Kryterium minimalnej p-wartości ilorazu wiarygodności

Porównanie p-wartości ilorazu wiarygodności oraz obiektywnych czynników Bayesa w modelach liniowych

Piotr Pokarowski i Jan Mielniczuk

Inst. Inf. UW IPI PAN, MiNI PW

<u>Przykład</u> Zbuduj model lin dla danych "samochody"

$$\text{ZużyciePaliwa} = \beta_0 + \beta_1 \text{Masa} + \beta_2 \text{PojemnSiln.} + \beta_3 \text{Moc} + \ldots$$

- Policzmy p-wartość testu F ($\Leftrightarrow$ iloraz wiarygodności) dla każdego modelu (podzbioru predyktorów)
- Sprawdźmy jaki jest związek $\mathbb{P}(>F)$ i BIC, AIC
- Okazało się, że $cor(\mathbb{P}(>F), BIC) \simeq 0.99$
- Dlaczego tak jest?

F, G dystrybuanty, $p(t|F) := 1 - F(t)$, $T \sim G$

$p(T|F)$ – p-wartość T przy hipotezie F

$X_{n,full}$ – plan modelu pełnego $n \times p_{full}$

0 – model minimalny (np. złożony tylko ze stałej)

$0 \subseteq j \subseteq full$

j: $y_n = X_{nj} \beta_j + \varepsilon_n$, $\varepsilon_n \sim N(0, \sigma^2 I_n)$

X_{nj} – plan o kolumnach ze zbioru j

$$S_{nj} := y_n^T (I - H_{nj}) y_n = \| y_n - X_{nj} \hat{\beta}_j \|^2$$

$$\text{Dewiacja} := 2 \log \frac{f_{\hat{\theta}_j}}{f_{\hat{\theta}_0}}$$

σ^2 znana $\quad D_{nj} := 2\log\frac{f_{\hat\beta_j}(y_n)}{f_{\hat\beta_0}(y_n)} = \frac{S_{n0}-S_{nj}}{\sigma^2}$

σ^2 nieznana $\quad D_{nj} := 2\log\frac{f_{\hat\beta_j\hat\sigma^2}(y_n)}{f_{\hat\beta_0\hat\sigma^2}(y_n)} = -n\log\frac{S_{nj}}{S_{n0}} \sim \mathcal{D}_{p_0,p_j,n}$

$\iff R^2_{nj} := 1-\frac{S_{nj}}{S_{n0}} \sim \mathcal{B}_{\frac{p_j-p_0}{2},\frac{n-p_j}{2}}$

$\iff F_{nj} = \frac{(S_{n0}-S_{nj})/(p_j-p_0)}{S_{nj}/(n-p_j)} \sim \mathcal{F}_{p_j-p_0,n-p_j}$

MPVC (minimal p-value criterion)

$$M_n = \operatorname*{argmin}_j p(D_{nj}\,|\,\mathcal{D}_{p_0,p_j,n}), \quad p(D_{n0}\,|\,\mathcal{D}_{p_0,p_0,n}) := \frac{1}{n}$$

$$= \operatorname*{argmin}_j p(D_{nj}\,|\,\chi^2_{p_j-p_0}), \quad p(D_{n0}\,|\,\chi^2_0) := \frac{1}{n}$$

Aproksymacja ogonów rozkładów χ^2 i $\mathcal{B}$

(1) $\forall p \geq 1 \qquad p(x \mid \chi^2_p) = e^{-x/2} \left(\frac{x}{2}\right)^{\frac{p}{2}-1} \Gamma^{-1}\left(\frac{p}{2}\right) \left[1 + O\left(\frac{1}{x}\right)\right], \; x \to \infty$

(2) $x > \frac{a-1}{a+b} \qquad p(x \mid \mathcal{B}_{a,b}) = (1-x)^b (bx)^{a-1} \Gamma^{-1}(a) \left[1 + O\left(\frac{1}{b}\right)\right], \; b \to \infty$

Jeśli $D_{nj} = d_j n + o_p(n)$, to

σ^2 znana $\qquad n\, p(D_{nj} \mid \chi^2_{p_j - p_0}) = e^{-D_{nj}/2}\, n^{\frac{p_j - p_0}{2}}\, c_1(p_j - p_0, d_j) \left[1 + O_p\left(\frac{1}{n}\right)\right]$

σ^2 nieznana $\qquad n\, p(D_{nj} \mid \mathcal{D}_{p_0, p_j, n}) = e^{-D_{nj}/2}\, n^{\frac{p_j - p_0}{2}}\, c_2(p_0, p_j, d_j) \left[1 + O_p\left(\frac{1}{n}\right)\right]$

$$\left\{ \begin{array}{l} d = -n \log(1-x) \\ e^{-d/n} = 1-x \end{array} \right\} \qquad S = D_{nj} - (p_j - p_0) \log n$$

$$-2 \log n\, p(D_{nj} \mid \chi^2_{p_j - p_0}) = S + \text{const} + O_p\left(\frac{1}{n}\right)$$

Regularny plan $\Rightarrow$ regularne dewiacje $\Rightarrow$ zgodność MPVC

(1) $\exists_{\, 0 \subseteq t \subseteq full} \quad y_n = X_{nt}\beta_t + \varepsilon_n \ , \ \varepsilon_n \sim N(0, \sigma^2 I_n)$

(2) t jest minimalnym w sensie "$\subseteq$" modelem spełniającym (1)

(3) $\forall_{j \notin t} \quad \frac{1}{n} X_{n,t\cup j}^T (I - H_{n_j}) X_{n,t\cup j} \longrightarrow \Sigma_j > 0$

(Casella et. al. Ann. Stat. 2009)

<u>Uwaga</u> (3) słabszy od "klasycznego"

(4) $\frac{1}{n} X_{n,full}^T X_{n,full} \longrightarrow \Sigma_{full} > 0$ (Nishii, Ann. Stat. 1984)

Wnioski

(1) $\frac{1}{n} S_{nj} \xrightarrow[a.s.]{} b_j^2 + \sigma^2$, $b_j^2 := \beta_{t,j}^T \Sigma_j \beta_{t,j}$

(2) σ^2 znana $\frac{1}{n} D_{nj} \xrightarrow[a.s.]{} \frac{b_0^2 - b_j^2}{\sigma^2} =: d_j$

(3) σ^2 nieznana $\frac{1}{n} D_{nj} \xrightarrow[a.s]{} \log \frac{b_0^2 + \sigma^2}{b_j^2 + \sigma^2} =: d_j$

Dewiacje (D_{nj}) są regularne

(1) $\forall_j \quad D_{nj} = d_j n + o_{\mathbb{P}}(n)$

(2) $\forall_{j \supseteq t} \quad D_{nj} - D_{nt} \xrightarrow{d} \chi^2_{p_j - p_t}$

Tw Jeśli dewiacje są regularne, to $\mathbb{P}(M_n = t) \to 1$.

Obiektywne czynniki Bayesa (BF - Bayes Factor)

- wykorzystują obiektywne (objective, noninformative, reference) rozkłady a priori

$$B_{1/0} = \frac{\int f(y|\theta_1)\,\pi_1(\theta_1)\,d\theta_1}{\int f(y|\theta_0)\,\pi_0(\theta_0)\,d\theta_0}$$

jeśli $j \supset 0$, to często wymaga się "zgodności" π i π_1:

$$B_{j/0} = \frac{\int f(y|\theta_1,\theta_2)\,\pi(\theta_1,\theta_2)\,d\theta_1\,d\theta_2}{\int f(y|\theta_1,\theta_2^0)\,\pi_1(\theta_1)\,d\theta_1}$$

$$= \frac{\int f(y|\theta_1,\theta_2)\,\pi_1(\theta_1)\,\pi_{2/1}(\theta_2|\theta_1)\,d\theta_1\,d\theta_2}{\int f(y|\theta_1,\theta_2^0)\,\pi_1(\theta_1)\,d\theta_1}$$

Obiektywne czynniki Bayesa dla modeli lin – σ^2 nieznana

$$BF = e^S c_k \left[1 + o_P(1)\right]$$

$$2S = D_{nj} - (p_j - p_0)\log n = -n \log \frac{RSS}{TSS} - (p_j - p_0)\log n$$

$S \iff$ kryterium Schwarza

c_k zależy od rodzaju BF

Fisher Factor $\quad FF := \dfrac{1}{n\, p(D_{nj} \mid \mathcal{D}_{p_0, p_j, n})} = e^S c_2 \left[1 + O_p\left(\frac{1}{n}\right)\right]$

Porovnanie $c_k \equiv c_k\left(p, \frac{RSS}{TSS}\right)$ $\quad r = \frac{RSS}{TSS}$

(1) Kass & Wasserman, JASA 1995
Pauler, Biometrika 1998

Unit information prior $\quad c_1 = 1$

Jeffreys prior $\quad c_2 = 2^{p/2}\,\Gamma\left(\frac{p+1}{2}\right) / \Gamma\left(\frac{1}{2}\right)$

(2) Zellner & Siow, Bayesian Statistics 1980
= Fractional BF O'Hagan, JRSS B, 1995

$$c_3 = 2^{p/2}\,\Gamma\left(\frac{p+1}{2}\right) r^{\frac{p+3}{2}} / \Gamma\left(\frac{1}{2}\right)$$

(3) Berger & Pericchi, JASA 1996
Intrinsic BF $\quad c_4 = \pi^{-1} r^{\frac{1}{2}} (p+2)^{\frac{p}{2}} B\left(\frac{1}{2}, \frac{p+1}{2}\right) \mathrm{hyp1f1}\left(\frac{p+1}{2}; \frac{p+2}{2}; \frac{p+2}{2}\left(1-\frac{1}{r}\right)\right)$
Casella et al, Ann. Stat. 2009

(4.) Statistic based BF
V. Johnson, JRSS B 2005

$$c_5 = r^{\frac{p+1}{2}} \left[\frac{p}{e(1-r)}\right]^{\frac{p}{2}}$$

(5) FF

$$c_6 = r^{\frac{p+1}{2}} \left[\frac{2}{1-r}\right]^{\frac{p-2}{2}} \Gamma\left(\frac{p}{2}\right)$$

Numeryczne porównanie funkcji $c_1 - c_6$:

- Zamieniamy c_k na wektor v_k
- Porównujemy wektory v_i, v_j odległością Canberra

$$CBR(v_i, v_j) = \sum_{l=1}^{m} \frac{|v_{li} - v_{lj}|}{|v_{li} + v_{lj}|} \overset{!}{=} CBR(v_i^{-1}, v_j^{-1})$$

lub $mCBR(v_i, v_j) = \underset{l}{\text{med}}$ .....

FF jest medoidem dla (CBR_{ij}) i $(mCBR_{ij})$