)\delta_l(j)+b-1} d\theta_{ij}} \\
&= \prod_j \frac{\theta_{ij}^{\sum_l \sum_{k' \neq k} R_{k'l}\delta_{k'}(i)\delta_l(j)+a-1} (1 - \theta_{ij})^{\sum_l \sum_{k' \neq k} (1-R_{k'l})\delta_{k'}(i)\delta_l(j)+b-1}}{B(\sum_l \sum_{k' \neq k} R_{k'l}\delta_{k'}(i)\delta_l(j) + a, \sum_l \sum_{k' \neq k} (1 - R_{k'l})\delta_{k'}(i)\delta_l(j) + b)} \\
&\equiv \prod_j \mathrm{Be}(\theta_{ij}; n_{(-k,+)(i,j)} + a, \bar{n}_{(-k,+)(i,j)} + b)
\end{aligned} \tag{S.13.2}
$$

ただし，

$$
\begin{aligned}
n_{(-k,+)(i,j)} &= \sum_{l} \sum_{k' \neq k} R_{k'l} \delta_{k'}(i) \delta_l(j) \\
\bar{n}_{(-k,+)(i,j)} &= \sum_{l} \sum_{k' \neq k} (1 - R_{k'l}) \delta_{k'}(i) \delta_l(j)
\end{aligned}
\tag{S.13.3}
$$

【演習問題 13.2】

ω_i^1 が既存クラスタのとき，式 (13.11)，式 (13.19) を式 (13.9) に代入すると，

$$
\begin{aligned}
&P(\mathbf{R}_{k,+}|s_k^1 = \omega_i^1, \mathbf{s}^2, \mathbf{R}_{-k,+}) \\
&= \int \left\{ \prod_j \theta_{ij}^{n_{(k,+)(i,j)}} (1 - \theta_{ij})^{\bar{n}_{(k,+)(i,j)}} \right\} \\
&\quad \times \left\{ \prod_j \frac{\theta_{ij}^{n_{(-k,+)(i,j)}+a} (1 - \theta_{ij})^{\bar{n}_{(-k,+)(i,j)}+b}}{B(n_{(-k,+)(i,j)} + a, \bar{n}_{(-k,+)(i,j)} + b))} \right\} d\theta_{ij} \\
&= \prod_j \frac{\int \theta_{ij}^{n_{(+,+)(i,j)}+a} (1 - \theta_{ij})^{\bar{n}_{(+,+)(i,j)}} d\theta_{ij}}{B(n_{(-k,+)(i,j)} + a, \bar{n}_{(-k,+)(i,j)} + b)} \\
&= \prod_j \frac{B(n_{(+,+)(i,j)} + a, \bar{n}_{(+,+)(i,j)} + b)}{B(n_{(-k,+)(i,j)} + a, \bar{n}_{(-k,+)(i,j)} + b)}
\end{aligned}
\tag{S.13.4}
$$

上記の式変形では，

$$
n_{(k,+)(i,j)} + n_{(-k,+)(i,j)} = n_{(+,+)(i,j)} \tag{S.13.5}
$$

$$
\bar{n}_{(k,+)(i,j)} + \bar{n}_{(-k,+)(i,j)} = \bar{n}_{(+,+)(i,j)} \tag{S.13.6}
$$

および，ベータ関数の定義式：$B(a, b) = \int x^a (1 - x)^b dx$ を用いた．

【演習問題 13.3】

ω_i^1 が新規クラスタのとき，式 (13.11)，および，$G_0(\boldsymbol{\theta}_{i,+}) = \prod_j \mathrm{Be}(\theta_{i,j}; a, b)$

を式 (13.9) に代入すると，

$$\begin{aligned}
&P(\mathbf{R}_{k,+}|s_k^1 = \omega_i^1, \mathbf{s}^2, \mathbf{R}_{-k,+}) \\
&\quad = \prod_j \int \theta_{ij}^{n_{(k,+)(i,j)}+a-1}(1-\theta_{ij})^{\bar{n}_{(k,+)(i,j)}+b-1} d\theta_{ij}/B(a,b) \\
&\quad = \prod_j \frac{B\big(n_{(k,+)(i,j)}+a, \bar{n}_{(k,+)(i,j)}+b\big)}{B(a,b)}
\end{aligned} \tag{S.13.7}$$

【演習問題 13.4】

$s_l^2 \in \mathbf{s}^2$ のサンプリングのための事後確率に対応する式は，式 (13.26) を導出した時と同様の計算をすれば良いが，より簡単には，$P(s_k^1 = \omega_i^1|\mathbf{R}_{k,+}, \mathbf{s}_{-k}^1, \mathbf{s}^2, \mathbf{R}_{-k,+})$ と $P(s_l^2 = \omega_j^2|\mathbf{R}_{+,l}, \mathbf{s}_{-l}^2, \mathbf{s}^1, \mathbf{R}_{+,-l})$ の対応関係に注意すると，$P(s_l^2 = \omega_j^2|\mathbf{R}_{+,l}, \mathbf{s}_{-l}^2, \mathbf{s}^1, \mathbf{R}_{+,-l})$ は，式 (13.26) において，

$$\begin{aligned}
&n_i^1 \to n_j^2 \\
&n_{(-k,+)(i,j)} \to n_{(+,-l)(i,j)} \\
&\bar{n}_{(-k,+)(i,j)} \to \bar{n}_{(+,-l)(i,j)}
\end{aligned} \tag{S.13.8}$$

とすることで，式 (13.27) が得られる．