

鈴木讓「ベイジアンネットワーク入門」(培風館)

### 3. 統計的学習

#### 3.4 条件付確率の推定 (2) 情報量基準の適用

鈴木讓

大阪大学

2009年12月24日(1)

# あらまし

## 情報量基準の適用

定理 3.6

定理 3.7

## 情報量基準

$$\mathcal{Y} := Y(\Omega)$$

経験的エントロピー: 例の状態分割  $\mathcal{S}$  への適合度

$$\mathcal{H}^{(\mathcal{S})}(z^n) := \sum_{s \in \mathcal{S}} \mathcal{H}_s(z^n)$$

パラメータ数: 状態分割  $\mathcal{S}$  の簡潔性

$$K(\mathcal{S}) := |(\mathcal{Y} - 1)| |\mathcal{S}|$$

情報量基準

$$IC^{(\mathcal{S})}(z^n) := \mathcal{H}^{(\mathcal{S})}(z^n) + \frac{K(\mathcal{S})}{2} d_n \quad (30)$$

## 情報量基準 (続)

$\{d_n\}_{n=1}^{\infty}$ : 非負実数列

$$\limsup_{n \rightarrow \infty} \frac{d_n}{n} = 0 \quad (31)$$

(31) を満足する実数列  $\{d_n\} = \{d_n^{(1)}\}, \{d_n^{(2)}\}$

1.  $\liminf_{n \rightarrow \infty} (d_n^{(1)} / d_n^{(2)}) > 1 \implies \{d_n^{(1)}\} \succ \{d_n^{(2)}\}$
2.  $\limsup_{n \rightarrow \infty} (d_n^{(1)} / d_n^{(2)}) < 1 \implies \{d_n^{(1)}\} \prec \{d_n^{(2)}\}$
3.  $\lim_{n \rightarrow \infty} (d_n^{(1)} / d_n^{(2)}) = 1 \implies \{d_n^{(1)}\} = \{d_n^{(2)}\}$

- ▶  $d_n = \log n \implies L^{(S)}(z^n)$  (MDL 基準)
- ▶  $d_n = 2 \implies \text{AIC}$  (Akaike's information criterion)

## 定理 3.6

$$A_n \sim B_n \stackrel{\text{def}}{\iff} \lim_{n \rightarrow \infty} A_n/B_n = 1$$

$$\Gamma_x(\alpha) := \int_0^x t^{\alpha-1} e^{-t} dt$$

定理 (Suzuki, 2006)

各  $S \in \text{Over}(S^*)$  で

$$\mu\{z^n \mid IC^{(S)}(z^n) < IC^{(S^*)}(z^n)\} \sim 1 - \frac{\Gamma_{\frac{K(S)-K(S^*)}{2}d_n} \left( \frac{K(S) - K(S^*)}{2} \right)}{\Gamma \left( \frac{K(S) - K(S^*)}{2} \right)} \quad (32)$$

## 証明の概略

$$\begin{aligned} & \mu\{z^n \mid IC^{(S)}(z^n) < IC^{(S^*)}(z^n)\} \\ & = \mu\{z^n \mid 2\{\mathcal{H}^{(S^*)}(z^n) - \mathcal{H}^{(S)}(z^n)\} > [K(S) - K(S^*)]d_n\}, \end{aligned}$$

$f_l: \chi_l^2$  の確率密度関数

$$\int_x^\infty f_l(t) dt = 1 - \frac{\Gamma_{x/2}(l/2)}{\Gamma(l/2)}$$

各  $S \in \text{Over}(S^*)$  で以下を示せば十分:

$$2\{\mathcal{H}^{(S^*)}(z^n) - \mathcal{H}^{(S)}(z^n)\} \sim \chi_{K(S)-K(S^*)}^2$$

## 証明の概略 (続)

$$\mathcal{S} \in \text{Over}(\mathcal{S}^*)$$

$$t \in \mathcal{S}^*$$

$$\mathcal{S}(t) := \{t \cap s \mid s \in \mathcal{S}\}$$

$$\alpha := |\mathcal{S}(t)|$$

$$\beta := |\mathcal{Y}|$$

$$V = (v_{y,s})$$

$$v_{y,s} := \frac{c_{y,s} - n_s \theta_{y,s}}{\sqrt{n_s \theta_{y,s}}}, \quad y \in \mathcal{Y}, s \in \mathcal{S}(t)$$

$$\theta_y := \theta_{y,s} \quad (s \in \mathcal{S}(t) \text{ によらず一定})$$

## 証明の概略 (続)

$$R := \left( \frac{\sqrt{n_s n_{s'}}}{n_t} \right)_{s, s' \in \mathcal{S}(t)}$$

固有値: 1 個は 1, 他  $\alpha - 1$  個は 0

$R$  のトレースは 1, 大きさ 2 以上の小行列はすべて 0

$$|\lambda I_\alpha - R| = \lambda^\alpha - \text{trace}(R)\lambda^{\alpha-1} + \cdots + (-1)^\alpha |R| = \lambda^{\alpha-1}(\lambda - 1)$$

固有ベクトル

$U_0$  固有値 1 に対応

$U_1 := (u_{1,s})_{s \in \mathcal{S}(t)}, \cdots, U_{\alpha-1} := (u_{\alpha-1,s})_{s \in \mathcal{S}(t)}$  固有値 0 に対応

$\tilde{U} := (U_0, U_1, \cdots, U_{\alpha-1})$  正規直交 ( $\alpha \times \alpha$ )

$U := (U_1, \cdots, U_{\alpha-1})$ : ( $\alpha \times (\alpha - 1)$ )



## 証明の概略 (続)

$$T := (\sqrt{\theta_y \theta_{y'}})_{y, y' \in \mathcal{Y}}$$

固有値: 1 個は 1, 他  $\beta - 1$  個は 0

固有ベクトル

$W_0 := (\sqrt{\theta_y})_{y \in \mathcal{Y}}$  固有値 1 に対応

$W_1 := (w_{y,1})_{y \in \mathcal{Y}}, \dots, W_{\beta-1} := (w_{y,\beta-1})_{y \in \mathcal{Y}}$  固有値 0 に対応

$\widetilde{W} := (W_0, W_1, \dots, W_{\alpha-1})$  正規直交 ( $\alpha \times \alpha$ )

$W := (W_1, \dots, W_{\beta-1})$ : ( $\beta \times (\beta - 1)$ )

$Q := {}^t UVW$ ,  $(\alpha - 1) \times (\beta - 1)$  行列

$$q_{i,j} := \sum_{s \in \mathcal{S}(t)} \sum_{y \in \mathcal{Y}} u_{i,s} v_{y,s} w_{y,j}$$

## 証明の概略 (続)

$$\begin{aligned}K(\mathcal{S}) - K(\mathcal{S}^*) &= \sum_{t \in \mathcal{S}^*} |\mathcal{S}(t)| (|\mathcal{Y}| - 1) - \sum_{t \in \mathcal{S}^*} (|\mathcal{Y}| - 1) \\ &= \sum_{t \in \mathcal{S}^*} (|\mathcal{S}(t)| - 1) (|\mathcal{Y}| - 1)\end{aligned}$$

$$\mathcal{A}_t := \sum_{s \in \mathcal{S}(t)} \sum_{y \in \mathcal{Y}} c_{y,s} \log \frac{c_{y,s}}{n_s} - \sum_{y \in \mathcal{Y}} c_{y,t} \log \frac{c_{y,t}}{n_t}$$

$$\mathcal{H}^{(\mathcal{S}^*)}(z^n) - \mathcal{H}^{(\mathcal{S})}(z^n) = \sum_{t \in \mathcal{S}^*} \mathcal{A}_t$$

$$\begin{aligned}\mathcal{B}_t &:= \sum_{s \in \mathcal{S}(t)} \sum_{y \in \mathcal{Y}} \frac{(c_{y,s} - n_s \theta_y)^2}{n_s \theta_y} - \sum_{y \in \mathcal{Y}} \frac{(c_{y,t} - n_t \theta_y)^2}{n_t \theta_y} \\ &= \sum_{s \in \mathcal{S}(t)} \sum_{y \in \mathcal{Y}} v_{y,s}^2 - \sum_{y \in \mathcal{Y}} v_{y,t}^2 = \sum_{i=1}^{\alpha-1} \sum_{j=1}^{\beta-1} q_{i,j}^2.\end{aligned}$$

## 証明の概略 (続)

▶  $2\mathcal{A}_t - \mathcal{B}_t \rightarrow 0 \ (n \rightarrow \infty)$

▶  $\mathcal{B}_t \sim \chi_{(\alpha-1)(\beta-1)}$

を証明すればよい (テキスト 119-121 ページ)

$$q_{i,j} \sim \mathcal{N}(0, 1)$$

を示せば、自由度  $(\alpha - 1)(\beta - 1)$  の  $\chi^2$  分布になる (命題 1.5)

## 定理 3.7

### 定理 3.7

各  $S \in \text{Under}(S^*)$  で、 $n \rightarrow \infty$  で、確率 1 で

$$\frac{IC^{(S)}(z^n) - IC^{(S^*)}(z^n)}{n} \rightarrow D(S^* || S)$$

証明: 確率 1 で、 $s \in \text{Over}(S^*) \cup \{S^*\}$  で

$$\frac{c_{y,s}}{n_s} \rightarrow \theta_{y,s}.$$

各  $S \in \text{Under}(S^*)$  で

$$\begin{aligned} \frac{c_{y,s}}{n_s} &= \sum_{t \in S^*} \frac{c_{y,snt}}{n_s} = \sum_{t \in S^*} \frac{n_{snt}/n}{n_s/n} \frac{c_{y,snt}}{n_{snt}} \\ &\rightarrow \sum_{t \in S^*} \frac{p(s \cap t)}{p(s)} \theta_{y,t}^* = \theta_{y,s} \end{aligned}$$

### 定理 3.7 (続)

$$\begin{aligned}
 & \frac{\mathcal{H}(\mathcal{S})(z^n) - \mathcal{H}(\mathcal{S}^*)(z^n)}{n} \\
 = & \sum_{t \in \mathcal{S}^*} \sum_{y \in \mathcal{Y}} \frac{c_{y,t}}{n_t} \log \frac{c_{y,t}}{n_t} - \sum_{s \in \mathcal{S}} \sum_{y \in \mathcal{Y}} \frac{c_{y,s}}{n_s} \log \frac{c_{y,s}}{n_s} \\
 = & \sum_{t \in \mathcal{S}^*} \sum_{s \in \mathcal{S}} \sum_{y \in \mathcal{Y}} \frac{c_{y,t} n_s}{n} \log \frac{c_{y,t}}{n_t \theta_{y,t}} - \sum_{t \in \mathcal{S}^*} \sum_{s \in \mathcal{S}} \sum_{y \in \mathcal{Y}} \frac{c_{y,t} n_s}{n} \log \frac{c_{y,s}}{n_s \theta_{y,s}} \\
 & + \sum_{t \in \mathcal{S}^*} \sum_{s \in \mathcal{S}} \sum_{y \in \mathcal{Y}} \frac{n_t n_s}{n} \cdot \frac{c_{y,t} n_s}{n_t n_s} \log \frac{\theta_{y,t}}{\theta_{y,s}} \\
 \rightarrow & \sum_{t \in \mathcal{S}^*} \sum_{s \in \mathcal{S}} \sum_{y \in \mathcal{Y}} p[s \cap t] \theta_{y,t} \log \frac{\theta_{y,t}}{\theta_{y,s}} \\
 = & D(\mathcal{S}^* \parallel \mathcal{S})
 \end{aligned}$$