

『渡辺澄夫ベイズ理論 100 問』の解答例 (数理)

鈴木 讓

プログラムに関しては、解答例 (R プログラム).html および解答例 (Python プログラム).html を参照してください。

第 1 章 渡辺ベイズ理論入門

1. $u = r \cos \theta$, $v = r \sin \theta$ とおくと、Jacobian が r であるから、

$$\begin{aligned} \left\{ \int_0^\infty \exp\left(-\frac{u^2}{2}\right) du \right\}^2 &= \int_0^\infty \int_0^\infty \exp\left(-\frac{u^2+v^2}{2}\right) dudv = \int_0^{\pi/2} \int_0^\infty \exp\left(-\frac{r^2}{2}\right) r dr d\theta \\ &= \frac{\pi}{2} \left[-\exp\left(-\frac{r^2}{2}\right) \right]_0^\infty = \frac{\pi}{2} \end{aligned}$$

となる。 $\int_0^\infty \exp\left(-\frac{u^2}{2}\right) du = \sqrt{\frac{\pi}{2}}$ より、 $\int_{-\infty}^\infty \exp\left(-\frac{u^2}{2}\right) du = 2 \int_0^\infty \exp\left(-\frac{u^2}{2}\right) du = 2\sqrt{\frac{\pi}{2}} = \sqrt{2\pi}$ を得る。

2. $y = x - \mu$ とおくと、 $dy = dx$ であり、

$$\mathbb{E}[X] = \int_{-\infty}^\infty xq(x)dx = \int_{-\infty}^\infty (x - \mu)q(x)dx + \mu \int_{-\infty}^\infty q(x)dx = \int_{-\infty}^\infty yq(y + \mu)dy + \mu = \mu$$

が成り立つ。 $z = \frac{x - \mu}{\sigma}$ とおくと、 $\frac{dz}{dx} = \frac{1}{\sigma}$ であり、 $\int_{-\infty}^\infty \frac{z^2}{\sqrt{2\pi}} \exp\left(-\frac{z^2}{2}\right) dz = 1$ は、

$$\int_{-\infty}^\infty \left(\frac{x - \mu}{\sigma}\right)^2 \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{(x - \mu)^2}{2\sigma^2}\right) \frac{dx}{\sigma} = 1$$

とできる。これは、

$$\mathbb{V}[X] = \int_{-\infty}^\infty (x - \mu)^2 \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x - \mu)^2}{2\sigma^2}\right) dx = \sigma^2$$

を意味する。したがって、

$$\begin{aligned} 1 &= \int_{-\infty}^\infty \frac{z^2}{\sqrt{2\pi}} \exp\left(-\frac{z^2}{2}\right) dz = \int_{-\infty}^\infty \left(\frac{z^3}{3}\right)' \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{z^2}{2}\right) dz \\ &= \left[\left(\frac{z^3}{3}\right)' \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{z^2}{2}\right) \right]_{-\infty}^\infty - \int_{-\infty}^\infty \frac{z^3}{3} \cdot (-z) \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{z^2}{2}\right) dz \\ &= \frac{1}{3} \int_{-\infty}^\infty z^4 \cdot \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{z^2}{2}\right) dz \end{aligned}$$

および

$$\int_{-\infty}^{\infty} \left(\frac{x-\mu}{\sigma}\right)^4 \cdot \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) \frac{dx}{\sigma} = 3$$

とできる。これは、

$$\mathbb{E}[(X-\mu)^4] = \int_{-\infty}^{\infty} (x-\mu)^4 \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) dx = 3\sigma^4$$

を意味する。

3. (1.9) は標準正規分布の確率密度関数なので、 $\Theta = \mathbb{R}$ で積分して 1、また各 $\theta \in \Theta$ で非負の値をとる。また、 $p(x_1|\mu) \cdots p(x_n|\mu)\varphi(\mu)$ は、

$$\begin{aligned} & \left(\frac{1}{\sqrt{2\pi}}\right)^n \exp\left\{-\sum_{i=1}^n \left(\frac{x_i-\mu}{2}\right)^2\right\} \left(\frac{1}{\sqrt{2\pi}}\right) \exp\left(-\frac{\mu^2}{2}\right) \\ &= \left(\frac{1}{\sqrt{2\pi}}\right)^n \exp\left\{-\frac{1}{2}\sum_{i=1}^n x_i^2 + \frac{1}{2(n+1)}\left(\sum_{i=1}^n x_i\right)^2\right\} \\ &\times \frac{1}{\sqrt{n+1}} \frac{1}{\sqrt{2\pi/(n+1)}} \exp\left\{-\frac{n+1}{2}\left(\mu - \sum_{i=1}^n x_i\right)^2\right\} \end{aligned}$$

と書ける。右辺の前半は μ に関して定数となる。後半は平均 $\sum_{i=1}^n x_i$ 、分散 $(n+1)^{-1}$ の正規分布の確率密度関数であるから、 μ で $\Theta = \mathbb{R}$ の範囲で積分して 1 になる。したがって、右辺の前半が周辺尤度の (1.10)、右辺の後半が事後確率の (1.11) となる。

4. (1.13) は、 $\varphi(\theta) \geq 0$ および $\int_{\Theta} \varphi(\theta) d\theta = \int_0^1 1 d\theta = 1$ を満足する。また、 $k = \sum_{i=1}^n x_i$ として、 $p(x_1|\theta) \cdots p(x_n|\theta)\varphi(\theta)$ は

$$\theta^{x_1}(1-\theta)^{1-x_1} \cdots \theta^{x_n}(1-\theta)^{1-x_n} \cdot 1 = \theta^k(1-\theta)^{n-k} = \frac{(n-k)!k!}{(n+1)!} \cdot \frac{(n+1)!}{(n-k)!k!} \theta^k(1-\theta)^{n-k}$$

と書ける。右辺の前半は θ に関して定数となる。右辺の後半は θ で $\Theta = \mathbb{R}$ の範囲で積分して 1 になる。したがって、前半が周辺尤度の (1.14)、後半が事後確率の (1.15) になる。

7. $\Theta_* = \{\pm 1\}$ となり、正則ではない。 $p(\cdot|\mu=1) \neq p(\cdot|\mu=-1)$ より、同質ではない。 $q(\cdot) \neq p(\cdot|\mu=1)$ 、 $p(\cdot|\mu=-1)$ より、実現可能ではない。

$$\log \frac{p(X|\mu_*)}{p(X|\mu)} = -\frac{X-\mu_*}{2} + \frac{X-\mu}{2} = (\mu_* - \mu) \left(X - \frac{\mu_* + \mu}{2}\right)$$

この値の平均、2乗平均はそれぞれ $\frac{1}{2}(\mu^2 - \mu_*^2)$ 、 $(\mu_* - \mu)^2 \left\{1 + \left(\frac{\mu + \mu_*}{2}\right)^2\right\}$ となるので、(1.21) は任意の $\mu_* \in \Theta_*$ 、 $\mu \in \Theta$ について、

$$(\mu_* - \mu)^2 \left\{1 + \left(\frac{\mu + \mu_*}{2}\right)^2\right\} \leq c \cdot \frac{1}{2}(\mu^2 - \mu_*^2) \quad (1)$$

なる $c > 0$ が存在することを意味する。 $\mu_* = 1$ のとき、前者は μ の 2 次関数 (2 次の係数が正)、後者は μ の 4 次関数 (2 次の係数が正) であって、 $|\mu| \geq 1$ の範囲で (1) は満足されない。したがって、相対的に有限な分散をもたない。

8. Θ_* が単一の要素を含むので同質である。 $q(\cdot) \neq p(\cdot|\mu = 1)$ より、実現可能ではない。 $\mu = 1$ は Θ の境界にあるので、正則ではない。また、 $\mu_* = 1$ として前問の μ の関数の値を比較すると、 $c \geq 13/6$ とすれば $1 \leq \mu \leq 2$ の範囲で (1) が成立する。したがって、相対的に有限な分散をもつ。
9. 真の分布が t 分布で統計モデルが正規分布なので、実現可能ではない。 $\mu = 0$ のみで $D(q||p(\cdot|\mu))$ が最小となり、 $0 \in (-\epsilon, \epsilon) \subseteq \Theta$ なる $\epsilon > 0$ が存在する。また、Fisher 情報量 (スカラー) はゼロにならないので、正則である。 $\mu_* = 0$ であるので、(1) は

$$\mu^2 \left(1 + \frac{\mu^2}{4}\right) \leq c \cdot \frac{1}{2}\mu^2$$

であり、 $c \geq 5/2$ とすれば $-1 \leq \mu \leq 1$ の範囲で成立する。したがって、相対的に有限な分散をもつ。

10. $(\alpha, \beta) = (\pm \frac{1}{\sigma}, \pm \frac{\mu}{\sigma})$ (複号同順) のときだけ $D(q||p) = 0$ となる。したがって、実現可能ではあるが、 Θ_* が複数の要素を含み、正則ではない。 $p(x|\frac{1}{\sigma}, \frac{\mu}{\sigma}) = p(x|-\frac{1}{\sigma}, -\frac{\mu}{\sigma})$ となるため、 $\Theta_* = \{(\frac{1}{\sigma}, \frac{\mu}{\sigma}), (-\frac{1}{\sigma}, -\frac{\mu}{\sigma})\}$ は同質である。実現可能であるので、命題 2 より相対的に有限な分散をもつ。

11. $(\alpha, \beta) = (1, 0)$ のとき $D(q||p) = 0$ となり、実現可能である。また、 $\left[\frac{\partial^2 D(q||p)}{\partial \theta_i \partial \theta_j} \Big|_{\theta=\theta_*} \right] = \begin{bmatrix} 2 & 0 \\ 0 & 1 \end{bmatrix}$ の

固有値がともに正となる。 $(1, 0)$ を中心とした半径 $\epsilon > 0$ の円を十分小さくとると、その円は Θ に含まれる。したがって、正則である。命題 2 より、相対的に有限な分散をもつ。

12. $a_* = 0$ もしくは $b_* = 0$ のとき、 $K(a, b) = -\int_{\mathcal{X}} \log\{1 + a\{\exp(bx - b^2/2) - 1\}\} = 0$ となる。このとき、「 $a = 0$ で b が任意」または「 $b = 0$ で a が任意」である。いずれにせよ、2 個以上の組合せとなる。

13. $\binom{n}{k} = B(k+1, n-k+1)$ が成立するので、 $l = k+1, m = n-k+1$ とおくと、 $k = l-1,$

$$n = m + k - 1 = m + l - 1 - 1 = l + m - 2 \text{ となり、} B(l, m) = \binom{l+m-2}{l-1} \text{ が得られる。}$$

第 2 章 MCMC と Stan

14. (2.7) は、ある時刻で θ のときに、 $\theta' \neq \theta$ であれば、 θ' の条件付き密度関数は $s(\theta'|\theta) \min\{1, \exp(-H(\theta') + H(\theta))\}$ となる。 $\theta' = \theta$ であれば、その条件付き密度関数の値は $1 - Q(\theta)$ となる ($\theta' = \theta$ が非ゼロの測度の値をもつ)。(2.8) は、 $\theta' \neq \theta$ であれば、 $s(\theta|\theta') = s(\theta'|\theta)$ であるから、第 1 項が等しく、第 2 項が 0 となる。 $\theta' = \theta$ であれば、第 2 項が等しく、第 1 項が 0 となる。
15. $-\frac{1}{2} \log(2\pi\theta_2) - \sum_{i=1}^n \frac{(x_i - \theta_1)^2}{2\theta_2}$ となる。関数 `accept(x, y)` は $y \geq x$ のとき 1、 $y < x$ のとき $\exp(y-x)$ の確率で 0、 $1 - \exp(y-x)$ の確率で 1 を出力する。
16. MCMC (この場合は `metropolis`) の出力が `theta[1]` (平均) と `theta[2]` (標準偏差) の対になっている。ただし、この処理では前者を固定して乱数を発生している。`output` では両者の対の列になり、`output2` では後者のみの列になる。

17. (2.4) および (2.5) を適用して、エネルギー保存則が得られる。

$$\frac{dH(p, \theta)}{dt} = \sum_{j=1}^d \frac{\partial V(p)}{\partial p_j} \frac{dp_j(t)}{dt} + \sum_{j=1}^d \frac{\partial U(\theta)}{\partial \theta_j} \frac{d\theta_j(t)}{dt} = \sum_{j=1}^d \frac{\partial V(p)}{\partial p_j} \left(-\frac{\partial U(\theta)}{\partial \theta_j} \right) + \sum_{j=1}^d \frac{\partial U(\theta)}{\partial \theta_j} \frac{\partial V(p)}{\partial p_j} = 0$$

20. 例 19 の次式に $k = D$, $n = N$ を代入して得られる。

$$p(\theta|x_1, \dots, x_n) = \frac{\theta^{k+a-1}(1-\theta)^{n-k+b-1}}{B(k+a, n-k+b)}$$

21. $V[x_*\hat{\beta}] = x_*^\top V[\hat{\beta}]x_* = x_*^\top \{\sigma^2(X^\top X)^{-1}\}x_*$ となる。 $\hat{\beta}$ の変動と e の変動は独立であるので、分散は両者の和、すなわち $\sigma^2 x_*^\top (X^\top X)^{-1} x_* + \sigma^2$ となる。

第 3 章 数学的準備

27. まず、 $(a+bi)(c+di) = ac - bd + (bc+ad)i = (a-bi)(c-di) = \overline{a+bi} \cdot \overline{c+di}$ が成立する。次に、 $A = (a_{i,j})$, $u = (u_i)$ として、 $Au = \lambda u$ の両辺で共役をとって $\overline{Au} = \overline{\lambda u}$ が成立する。さらに、

$$\overline{\sum_{j=1}^n a_{i,j}u_j} = \sum_{j=1}^n \overline{a_{i,j}u_j} = \sum_{j=1}^n \overline{a_{i,j}} \cdot \overline{u_j} = \sum_{j=1}^n \overline{a_{i,j}} \cdot \overline{u_j}, \quad \overline{\lambda u_i} = \overline{\lambda} \cdot \overline{u_i}$$

より、

$$\overline{Au} = A\overline{u}, \quad \overline{\lambda u} = \overline{\lambda} \cdot \overline{u}$$

が成立する。

28. $U^\top U$ の (i, j) 成分は $u_i^\top u_j = \langle u_i, u_j \rangle = \delta_{i,j}$ なので、 $U^\top U = I_n$ である。したがって、 $\det(U^\top U) = \det(U)^2 = 1$ が成立し、 U は正則であり、逆行列 U^{-1} が U^\top と一致する。したがって、 $UU^\top = UU^{-1} = I_n$ が成立する。

29. (a) 各 $x \in (a, b)$ について、 $(x - \epsilon, x + \epsilon) \subset (a, b)$ なる $\epsilon > 0$ が存在する。各 $x \notin [a, b]$ について、 $(x - \epsilon, x + \epsilon) \cap [a, b]$ が空集合となる $\epsilon > 0$ が存在する。

(b) \mathbb{R} は数直線上のすべての点からなるので、触点をすべて含む。また、 \mathbb{Z} は孤立点からなり、触点はそれらしかなく、触点をすべて含んでいる。

(c) 各 $x \in \mathbb{R} \cap \mathbb{Z}^C$ について、 $(x - \epsilon, x + \epsilon) \subset \mathbb{R} \cap \mathbb{Z}^C$ なる $\epsilon > 0$ が存在する。

(d) 各 $x \in \mathbb{R}$, $\epsilon > 0$ について、 x が有理数ならば $(x - \epsilon, x + \epsilon)$ が無理数を、 x が無理数ならば $(x - \epsilon, x + \epsilon)$ が有理数を含む。前者より \mathbb{Q} は開集合ではないことがわかり、後者より \mathbb{Q}^C が開集合ではないことがわかる。したがって、 \mathbb{Q} は閉集合ではない。

(e) $(0, 0, 0)$ はその領域に含まれるが、半径 $\epsilon > 0$ をいくら小さくしても $(0, 0, 0)$ を中心とする球はその領域の中に含まれない。したがって、その領域は開集合ではない。 $\left(\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}, 0\right)$ はその領域に含まれないが、半径 $\epsilon > 0$ をいくら小さくしても $\left(\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}, 0\right)$ を中心とする球はその領域の補集合の中に含まれない。したがって、その領域は閉集合ではない。

30. $f^{(n)}(x) = \{e^x\}^{(n)} = e^x$, $f^{(n)}(x) = \{\log(1+x)\}^{(n)} = (-1)^{n+1} \frac{(n-1)!}{(1+x)^n}$ に $x=0$ を代入すると、それぞれ $f^{(n)}(0) = 1$, $f^{(n)}(0) = (-1)^{n+1}(n-1)!$ となる。それらを Maclaurin 展開の定義に代入して得られる。

31. $f(x) = e^{-1/x}$ について、 $f^{(n)}(x) = g_n(x)x^{-2n}e^{-1/x}$ なる次数が $n-1$ 以下の多項式 g_n が存在する。実際、 $n=1$ で成立し ($f'(x) = x^{-2}e^{-1/x}$)、 $n=k$ で成立すると、

$$\begin{aligned} f^{(k+1)}(x) &= \{f^{(k)}(x)\}' = g_k'(x)x^{-2k}e^{-1/x} - 2kg_k(x)x^{-(2k+1)}e^{-1/x} + g_k(x)x^{-2k} \cdot x^{-2}e^{-1/x} \\ &= \{x^2g_k'(x) - 2kxg_k(x) + g_k(x)\}x^{-2(k+1)}e^{-1/x} \end{aligned}$$

となり、 $x^2g_k'(x) - 2kxg_k(x) + g_k(x)$ は次数が k 以下の多項式となる。

32. $x, y, z \in V = \mathbb{R}$, $a \in \mathbb{R}$ について、

$$|x| \geq 0, \quad \|ax\| = |a| \|x\|, \quad |x-y| + |y-z| \geq |x-z|, \quad |x| = 0 \implies x = 0$$

が成立するのでノルムである。

$$\begin{aligned} \int_{\mathcal{X}} f(x)^2 q(x) dx, \int_{\mathcal{X}} g(x)^2 q(x) dx &< \infty \\ \implies \int_{\mathcal{X}} \{f(x) + g(x)\}^2 q(x) dx &\leq 2 \left\{ \int_{\mathcal{X}} f(x)^2 q(x) dx + \int_{\mathcal{X}} g(x)^2 q(x) dx \right\} < \infty \\ a \in \mathbb{R}, \int_{\mathcal{X}} f(x)^2 q(x) dx < \infty &\implies \int_{\mathcal{X}} \{af(x)\}^2 q(x) dx = a^2 \int_{\mathcal{X}} \{f(x)\}^2 q(x) dx < \infty \end{aligned}$$

より、 $L^2(q)$ は線形空間。次に、 $f \sim g \iff \int_{\mathcal{X}} \{f(x) - g(x)\}^2 q(x) dx = 0$ は同値関係。実際、

$$\begin{aligned} \int_{\mathcal{X}} \{f(x) - g(x)\}^2 q(x) dx &= \int_{\mathcal{X}} \{g(x) - h(x)\}^2 q(x) dx = 0 \\ \implies \int_{\mathcal{X}} \{f(x) - h(x)\}^2 q(x) dx &= \int_{\mathcal{X}} \{f(x) - g(x) + g(x) - h(x)\}^2 q(x) dx \\ &\leq 2 \int_{\mathcal{X}} \{f(x) - g(x)\}^2 q(x) dx + 2 \int_{\mathcal{X}} \{g(x) - h(x)\}^2 q(x) dx = 0 \end{aligned}$$

より $f \sim g, g \sim h \implies f \sim h$ が成立する。そして、 $L^2(q)$ を \sim で割った同値類 $L^2(q)/\sim$ についてのノルムとみなす。

$$\int_{\mathcal{X}} \{f(x)\}^2 q(x) dx \geq 0, \quad \int_{\mathcal{X}} \{af(x)\}^2 q(x) dx = a^2 \int_{\mathcal{X}} \{f(x)\}^2 q(x) dx$$

三角不等式は、以下のように確かめられる。

$$\begin{aligned} &\left\{ \left[\int_{\mathcal{X}} \{f(x) - g(x)\}^2 q(x) dx \right]^{1/2} + \left[\int_{\mathcal{X}} \{g(x) - h(x)\}^2 q(x) dx \right]^{1/2} \right\}^2 \\ &- \int_{\mathcal{X}} \{f(x) - g(x) + g(x) - h(x)\}^2 q(x) dx \\ &= \left[\int_{\mathcal{X}} \{f(x) - g(x)\}^2 q(x) dx \right]^{1/2} \cdot \left[\int_{\mathcal{X}} \{g(x) - h(x)\}^2 q(x) dx \right]^{1/2} \\ &- \int_{\mathcal{X}} \{f(x) - g(x)\} \{g(x) - h(x)\} q(x) dx \geq 0 \end{aligned}$$

そして、上記の3条件は、 f, g, h を $L^2(q)/\sim$ の要素として見たときも成立する。最後に、 $\int_{\mathcal{X}} \{f(x)\}^2 q(x) dx = 0 \implies f \sim 0$ が成立するので、ノルムになる。

33. 確率空間 $(\Omega, \mathcal{F}, \mu)$ で、 \mathcal{F} が $\{\omega \in \Omega \mid X(\omega) = 1\}$ または $\{\omega \in \Omega \mid X(\omega) = 0\}$ のいずれかの事象を要素として含む。

34. 最初の不等式は、 $(X - \mu)^2 - (X - \mu)^2 I(|X - \mu| \geq k) = (X - \mu)^2 I(|X - \mu| < k) \geq 0$ による。最後の不等式は、 $\{(X - \mu)^2 - k^2\} I(|X - \mu| \geq k) \geq 0$ による。
36. n を少なくすると、中心極限定理の極限から遠ざかるので、正規分布から乖離する。 m が少ないと、相対頻度の情報が不安定になり、結果的に正規分布から乖離する。
38. (a) $x \leq a_0$, $a_k < x$ では 0、それ以外では $h(x) = g(a_i)$ ($a_{i-1} < x \leq a_i$) となる $h: \mathbb{R} \rightarrow \mathbb{R}$ を、

$$|\mathbb{E}[g(X_n)] - \mathbb{E}[g(X)]| \leq |\mathbb{E}[g(X_n)] - \mathbb{E}[h(X_n)]| + |\mathbb{E}[h(X_n)] - \mathbb{E}[h(X)]| + |\mathbb{E}[h(X)] - \mathbb{E}[g(X)]|$$

に適用する。右辺第 1 項、第 3 項はそれぞれ ϵ 未満になる。次に $|g(x)| < M$ とすると、右辺第 2 項は高々

$$\begin{aligned} & \left| \sum_{i=1}^k P(X \in (a_{i-1}, a_i]) g(a_i) - \sum_{i=1}^k P(X_n \in (a_{i-1}, a_i]) g(a_i) \right| \\ & \leq \sum_{i=1}^k |g(a_i)| \cdot |P(X \in (a_{i-1}, a_i]) - P(X_n \in (a_{i-1}, a_i])| \end{aligned}$$

となる。 $X_n \xrightarrow{d} X$ より、 $|P(X \in (a_{i-1}, a_i]) - P(X_n \in (a_{i-1}, a_i])|$ の値は ϵ 以下に抑えることができる。したがって、右辺全体で $(Mk + 2)\epsilon$ となる。 $\epsilon > 0$ は任意であったので、 $|\mathbb{E}[g(X_n)] - \mathbb{E}[g(X)]| \rightarrow 0$ が示された。

- (b) $g: \mathbb{R} \rightarrow \mathbb{R}$ は有界で連続な関数であるので、たとえば、 $a \in \mathbb{R}$ を X の分布関数の連続点、 $m \geq 1$ として、

$$g_{a,m} = \begin{cases} 1, & x \leq a \\ -m(x - a) + 1, & a < x < a + \frac{1}{m} \\ 0, & x \geq a + \frac{1}{m} \end{cases}$$

なる $g_{a,m}: \mathbb{R} \rightarrow \mathbb{R}$ について、

$$F_X(a) \leq \mathbb{E}[g_{a,m}(X)] \leq F_X\left(a + \frac{1}{m}\right), \quad F_{X_n}(a) \leq \mathbb{E}[g_{a,m}(X_n)] \leq F_{X_n}\left(a + \frac{1}{m}\right)$$

が得られる。 $m \rightarrow \infty$ および $\mathbb{E}[g_{a,m}(X_n)] \rightarrow \mathbb{E}[g_{a,m}(X)]$ ($n \rightarrow \infty$) を用いると、 X, X_n の分布関数の $X = a$ における右極限が一致する。 $a \in \mathbb{R}$ は分布関数の任意の連続点であることから、 $X_n \rightarrow X$ が成立する。

39. (3.16) に $\theta = \theta_*$ を代入して、(3.18) を適用する。
40. (3.23) より、 $\mu = \mu_{**}$, $\sigma^2 = \sigma_{**}^2$ が成立する。さらに、 $\mu = \mu_*$, $\sigma^2 = \sigma_*^2$ であれば、それらを (3.25) に代入して (3.27) が得られる。

$$41. v(\mu) = \begin{bmatrix} \mu \\ -\mu^2/2 \end{bmatrix}, w(x) = \begin{bmatrix} x \\ 1 \end{bmatrix}, \nabla[v(\mu)^\top w(x)] = \nabla[\mu x - \mu^2/2] = x - \mu, \nabla^2[v(\mu)^\top w(x)] = 1$$

が成立するので、 $I(\mu) = \mathbb{V}[X - \mu] = 1$ および $J(\mu) = -\mathbb{E}[-1] = 1$ が得られる。

第 4 章 正則な統計モデル

42. 実現可能で正則であるので、命題 9 より $I(\theta_*) = J(\theta_*)$ となる。したがって、

$$J(\theta) = \theta_* \frac{\partial^2}{\partial \theta^2} \{-\log \theta\} + (1 - \theta_*) \frac{\partial^2}{\partial \theta^2} \{-\log(1 - \theta)\} = \theta_* \frac{1}{\theta^2} + (1 - \theta_*) \frac{1}{(1 - \theta)^2} = \frac{\theta_* - 2\theta\theta_* + \theta^2}{\theta^2(1 - \theta)^2}$$

$$I = J = J(\theta_*) = \frac{1}{\theta_*(1 - \theta_*)}$$

43. (a) 絶対値はノルムである (問題 32) ので、三角不等式が成立する。

$$|\phi_2(\theta) + \{\phi_1(\theta) - \phi_2(\theta)\}| \leq |\phi_2(\theta)| + |\phi_1(\theta) - \phi_2(\theta)|$$

(b) (a) の両辺で上限をとると、右辺はそれぞれの項で上限をとった方が大きくなる。

$$\sup_{\theta \in \Theta} |\phi_1(\theta)| \leq \sup_{\theta \in \Theta} \{|\phi_2(\theta)| + |\phi_1(\theta) - \phi_2(\theta)|\} \leq \sup_{\theta \in \Theta} |\phi_2(\theta)| + \sup_{\theta \in \Theta} |\phi_1(\theta) - \phi_2(\theta)|$$

(b) の不等式は ϕ_1, ϕ_2 を入れかえても成立する。したがって、

$$\max\{\sup_{\theta \in \Theta} |\phi_1(\theta)| - \sup_{\theta \in \Theta} |\phi_2(\theta)|, \sup_{\theta \in \Theta} |\phi_2(\theta)| - \sup_{\theta \in \Theta} |\phi_1(\theta)|\} \leq \sup_{\theta \in \Theta} |\phi_1(\theta) - \phi_2(\theta)|$$

となり、左辺が $|\sup_{\theta \in \Theta} |\phi_1(\theta)| - \sup_{\theta \in \Theta} |\phi_2(\theta)||$ となるので、(4.4) が成立する。

44. 最初の不等式は、

$$\frac{1}{n} \sum_{i=1}^n \log \frac{p(x_i | \theta_*)}{p(x_i | \theta)} \geq \epsilon_n - \frac{\eta_n(\theta)}{\sqrt{n}}$$

から成り立つ。2 番目の不等式は、 $\exp(-n\epsilon_n + \sqrt{n}\eta_n(\theta))$ を積分の外に出すために θ に関して上限をとった。最後の不等式は、相加相乗平均の大小関係

$$\sup_{\theta \in \Theta} \sqrt{n}\eta_n(\theta) \leq \frac{1}{2} \left(n\epsilon_n + \frac{\sup_{\theta \in \Theta} \eta_n(\theta)^2}{\epsilon_n} \right)$$

によった。

45. 最初の等式は、 $\theta = \theta_*$ の回りの Taylor 展開

$$-\log p(x | \theta) = -\log p(x | \theta_*) + (\theta - \theta_*)^\top \nabla \{-\log p(x | \theta)\} + \frac{1}{2}(\theta - \theta_*)^\top J(\theta^1)(\theta - \theta_*)$$

について X で平均をとり、正則ゆえ $\mathbb{E}_X[\nabla \log p(X | \theta)] = 0$ を適用した。2 番目の等式も $\theta = \theta_*$ の周りの Taylor 展開で、

$$\eta_n(\theta) = \eta_n(\theta_*) + (\theta - \theta_*)^\top \nabla \eta_n(\theta^2) = (\theta - \theta_*)^\top \nabla \eta_n(\theta^2)$$

となる。ただし、(4.1) の定義より、 $\eta_n(\theta_*) = 0$ となることを用いた。最後の等式は、左辺が

$$\left\{ \theta - \theta_* - \frac{J(\theta^1)^{-1} \nabla \eta_n(\theta^2)}{\sqrt{n}} \right\}^\top J(\theta^1) \left\{ \theta - \theta_* - \frac{J(\theta^1)^{-1} \nabla \eta_n(\theta^2)}{\sqrt{n}} \right\}$$

となり、展開すると

$$(\theta - \theta_*)^\top J(\theta^1)(\theta - \theta_*), \quad -\frac{1}{\sqrt{n}}(\theta - \theta_*)^\top J(\theta^1)J(\theta^1)^{-1} \nabla \eta_n(\theta^2), \quad \left\| J(\theta^1)^{-1/2} \frac{\nabla \eta_n(\theta^2)}{\sqrt{n}} \right\|^2$$

の 3 種類になる (上記の 2 番目の項は 2 回) ことから導かれる。

46. (a) Θ_* が単一の要素を含むとして、 Θ を $B(\epsilon_n, \theta_*)$ とその補集合 $B(\epsilon_n, \theta_*)^C$ に分割して議論を進めている。
- (b) 真の分布が統計モデルに対して正則であることを仮定しているので、行列 $J(\theta_*)$ が正則であるとしている。
- (c) θ_* が Θ の境界にはないので、 $\nabla \mathbb{E}_X[-\log p(X|\theta_*)] = 0$ が保証されている。
47. 事後分布の分散がサンプルサイズに反比例するから。正則ではないから。
48. 例 4 で得られた (1.12) から

$$-\log r(x|x_1, \dots, x_n) = \frac{1}{2} \log \left(2\pi \frac{n+2}{n+1} \right) + \frac{n+1}{2(n+2)} \left\{ x - \frac{1}{n+1} \sum_{i=1}^n x_i \right\}^2$$

$$\mathbb{E}_X \left[\left(X - \frac{1}{n+1} \sum_{i=1}^n x_i \right)^2 \right] = 1 + \left(\mu - \frac{1}{n+1} \sum_{i=1}^n x_i \right)^2$$

が得られる。これらを G_n, T_n の定義に代入すると、例 49 にある G_n, T_n の公式が得られる。図の生成は、プログラムを参照のこと。

49. 1 行目の `y_pred` には $r(x \cdot | x_1, \dots, x_n)$ に従って発生した $x = x^{(1)}, \dots, x^{(m)}$ が格納されている。2 行目は、`y_pred` から $r(\cdot | x_1, \dots, x_n)$ の分布を推定し、各 $j = 1, \dots, m$ について、 $x^{(j)}$ に対応する $r(x^{(j)} | x_1, \dots, x_n)$ を求め、それぞれを `dens$x`, `dens$y` に格納する。3 行目は、`dens` で表現されるプロットを関数 $r(\cdot | x_1, \dots, x_n)$ の形式で求めている。4 行目と 5 行目は汎化損失の平均を求める操作を行っていて、4 行目で $r(\cdot | x_1, \dots, x_n)$ に負の対数をとって真の分布 $q(\cdot)$ をかけて、5 行目で積分 $\int_{\mathcal{X}} -\log r(x | x_1, \dots, x_n) q(x) dx$ を計算している。
50. (4.19) に $\alpha = 0, 1$ を代入すると、

$$s(x, 0) = 0, \quad s(x, 1) = \log \int_{\Theta} p(x | \theta) p(\theta | x_1, \dots, x_n) - \log p(x | \theta_*)$$

となる。(4.22) と (4.23) に $\alpha = 0$ を代入すると、

$$s'(x, 0) = -\mathcal{E}(x) - \log p(x|\theta_*)$$

$$s''(x, 0) = \mathcal{V}(x)$$

となる。これらを (4.20) に $\alpha = 1$ を代入した式に代入すると、

$$\log \int_{\Theta} p(x | \theta) p(\theta | x_1, \dots, x_n) - \log p(x | \theta_*) = -\mathcal{E}(x) - \log p(x|\theta_*) + \frac{1}{2} \mathcal{V}(x) + \sum_{k=3}^{\infty} \frac{1}{k!} s^{(k)}(x, 0) \alpha^k$$

となり、(4.21) が得られる。

51. $s^{(k)}(x, \alpha)$ の各項が $s_h(x, \alpha)$ ($h \leq k$) の積で書けることを示す。 $k = 1$ のとき $s' = s_1(x, \alpha)$ より正しい。ある k で正しいとき、 $s^{(k+1)}(x, \alpha)$ を求めるときは、 $s_h(x, \alpha)$ ($h \leq k$) を α で微分したものが $s_h(x, \alpha)$ ($h \leq k + 1$) の積で書いていれば十分である。それは、以下のように示される。

$$s'_h(x, \alpha) = \frac{\int_{\Theta} \{\log p(x|\theta)\}^{h+1} p(x|\theta)^\alpha p(\theta|x_1, \dots, x_n) d\theta}{\int_{\Theta} p(x|\theta)^\alpha p(\theta|x_1, \dots, x_n) d\theta}$$

$$= \frac{\int_{\Theta} \{\log p(x|\theta)\}^h p(x|\theta)^\alpha p(\theta|x_1, \dots, x_n) d\theta \{ \int_{\Theta} \{\log p(x|\theta)\} p(x|\theta)^\alpha p(\theta|x_1, \dots, x_n) d\theta \}}{\{ \int_{\Theta} p(x|\theta)^\alpha p(\theta|x_1, \dots, x_n) d\theta \}^2}$$

$$= s_{h+1}(x, \alpha) - s_h(x, \alpha) s_1(x, \alpha)$$

52. 次式の左辺の第1項と第2項の $o_P\left(\frac{1}{n}\right)$ を加えたものを $o_P\left(\frac{1}{n}\right)$ としている。

$$\frac{1}{n} \sum_{i=1}^n \frac{1}{n} \text{tr}\{(I + o_P(1))J^{-1}\} + o_P\left(\frac{1}{n}\right) = \frac{1}{n} \text{tr}(IJ^{-1}) + o_P\left(\frac{1}{n}\right)$$

53. $A = (a_{i,j}), B = (b_{k,l})$ の積はそれぞれ、 $AB = (\sum_{j=1}^n a_{i,j} b_{k,l} \delta_{j,k})$, $BA = (\sum_{j=1}^m b_{k,l} a_{i,j} \delta_{l,i})$ となる。
それぞれのトレースは、 $\sum_{i=1}^m \sum_{j=1}^n a_{i,j} b_{j,i} = \sum_{k=1}^n \sum_{l=1}^m b_{k,l} a_{l,k}$ となり、一致する。

第5章 情報量規準

54. $I_n = J_n$ のとき、大きさが d の単位行列のトレースを n で割った値なので、 d/n となる。

55. 最尤推定量を求める際に、

$$\frac{\partial l}{\partial \sigma^2} = \frac{1}{2\sigma^2} - \frac{1}{n} \sum_{i=1}^n \frac{(y_i - \sum_{k=1}^p x_{i,k} \beta_k)^2}{2(\sigma^2)^2}$$

が必要となるが、この関係式を

$$l = \frac{1}{n} \sum_{i=1}^n -\log p(x_i, y_i | \beta, \sigma^2) = \frac{1}{2} \log 2\pi\sigma^2 + \frac{1}{n} \sum_{i=1}^n \frac{(y_i - \sum_{k=1}^p x_{i,k} \beta_k)^2}{2\sigma^2}$$

に代入すると、第2項が $1/2$ となる。

56. X, y は、それぞれ大きさ $n \times p, n$ で、説明変数、目的変数の値をもつ。 T の各列は、 X の列を同じ個数 (m 個) 選択する組み合わせからなる。たとえば、 X の列数が $p = 5$ であって $m = 2$ 変数を選択する場合、 T は 2×10 の行列になる。 m 個の説明変数を選択したもので、RSS の値を比較して、その最小値が $S.\min$ に格納される。その際に選択された m 個の列からなるベクトルが set.q に格納される。

57. 最尤推定 $\hat{\theta}$ は、 $\mathcal{L}(\cdot)$ を最小にする。しかし、

$$\mathbb{E}_X \left[\log \frac{p(X|\theta_*)}{p(X|\hat{\theta})} \right] \geq \epsilon_n$$

であれば、 $\mathcal{L}(\hat{\theta}) > \mathcal{L}(\theta_*)$ となる。したがって、少なくとも確率1で、

$$\mathbb{E}_X \left[\log \frac{p(X|\theta_*)}{p(X|\hat{\theta})} \right] < \epsilon_n$$

とならなければならない。また、 $\nabla \mathcal{L}(\hat{\theta}) = 0$ であるので、平均値の定理より

$$0 = \nabla \mathcal{L}(\theta_*) + \nabla^2 \mathcal{L}(\theta^1)(\hat{\theta} - \theta_*)$$

なる θ^1 が $\theta_*, \hat{\theta}$ の間に存在し、

$$\hat{\theta} = \theta_* - \nabla^2 \mathcal{L}(\theta^1)^{-1} \nabla \mathcal{L}(\theta_*)$$

とできる。次に、 \mathcal{L} を θ で2回微分して $\theta = \theta^1$ を代入すると、前半の主張 $\hat{\theta} \xrightarrow{P} \theta_*$ より、

$$\nabla^2 \mathcal{L}(\theta^1) = \nabla^2 \mathbb{E}_X [-\log p(X|\theta)] \Big|_{\theta=\theta^1} + o_P(1) = J + o_P(1)$$

が成立する。さらに、 $\mathcal{L}(\theta)$ の定義を θ で 1 回微分して $\theta = \theta_*$ を代入すると、(4.1) と (4.2) の定義より、

$$-\nabla \mathcal{L}(\theta_*) = \frac{1}{\sqrt{n}} \nabla \eta_n(\theta_*) + o_P \left(\frac{1}{\sqrt{n}} \right) = J \Delta_n + o_P \left(\frac{1}{\sqrt{n}} \right)$$

が成立する。

58. (a) $X_i \sim N(0, 1)$ なので、

$$\mathbb{E}[X_i^2] = \mathbb{V}[X_i] = 1, \mathbb{E}[(X_i^2)^2] = \mathbb{E}[X_i^4] = 3, \mathbb{V}[X^2] = \mathbb{E}[(X^2)^2] - \{\mathbb{E}[X^2]\}^2 = 3 - 1 = 2$$

が成立する。

(b) それぞれが独立なので、

$$\mathbb{E} \left[\sum_{i=1}^m X_i^2 \right] = \sum_{i=1}^m \mathbb{E}[X_i^2] = m \quad \mathbb{V} \left[\sum_{i=1}^m X_i^2 \right] = \sum_{i=1}^m \mathbb{V}[X_i^2] = 2m$$

59. (a) 命題 21 によって、「(5.8) \implies (5.10)」「(5.9) \implies (5.11)」が成立する。そして、命題 20 によって、

$$\begin{aligned} \mathbb{E}_X[-\log p(X|\hat{\theta}_n)] &= \mathbb{E}_X[-\log p(X|\theta_*)] + \frac{1}{2} \Delta_n^\top J \Delta_n + o_P \left(\frac{1}{n} \right) \\ \frac{1}{n} \sum_{i=1}^n \{-\log p(x_i|\theta_*)\} &= \frac{1}{n} \sum_{i=1}^n \{-\log p(x_i|\hat{\theta}_n)\} + \frac{1}{2} \Delta_n^\top J \Delta_n + o_P \left(\frac{1}{n} \right) \end{aligned}$$

が得られる。これらから、命題 22 の (5.8)(5.9) が得られる。

(b) 命題 22 の (5.10) および (5.11) が (5.14) を意味している。また、AIC に関しては、命題 9 (実現可能かつ正則であれば $I = J$) を用いている。

(c) 命題 19 と命題 21 より、ただちに命題 24 が得られる。

60. (5.20) の第 2 項について、 $\nu(x_i)$ は $-\log p(x|\theta)$ の事後分散であるから、 $-\log p(x_i|\theta)$ の 2 乗事後平均から $-\log p(x_i|\theta)$ の事後平均の 2 乗を引いた値になる。それらが **mean** の内側の処理になる。それらを各 $i = 1, \dots, n$ で計算して平均をとるのが **mean** の操作になる。その値に (5.20) の第 1 項である経験損失の値を加えるのが関数 **WAIC** の処理である。

63. (5.22) は (4.8) に対して負の対数をとって得られる。(5.23) の $\varphi(\theta_*)$ を代入して、もう一つの式が得られる。

65. 1 回のサンプル取得で $-\log p(x_1|\theta), \dots, -\log p(x_n|\theta)$ およびそれらの和 $\sum_{i=1}^n \{-\log p(x_i|\theta)\}$ が得られる。ただし、 x_1, \dots, x_n は独立で、事後分布に従う乱数系列である。このサンプルを何度も発生させて平均をとり、**WBIC** を得る。また、逆温度 $\beta > 0$ の生成は、事後分布に従う乱数を生成する段階で指定が必要となる。

第 6 章 代数幾何

67. $2\mathbb{Z} + 3\mathbb{Z} = \mathbb{Z}$ より、 \mathbb{Z} 全体になる。

68. (a) はプログラムを参照のこと。(b) は、 $(a, b) = (-3, 3), (1, 0), (-1, 0), (0, 0), (-3, 2), a^3 + 27b^2 = 135, 4, -4, 0, 0$ となるので、最初の 3 組は非特異。 $(a, b) = (0, 0)$ では $(x, y) = (0, 0)$ が特異点、 $(a, b) = (-3, 2)$ は $(x, y) = (-\frac{3b}{2a}, 0) = (0, 1)$ が特異点。

$$\begin{array}{ccccc}
y^2 - x^5 & \xrightarrow{(x, y) = (u_1 u_2, u_2)} & u_1^2(u_2^2 - u_1^3) & \xrightarrow{(u_1, u_2) = (v_1 v_2, v_2)} & v_1^2 v_2^4 (1 - v_1^3 v_2) \\
\downarrow & & \downarrow & & \downarrow \\
& & (x, y) = (u_1 u_2, u_2) & & (u_1, u_2) = (v_1, v_1 v_2) \\
u_2^2(1 - u_1^5 u_2^3) & & & & v_1^4(v_2^2 - v_1) \xrightarrow{(v_1, v_2) = (w_1, w_1 w_2)} -w_1^5(1 - w_1 w_2^2) \\
& & & & \downarrow \\
& & & & (v_1, v_2) = (w_1 w_2, w_2) \\
z_1^{10} z_2^5 (1 - z_2) & \xrightarrow{(u_1, u_2) = (z_1, z_1 z_2)} & w_1^4 w_2^5 (w_2 - w_1) & \xrightarrow{(w_1, w_2) = (z_1 z_2, z_2)} & z_1^4 z_2^{10} (1 - z_1) \\
& & & & \downarrow \\
& & & & (v_1 v_2, v_1 v_2^2) \\
(x, y) = & \left\{ \begin{array}{l} (u_1, u_1 u_2) \\ (u_1, u_1 u_2) \end{array} \right. = & \left\{ \begin{array}{l} (v_1 v_2, v_1 v_2^2) \\ (v_1, v_1^2 v_2) \end{array} \right. = & \left\{ \begin{array}{l} (w_1, w_1^3 w_2) \\ (w_1 w_2, w_1^2 w_2^3) \end{array} \right. = & \left\{ \begin{array}{l} (z_1 z_2^2, z_1^2 z_2^5) \\ (z_1^2 z_2, z_1^5 z_2^3) \end{array} \right.
\end{array}$$

69. (a) $\epsilon > 0$, $x \in M$ によって開集合 $B(\epsilon, x) := \{y \in M \mid \text{dist}(x, y) < \epsilon\}$ を定めると、 $x \neq y$ であれば、 $\epsilon > 0$ を十分小さくとることによって、 $B(\epsilon, x) \cap B(\epsilon, y)$ を空集合にすることができる。

(b) $2n, 2(n+1)$ を含む開集合はともに $2n+1$ を要素として含むから。

70. (a) $i < j$ を仮定すると、 $x_i, x_j \neq 0$ として、

$$\begin{aligned}
& \phi_i^{-1}(x_0, x_1, \dots, x_{i-1}, x_{i+1}, \dots, x_d) = [x_0 : x_1 : \dots : x_{i-1} : 1 : x_{i+1} : \dots : x_d] \\
& \phi_j \circ \phi_i^{-1}(x_0, x_1, \dots, x_{i-1}, x_{i+1}, \dots, x_d) = \phi_j([x_0 : x_1 : \dots : x_{i-1} : 1 : x_{i+1} : \dots : x_d]) \\
& = \left(\frac{x_0}{x_j}, \frac{x_1}{x_j}, \dots, \frac{x_{i-1}}{x_j}, \frac{1}{x_j}, \frac{x_{i+1}}{x_j}, \dots, \frac{x_{j-1}}{x_j}, \frac{x_{j+1}}{x_j}, \dots, \frac{x_d}{x_j} \right)
\end{aligned}$$

$j < i$ の場合に j の位置が i より前にくる以外には差異はない。

(b) $(u_x, v_x) \in U_x$ と $(u_y, v_y) \in U_y$ の間に

$$(u_x, u_x v_x, [1 : v_x]) = (u_y v_y, v_y, [u_y : 1]) \implies v_x u_y = 1, u_x = u_y v_y, v_y = u_x v_x$$

が成立するとき、 $(u_x, v_x) = (u_y v_y, \frac{1}{u_y})$, $(u_y, v_y) = (\frac{1}{v_x}, u_x v_x)$ が成立する。

71. 以下では、 $[x', y'] \in \mathbb{P}^1$ に関して、 $(x', y') \neq (0, 0)$ とする。まず、 $xy' = x'y$ を仮定する。 $x, x' \neq 0$ のとき、 $y' = x'y/x$, $[x' : y'] = [x' : x'y/x] = [x : y]$ が成立する。 $x = 0, x' \neq 0$ のとき、 $y = 0$ が成立する。 $x' = 0, x \neq 0$ のとき、 $y' = 0$ となり不適。 $x = x' = 0$ のとき、同様の理由で $y' \neq 0$ となる。このとき、 $y = 0$ であれば $(x, y) = (0, 0)$ 、 $y \neq 0$ であれば $[x : y] = [x' : y'] = [0 : 1]$ が成立する。逆に、 $(x, y) = (0, 0)$ を仮定すると、 $xy' = x'y$ が成立する。また、 $[x : y] = [x' : y']$, $(x, y) \neq (0, 0)$ を仮定すると、 $xy' = x'y$ が成立する。

72. 上記のようになる。

73. (u_i, v_i, w_i) を用いずに書くと以下のようなになる。

	$g(u)$	$f(g(u))$	$ g'(u) $	$(\kappa_1, \kappa_2, \kappa_3)$	(h_1, h_2, h_3)
(3)	(α_2, y, z)	$z^2\{(1 + \alpha_2 y)^2 + \alpha_2^2 y^4\}$	$ w_1 $	$(0, 0, 2)$	$(0, 0, 1)$
(5)	$(x, \beta_2, y + \alpha_1)$	$x^2(y + \alpha_1)^2\{1 + \beta_2^4(y + \alpha_1)^2\}$	$ u_2 w_2 $	$(2, 0, 2)$	$(1, 0, 1)$
(6)	(x, y, γ_1)	$x^2 y^4 (1 + \gamma_1^2)$	$ u_3 v_3^2 $	$(2, 4, 0)$	$(1, 2, 0)$
(7)	(x, γ_2, β_1)	$x^2 \gamma_2^2 \beta_1^4 (1 + \gamma_2^2)$	$ u_4 v_4 w_4^2 $	$(2, 2, 4)$	$(1, 1, 2)$

74. 以下のようになる。

	u	$g(u)$	$g'(u)$
(5)	$(x, \alpha_2, \xi_1, \xi_3)$	$(x, \xi_1 \alpha_2 \xi_3, \xi_3 - x \xi_1, \alpha_2 \xi_3)$	$-\alpha_2 \xi_3^2$
(7)	$(\alpha_1, \beta_2, w, \xi_1)$	$(\alpha_1 \beta_2, \xi_1 w, \alpha w - \alpha_1 \beta_2 \xi_1, w)$	$\alpha_1 \xi_1 w$
(9)	$(x, \beta_1, \gamma_2, w)$	$(x, \beta_1 \gamma_2 w, \beta_1 x w - x \gamma_2 \beta_1, w)$	$-x \beta_1 w^2$
(10)	$(x, \gamma_1, \delta_1, w)$	$(x, (1 + \delta_1 \gamma_1)w, x(1 + \delta_1 \gamma_1)(w \gamma_1 - 1), w)$	$x w^2 \gamma_1 (1 + \delta_1 \gamma_1)$
(11)	(x, δ_2, ξ_1, w)	$(x, \xi w, x w \delta_2 \xi_1 (\xi_1 - 1) - x \xi_1, w)$	$-x w^2 \xi_1 (1 - \xi_1)$

第7章 WAIC の本質

75. 例 72 より、各 $1 \leq m \leq d$ で

$$\int_0^1 t^z t^{\lambda-1} (-\log t)^{m-1} dt = \frac{(m-1)!}{(z+\lambda)^m}$$

が成立する。 $d = m$ とおき、 $\gamma_d = [(d-1)! \prod_{j=1}^d \kappa_j]^{-1}$ とおいたものをかけると、

$$\int_0^1 t^z \gamma_d t^{\lambda-1} (-\log t)^{d-1} dt = \frac{1}{(z+\lambda)^d} \cdot \frac{1}{\prod_{j=1}^d \kappa_j}$$

が成立する。また、この右辺は例 73 の $\zeta(z)$ に等しい。また、具体的には (7.4) によって与えられるので、問題の等号が得られる。

76. (a) 右辺は以下のように書ける。

$$\frac{(\beta + \gamma)z^2 + \{\beta(p+q) + \alpha + 2\gamma p\}z + \alpha q^2 + \beta p q + \gamma p^2}{(z+p)^2(z+q)}$$

したがって、 $\beta + \gamma = 0$, $\beta(p+q) + \alpha + 2\gamma p = 0$, $\beta p q + \gamma p^2 = 1$ の 3 式が成立し、これを解くと、次式が得られる。

$$\alpha = \frac{p-q}{p(p+a)}, \quad \beta = \frac{1}{p(p+a)}, \quad \gamma = -\frac{1}{p(p+a)}$$

(b) 式の個数、変数の個数ともに $\sum_{k=1}^s m_k$ となる。

77. $\int_{-a}^a \frac{1}{2a} dx = 2a \cdot \frac{1}{2a} = 1$ となる。グラフはプログラムを参照のこと。

78. (7.7) は、 $\sum_{k=1}^s m_k$ 個の項からなるが、 $t \rightarrow 0$ では t^{λ_k-1} の λ_k を最小にしている k がオーダーを支配する。同じ λ_k の値をもつときは、 $(-\log t)^{j-1}$ で $j = m_k$ における m_k の最大の項がオーダーを支配する。

79. (a) 命題 30 の

$$d\tau\delta(\tau - u^{2k})u^h b(u)du = \tau^{\lambda-1} (-\log \tau)^{m-1} du^* + o(-\tau^{\lambda-1} \log \tau)$$

に $\tau = t/n$ を代入した。

(b)

$$\begin{aligned} & \log \left(\sum_{\alpha} \frac{(\log n)^{m(\alpha)-1}}{n^{\lambda(\alpha)}} I_{\alpha} \right) \\ &= \log \left[\frac{(\log n)^{m-1}}{n^{\lambda}} \left(\sum_{\alpha \in A} I_{\alpha} \right) \left(1 + \sum_{\alpha \notin A} \frac{(\log n)^{m(\alpha)-m}}{n^{\lambda(\alpha)-\lambda}} \cdot \frac{I_{\alpha}}{\sum_{\alpha' \in A} I_{\alpha'}} \right) \right] \\ &= \log \left(\frac{(\log n)^{m-1}}{n^{\lambda}} \sum_{\alpha \in A} I_{\alpha} \right) + o_P(1) \\ &= -\lambda \log n + (m-1) \log \log n + \log \sum_{\alpha \in A} I_{\alpha} + o_P(1) \end{aligned}$$

80. (a) $S'_{\lambda}(a) = \int_0^{\infty} t^{\lambda-1} \sqrt{t} e^{-t+a\sqrt{t}} dt = S_{\lambda+\frac{1}{2}}(a)$

(b) $S''_{\lambda}(a) = \int_0^{\infty} t^{\lambda-1} t e^{-t+a\sqrt{t}} dt = S_{\lambda+1}(a)$

(c)

$$\begin{aligned} S_{\lambda+1}(a) &= \int_0^{\infty} (-e^{-t})' (t^{\lambda} e^{a\sqrt{t}}) dt = \int_0^{\infty} e^{-t} (t^{\lambda} e^{a\sqrt{t}})' dt \\ &= \int_0^{\infty} e^{-t} \frac{a}{2\sqrt{t}} t^{\lambda} e^{a\sqrt{t}} dt + \int_0^{\infty} e^{-t} \lambda t^{\lambda-1} e^{a\sqrt{t}} dt = \frac{a}{2} S_{\lambda+\frac{1}{2}}(a) + \lambda S_{\lambda}(a) \end{aligned}$$

$a = \xi_n(u)$ を代入して、両辺に $T[\cdot] = \sum_{\alpha \in A} \int_{[0,1]^d} du_{\alpha}^*[\cdot]$ を作用させ、 $T[S_{\lambda}(\xi_n(u))]$ で割ると、

$$\frac{T[S_{\lambda+1}(\xi_n(u))]}{T[S_{\lambda}(\xi_n(u))]} = \frac{T[\lambda S_{\lambda}(\xi_n(u))]}{T[S_{\lambda}(\xi_n(u))]} + \frac{T[\frac{\xi_n(u)}{2} S_{\lambda+\frac{1}{2}}(\xi_n(u))]}{T[S_{\lambda}(\xi_n(u))]}$$

となる。一方、分母は $T[S_{\lambda}(\xi_n(u))]$ と表せて、

$$\begin{aligned} T[S_{\lambda+1}(\xi_n(u))] &= \mathcal{E}[t|x_1, \dots, x_n] \cdot T[S_{\lambda}(\xi_n(u))] \\ T[\xi_n(u) S_{\lambda+1/2}(\xi_n(u))] &= \mathcal{E}[\sqrt{t} \xi_n(u) | x_1, \dots, x_n] \cdot T[S_{\lambda}(\xi_n(u))] \end{aligned}$$

が成立するので、(7.27) が得られる。

81. (a) (7.50) の両辺の対数を取り、

$$\log \left(\frac{\int_{\Theta} p(x_n|\theta)^{\beta} \prod_{i=1}^{n-1} p(x_i|\theta)^{\beta} \varphi(\theta) d\theta}{Z_{n-1}(\beta)} \right) = -\log \left(\frac{\int_{\Theta} p(x_n|\theta)^{-\beta} \prod_{i=1}^n p(x_i|\theta)^{\beta} \varphi(\theta) d\theta}{Z_n(\beta)} \right)$$

を得る。右辺に関しては、最初の x_n の箇所を x_1, \dots, x_n にそれぞれ置き換えて和をとって n で割る。

$$-\frac{1}{n} \sum_{i=1}^n \log \left(\frac{\int_{\Theta} p(x_i|\theta)^{-\beta} \prod_{i=1}^n p(x_i|\theta)^{\beta} \varphi(\theta) d\theta}{Z_n(\beta)} \right)$$

そして、

$$\log \left(\mathcal{E}_\beta^{-n} p(x_n | \theta)^\beta \right) = -\frac{1}{n} \sum_{i=1}^n \log \left(\mathcal{E} [p(x_i | \theta)^{-\beta}] \right)$$

について、平均 $\mathbb{E}_{X_1 \dots X_n} [\cdot]$ をとると、

$$\mathbb{E}_{X_1 \dots X_{n-1}} \mathbb{E}_X \left[\log \mathcal{E}_\beta^{-n} [p(X | \theta)^\beta] \right] = -\mathbb{E}_{X_1 \dots X_n} \left[\frac{1}{n} \sum_{i=1}^n \log \mathcal{E}_\beta [p(X | \theta)^{-\beta}] \right]$$

となる。最後に $\beta \mathbb{E}_X [\log p(X | \theta_*)]$ を両辺から引くと、(7.51) が得られる。

(b)

$$\mathcal{G}'_n(0) = -\frac{1}{n} \left(\lambda + \frac{1}{2} \mathcal{E} [\sqrt{t} \xi_n(u) | x_1, \dots, x_n] \right)$$

$$\mathcal{T}'_n(0) = -\frac{1}{n} \left(\lambda - \frac{1}{2} \mathcal{E} [\sqrt{t} \xi_n(u) | x_1, \dots, x_n] \right)$$

$$\mathcal{G}''_n(0) = \mathbb{E}_X [\mathcal{V}(X)]$$

$$\mathcal{T}''_n(0) = \frac{1}{n} \sum_{i=1}^n \mathcal{V}(x_i) = \mathbb{E}_X [\mathcal{V}(X)] + o_P(1)$$

82. x_i を固定して、 x_1, \dots, x_n の n を大きくすると、事後分散 $\mathcal{V} [\sqrt{t} a(x_i, u) | \xi_n]$ が確率収束する。 x_i をすべての x_1, \dots, x_n に置き換えて、それらを加えて n で割った値 V_n は、大数の法則より、 $\mathbb{E}_X \left[\mathcal{V} \left[\sqrt{\frac{t}{n}} a(X, u) \right] \right] = V(\xi)$ に確率収束する。次に、この確率収束を $\mathbb{E}_{X_1 \dots X_n}$ で平均をとった場合、有限の n でも収束後も確率変動はない。そのため極限が通常の収束になる。

83. 命題 37 より、

$$\mathbb{E}_{X_1 \dots X_n} [G_n] - \mathbb{E}_{X_1 \dots X_n} \left[T_n + \frac{V_n}{n} \right] = \frac{2\nu}{n} - \mathbb{E}_{X_1 \dots X_n} \left[\frac{V_n}{n} \right] + o \left(\frac{1}{n} \right)$$

が成立する。さらに、命題 38 の $\mathbb{E}_{X_1 \dots X_n} [V_n] = 2\nu + o_P(1)$ を適用すると、主張が得られる。

84. 事後分布を計算する際の $\prod_{i=1}^n p(x_i | \theta)$ が $\prod_{i=1}^n p(x_i | \theta)^\beta$ になる。したがって、(7.15) の $\exp\{-t + \sqrt{t} \xi_n(u)\}$ を $\exp\{\beta(-t + \sqrt{t} \xi_n(u))\}$ としてその先の議論を進めればよい。

$$S_\lambda(a) := \int_0^\infty t^{\lambda-1} e^{\beta(-t + a\sqrt{t})} dt, \quad \lambda > 0$$

$$\begin{aligned} S_{\lambda+1}(a) &= \int_0^\infty \left(-\frac{e^{-\beta t}}{\beta} \right)' (t^\lambda e^{a\beta\sqrt{t}}) dt = \frac{1}{\beta} \int_0^\infty e^{-\beta t} (t^\lambda e^{a\beta\sqrt{t}})' dt \\ &= \frac{1}{\beta} \int_0^\infty e^{-\beta t} \frac{a\beta}{2\sqrt{t}} t^\lambda e^{a\beta\sqrt{t}} dt + \frac{1}{\beta} \int_0^\infty e^{-\beta t} \lambda t^{\lambda-1} e^{a\beta\sqrt{t}} dt = \frac{a}{2} S_{\lambda+\frac{1}{2}}(a) + \frac{\lambda}{\beta} S_\lambda(a) \end{aligned}$$

が成立する。同様の議論によって、(7.27) が

$$\mathcal{E} [t | x_1, \dots, x_n] = \frac{\lambda}{\beta} + \frac{1}{2} \mathcal{E} \left[\sqrt{t} \xi_n(u) | x_1, \dots, x_n \right]$$

となり、事後平均と事後分散の定義が一般化されるので、 $\mathcal{E}(\cdot)$, $\mathcal{V}(\cdot)$ が、 $\mathcal{E}_\beta(\cdot)$, $\mathcal{V}_\beta(\cdot)$ となる。したがって、命題 35 は

$$\begin{aligned} G_n &= \mathbb{E}_X [-\log p(X | \theta_*)] + \frac{1}{n} \left(\frac{\lambda}{\beta} + \frac{1}{2} \mathcal{E}_\beta [\sqrt{t} \xi_n(u) | x_1, \dots, x_n] \right) - \frac{1}{2} \mathbb{E}_X [\mathcal{V}_\beta(X)] + o_P \left(\frac{1}{n} \right) \\ T_n &= \frac{1}{n} \sum_{i=1}^n \{-\log p(x_i | \theta_*)\} + \frac{1}{n} \left(\frac{\lambda}{\beta} - \frac{1}{2} \mathcal{E}_\beta [\sqrt{t} \xi_n(u) | x_1, \dots, x_n] \right) - \frac{1}{2} \mathbb{E}_X [\mathcal{V}_\beta(X)] + o_P \left(\frac{1}{n} \right) \end{aligned}$$

となる。(7.44)–(7.47) はそれらから求まる。

85. (7.41) に $w(\theta) = p(x_i | \theta)$ を代入して、(7.48) が得られる。CV の値は (7.48) の左辺で定義されるが、このプログラムはその右辺を計算している。
86. グラフはプログラムを参照されたい。 $\beta = 1$ のとき、(7.43) の第 1 項が消える。

第 8 章 WBIC と機械学習への応用

87. 逆温度が $\beta > 0$ のときの WBIC_n 、自由エネルギー F_n をそれぞれ (8.1) および (8.2) で定義する。

$$F'_n(\beta) = \mathcal{E}_\beta \left[-\sum_{i=1}^n \log p(x_i | \theta) \right] = \text{WBIC}_n$$

$$F''_n(\beta) = -\mathcal{E}_\beta \left[\left\{ \sum_{i=1}^n \log p(x_i | \theta) \right\}^2 \right] + \left(\mathcal{E}_\beta \left[\sum_{i=1}^n \log p(x_i | \theta) \right] \right)^2 < 0$$

であるので、 $\text{WBIC}_n = F'_n(\beta)$ は β の単調減少である。また、 $F_n(0) = 0$ であり、

$$F_n = F_n(1) = \int_0^1 F'_n(\beta) d\beta$$

に平均値の定理を適用すると、

$$F_n = F'_n(\beta^1) = \mathcal{E}_{\beta^1} \left[-\sum_{i=1}^n \log p(x_i | \theta) \right]$$

なる $\beta = \beta^1 (0 < \beta^1 < 1)$ が存在する。また、 $F'_n(\beta)$ は β の単調減少関数であるから、自由エネルギー F_n と WBIC_n の値が等しくなるような逆温度 $\beta > 0$ が存在する。

88. $p = 0$ のとき、

$$\int_0^\infty dt t^{\lambda-1} e^{-t} \{1 + \sqrt{\beta t} \xi_n(u)\} = \int_0^\infty t^{\lambda-1} e^{-t} dt + \xi_n(u) \int_0^\infty t^{\lambda-1/2} e^{-t} dt$$

となり、 $p = 1$ のとき

$$\int_0^\infty dt t^{\lambda-1} e^{-t} \left\{ \frac{t - \sqrt{\beta t} \xi_n(u)}{\beta} \right\} \{1 + \sqrt{\beta t} \xi_n(u)\}$$

$$= \frac{1}{\beta} \int_0^\infty t^\lambda e^{-t} dt + \xi_n(u) \frac{1}{\sqrt{\beta}} \int_0^\infty t^{\lambda+1/2} e^{-t} dt - \xi_n(u) \frac{1}{\sqrt{\beta}} \int_0^\infty t^{\lambda-1/2} e^{-t} dt + \xi_n(u)^2 \int_0^\infty t^\lambda e^{-t} dt$$

となる。ただし、最後の項は $\beta \rightarrow 0$ で相対的に無視できる。

89. $\lambda > d/2$ を仮定すると、 $Z(n) \frac{n^\lambda}{(\log n)^{m-1}} > Z(n) n^{d/2}$ の左辺が正の値をとり、右辺が 0 に収束し、(8.10) の右辺が 0 に収束する。これは矛盾であり、 $\lambda \leq d/2$ が必要となる。
90. ある 1 個の最適なパラメータ θ_* について $K(\theta_*) = 0$ であり、 $\epsilon > 0$ を十分に小さくとると、 $K(\theta) < \epsilon$ なる開集合が Θ に含まれていることを仮定している。そして、 $J(\theta^1)$ を正定値にすることができ、(8.11) が得られる。

91. (8.12) および (8.15) の導出はそれぞれ以下のようになる。

$$\begin{aligned}\mathbb{E}_Y \left[\log \frac{q(Y)}{p(Y|x)} \right] &= \mathbb{E} \left[Y \{a\sigma(bx) + c\sigma(dx)\} + \frac{1}{2} \{a\sigma(bx) + c\sigma(dx)\}^2 \right] = \frac{1}{2} \{a\sigma(bx) + c\sigma(dx)\}^2 \\ p_k &= x(\xi_1 w)^k + w^k \{xw(\xi_1 - 1)\xi_1 \delta_2 - x\xi_1\} = x\xi_1 \{\xi_1^{k-1} - 1 + \delta_2 w(\xi_1 - 1)\} w^k \\ &= x\xi_1(\xi_1 - 1) \{1 + \xi_1 + \dots + \xi_1^{k-2} + \delta_2 w\} w^k\end{aligned}$$

(8.14) より、

$$\begin{aligned}& \begin{vmatrix} \frac{\partial a}{\partial x} & \frac{\partial a}{\partial w} & \frac{\partial a}{\partial \delta_2} & \frac{\partial a}{\partial \delta_2} \\ \frac{\partial b}{\partial x} & \frac{\partial b}{\partial w} & \frac{\partial b}{\partial \delta_2} & \frac{\partial b}{\partial \delta_2} \\ \frac{\partial c}{\partial x} & \frac{\partial c}{\partial w} & \frac{\partial c}{\partial \delta_2} & \frac{\partial c}{\partial \delta_2} \\ \frac{\partial d}{\partial x} & \frac{\partial d}{\partial w} & \frac{\partial d}{\partial \delta_2} & \frac{\partial d}{\partial \delta_2} \end{vmatrix} \\ &= \begin{vmatrix} 1 & 0 & 0 & 0 \\ 0 & \xi_1 & 0 & w \\ w(\xi_1 - 1)\xi_1 \delta_2 - \xi_1 & x(\xi_1 - 1)\xi_1 \delta_2 & xw(\xi_1 - 1)\xi_1 & 2xw\xi_1 \delta_2 - xw\delta_2 - x \\ 0 & 1 & 0 & 0 \end{vmatrix} \\ &= -xw^2(\xi_1 - 1)\xi_1\end{aligned}$$

92. (5), (7), (9), (10) の $|g'(u)|$ は、それぞれ以下の行列式に絶対値をつけたものである。

$$\begin{aligned}& \begin{vmatrix} \frac{\partial x}{\partial x} & \frac{\partial x}{\partial \alpha_2} & \frac{\partial x}{\partial \xi_1} & \frac{\partial x}{\partial \xi_3} \\ \frac{\partial y}{\partial y} & \frac{\partial y}{\partial \alpha_2} & \frac{\partial y}{\partial \xi_1} & \frac{\partial y}{\partial \xi_3} \\ \frac{\partial z}{\partial z} & \frac{\partial z}{\partial \alpha_2} & \frac{\partial z}{\partial \xi_1} & \frac{\partial z}{\partial \xi_3} \\ \frac{\partial w}{\partial w} & \frac{\partial w}{\partial \alpha_2} & \frac{\partial w}{\partial \xi_1} & \frac{\partial w}{\partial \xi_3} \end{vmatrix} = \begin{vmatrix} 1 & 0 & 0 & 0 \\ 0 & \xi_1 \xi_3 & \alpha_2 \xi_3 & \alpha_2 \xi_1 \\ -\xi_1 & 0 & -x & 1 \\ 0 & \xi_3 & 0 & \alpha_2 \end{vmatrix} \\ & \begin{vmatrix} \frac{\partial x}{\partial w} & \frac{\partial x}{\partial \alpha_1} & \frac{\partial x}{\partial \beta_2} & \frac{\partial x}{\partial \xi_1} \\ \frac{\partial y}{\partial y} & \frac{\partial y}{\partial \alpha_1} & \frac{\partial y}{\partial \beta_2} & \frac{\partial y}{\partial \xi_1} \\ \frac{\partial z}{\partial z} & \frac{\partial z}{\partial \alpha_1} & \frac{\partial z}{\partial \beta_2} & \frac{\partial z}{\partial \xi_1} \\ \frac{\partial w}{\partial w} & \frac{\partial w}{\partial \alpha_1} & \frac{\partial w}{\partial \beta_2} & \frac{\partial w}{\partial \xi_1} \end{vmatrix} = \begin{vmatrix} 0 & \beta_1 & \alpha_1 & 0 \\ \xi_1 & 0 & 0 & w \\ \alpha_1(1 - \beta_2 \xi_1) & w - \beta_2 \xi_1 & -\alpha_1 \xi_1 & 1 \\ 1 & 0 & 0 & 0 \end{vmatrix} \\ & \begin{vmatrix} \frac{\partial x}{\partial x} & \frac{\partial x}{\partial w} & \frac{\partial x}{\partial \beta_1} & \frac{\partial x}{\partial \gamma_2} \\ \frac{\partial y}{\partial y} & \frac{\partial y}{\partial w} & \frac{\partial y}{\partial \beta_1} & \frac{\partial y}{\partial \gamma_2} \\ \frac{\partial z}{\partial z} & \frac{\partial z}{\partial w} & \frac{\partial z}{\partial \beta_1} & \frac{\partial z}{\partial \gamma_2} \\ \frac{\partial w}{\partial w} & \frac{\partial w}{\partial w} & \frac{\partial w}{\partial \beta_1} & \frac{\partial w}{\partial \gamma_2} \end{vmatrix} = \begin{vmatrix} 1 & 0 & 0 & 0 \\ 0 & \beta_1 \gamma_2 & w \gamma_2 & w \beta_1 \\ w \beta_1 - \beta_1 \gamma_2 & x \beta_1 & xw - x \gamma_2 & -x \beta_1 \\ 0 & 1 & 0 & 0 \end{vmatrix}\end{aligned}$$

$$= \begin{vmatrix} \frac{\partial x}{\partial x} & \frac{\partial x}{\partial w} & \frac{\partial x}{\partial \gamma_1} & \frac{\partial x}{\partial \delta_1} \\ \frac{\partial y}{\partial x} & \frac{\partial y}{\partial w} & \frac{\partial y}{\partial \gamma_1} & \frac{\partial y}{\partial \delta_1} \\ \frac{\partial z}{\partial x} & \frac{\partial z}{\partial w} & \frac{\partial z}{\partial \gamma_1} & \frac{\partial z}{\partial \delta_1} \\ \frac{\partial w}{\partial x} & \frac{\partial w}{\partial w} & \frac{\partial w}{\partial \gamma_1} & \frac{\partial w}{\partial \delta_1} \end{vmatrix} = \begin{vmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 + \delta_1 \gamma_1 & w \delta_1 & w \gamma_1 \\ (1 + \delta_1 \gamma_1)(w \gamma_1 - 1) & x \gamma_1 (1 + \delta_1 \gamma_1) & x w (1 + \delta_1 \gamma_1) + x \delta_1 (w \gamma_1 - 1) & x \gamma_1 (w \gamma_1 - 1) \\ 0 & 1 & 0 & 0 \end{vmatrix}$$

また、極は $-(h_i + 1)/\kappa_i$ ($i = 1, \dots, d$) が成分となる。

	κ	h	極 $-(h + 1)/\kappa$
(5)	(0, 2, 0, 4)	(0, 1, 0, 2)	$(-\infty, -1, -\infty, -3/4)$
(7)	(4, 2, 0, 0)	(2, 1, 0, 0)	$(-3/4, -1, -\infty, -\infty)$
(9)	(2, 4, 2, 0)	(1, 2, 1, 0)	$(-1, -3/4, -1, -\infty)$
(10)	(2, 4, 2, 0)	(1, 2, 1, 0)	$(-1, -3/4, -1, -\infty)$

93. 答え slope。 $O_P(\log n)$ の変動を無視すれば同じ値になる。

94. $H = 1$ のとき左辺は $1/2$ 、右辺は 1 となる。 $H \geq 2$ のとき、右辺は $(H + 4)/3$ である。また、問題の仮定より、 \sqrt{H} は整数であり、左辺は $\sqrt{H}/2$ となる。両辺を 2 乗して右辺から左辺を引くと、 $4H^2 + 7H + 16$ となり、つねに正の値をとる。

95. まず、導出は以下のようになる。

$$\begin{aligned} & (2\pi)^{-N/2} \frac{2}{\Gamma(\nu/2)} \left(\frac{\nu\sigma^2}{2}\right)^{\nu/2} \cdot \frac{1}{2} \Gamma\left(\frac{\nu+N}{2}\right) \left(\frac{\nu\sigma^2 + z^\top z}{2}\right)^{-(\nu+N)/2} \\ &= \frac{\Gamma(\frac{\nu+N}{2})}{\Gamma(\frac{\nu}{2})} (2\pi)^{-N/2} \left(\frac{\nu\sigma^2}{2}\right)^{\nu/2} 2^{\frac{\nu+N}{2}} \left(1 + \frac{z^\top z}{\nu\sigma^2}\right)^{-\frac{\nu+N}{2}} (\nu\sigma^2)^{-\frac{\nu+N}{2}} \\ &= \frac{\Gamma(\frac{\nu+N}{2})}{\Gamma(\frac{\nu}{2})} \frac{1}{(\nu\pi)^{N/2} \sigma^N} \left(1 + \frac{z^\top z}{\nu\sigma^2}\right)^{-(\nu+N)/2} \end{aligned}$$

40 行目は `diag-post_multiply(X, lambda)` で、 $X \begin{bmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_d \end{bmatrix}$ を表す。ただし X は行列、

$\lambda = \begin{bmatrix} \lambda_1 \\ \vdots \\ \lambda_d \end{bmatrix}$ である。ここまでの積が $\mathbb{R}^{n \times M}$ であるが、これに $L \in \mathbb{R}^{M \times H}$ および $\hat{B}^\top \in \mathbb{R}^{H \times N}$ を順

次かけている。(8.19) のパラメータ (ν, σ) の student t 分布の対数尤度を加えている。 $\mu \in \mathbb{R}^N$ は z 平均である。

96. `lq_decomp` 関数は、QR 分解を実行するための関数で、QR 分解は、行列 A を直交行列 Q と上三角行列 R の積 $A = QR$ として表す (`qr.Q`, `qr.R` がそれぞれ Q , R を抽出する)。 $A^\top = QR$ となる直交行列 Q および上三角行列 R が存在するので、 $L := R^\top$ (下三角行列), $S := Q^\top$ (直交行列) とおけば、 $A = R^\top Q^\top = LS$ とできる。

97. $l(s)$ は、 $0 \leq s \leq \min\{H-r, M-r\}$ の 2 次関数であるから、極小点が区間内に含まれる場合、 $M+H-N-r$ が偶数の場合 ($M+H+N+r$ が偶数であることと同値) は $s = (M+H-N-r)/2$ で最小、奇数の場合は $s = (M+H-N-r \pm 1)/2$ で最小となる。 $M+H < N-r$ のとき $s = 0$ 、 $(M+H-N-r)/2 > H-r$ かつ $H < M$ のとき $s = H-r$ 、 $(M+H-N-r)/2 > M-r$ かつ $M < H$ のとき $s = M-r$ で最小となる。証明中に 2 種類のブローアップがあり、前者だけでは $A^{(s)}$ が消滅するが、

$$u_1^2 \cdots u_s^2 \left(\|C^{(s)}\|^2 + \sum_{i=1}^s \|b_i^{(s)}\|^2 + \left\| \sum_{i=1}^s b_i^{(s)} D_i^{(s)} \right\|^2 \right)$$

となり、 $s = M-r$ なら $D_i^{(s)}$ も消滅する。したがって、 $C^{(s)}$ もしくは $b^{(s)}$ のどこかの成分に関してブローアップする必要がある。

99. $\theta = t^4$, $L := \log\{t^{4k}(1-t^4)^{n-k}\}$ とおくと、

$$L' = \frac{4k}{t} + \frac{4t^3}{1-t^4}(n-k) = \frac{4(k-nt^4)}{t(1-t^4)}, \quad \mathbb{E}[(L')^2] = \frac{16}{t^2(1-t^4)^2} nt^4(1-t^4) = \frac{16t^2n}{1-t^4}$$

より、Fisher 情報量が $\frac{16t^2}{1-t^4}$ となり、Jeffrey の事前分布は $\frac{t}{\sqrt{1-t^4}}$ の定数倍となる。他方、

$$\frac{d\theta}{\sqrt{\theta(1-\theta)}} = \frac{4t^3 dt}{\sqrt{t^4(1-t^4)}} = \frac{4t dt}{\sqrt{1-t^4}}$$

が成立し、題意が満たされた。

100. $k_i = 0$ なる i に関しては、 $\theta_i^{2k_i}$ および $\theta_i^{-I(k_i \neq 0)}$ がともに 1 となり、積から除外される。そして、 $I_{i,j}(\theta)$ の定義の中にあられる $\theta_i^{-I(k_i \neq 0)}$ がなければ、 θ_i の指数は $2dk_i$ であるが、非ゼロの各 h は、 $i_k = h$, $j_k = h$ の 2 回だけ指数から引かれることになる。