統計数理研究所 75 周年記念事業（2019）

変化点モデルのための AIC

AIC for change-point models

数理・推論研究系 二宮 嘉行 （Yoshiyuki Ninomiya）

要　旨

変化点問題は長い間議論されているが，その理由の一つは，変化点モデルに通常の統計的漸近理論を満たさせない非正則性が存在することにある．本研究の目的は，そのような変化点モデルの AIC を導出することである．AIC の罰則項は最大対数尤度の漸近バイアスの二倍であり，正則条件を満たすモデルであればそれはパラメータの数の二倍となる．一方変化点モデルでは，その非正則性がゆえにそれは $2m+2p_m$ とならない．ここで m と p_m はそれぞれ変化点の数と変化点以外のパラメータの数である．本稿では，その漸近バイアスが $6m+2p_m$ で評価できることを示した Ninomiya (2015) の結果を紹介する．

キーワード：構造変化，情報量規準，非正則性，Brown 運動，ランダムウォーク

1. モデル

独立観測系列 $\{\boldsymbol{x}_i,\ 1\le i\le n\}$ に対し，m 個の変化点 $k^{(1)},\dots,k^{(m)}$ をもつモデルを考える．簡単のため，$\boldsymbol{x}_i$ の分布は指数型分布族に属する，つまり確率関数は

$$k^{(j-1)}+1\le i\le k^{(j)} \quad \text{のとき} \quad \exp\{\boldsymbol{\theta}^{(j)\mathrm{T}}\boldsymbol{T}(\cdot)+S(\cdot)-A(\boldsymbol{\theta}^{(j)})\} \tag{1.1}$$

と書けるものとする（$1\le j\le m+1$）．ここで $k^{(0)}=0$ かつ $k^{(m+1)}=n$ としている．また，$\boldsymbol{\theta}^*=(\boldsymbol{\theta}^{*(1)\mathrm{T}},\dots,\boldsymbol{\theta}^{*(m+1)\mathrm{T}})^{\mathrm{T}}$ と $\boldsymbol{k}^*=(k^{*(1)},\dots,k^{*(m)})^{\mathrm{T}}$ を，$\boldsymbol{\theta}=(\boldsymbol{\theta}^{(1)\mathrm{T}},\dots,\boldsymbol{\theta}^{(m+1)\mathrm{T}})^{\mathrm{T}}$ と $\boldsymbol{k}=(k^{(1)},\dots,k^{(m)})^{\mathrm{T}}$ の真値とする．いま，

$$\boldsymbol{\theta}^{*(1)}\ne\boldsymbol{\theta}^{*(2)}\ne\cdots\ne\ \boldsymbol{\theta}^{*(m+1)}, \tag{1.2}$$

を仮定し，$\boldsymbol{\theta}^*$ と $\boldsymbol{k}^*$ は未知であるとする．加えて，$\boldsymbol{\theta}^{*(1)},\dots,\boldsymbol{\theta}^{*(m+1)}$ は自然パラメータ空間に含まれるパラメータ集合の内点であり，そのパラメータ集合上で $\partial^2A(\boldsymbol{\theta})/\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^{\mathrm{T}}$ は正定値であるとする．さらに，後の漸近論のため，$1\le j\le m$ に対して $\lim_{n\to\infty}k^{*(j)}/n=\kappa^{(j)}$ であることを仮定する．ただし $0<\kappa^{(1)}<\cdots<\kappa^{(m)}<1$ としている．

2. 結果

$\hat{\boldsymbol{k}}_{\boldsymbol{x}}$ と $\hat{\boldsymbol{\theta}}_{\boldsymbol{x}}$ を $\boldsymbol{k}^*$ と $\boldsymbol{\theta}^*$ の $\boldsymbol{x}=(\boldsymbol{x}_1^{\mathrm{T}},\dots,\boldsymbol{x}_n^{\mathrm{T}})^{\mathrm{T}}$ に基づく最尤推定量とし，$f(\boldsymbol{x}|\boldsymbol{k}^*,\boldsymbol{\theta}^*)$ を $\boldsymbol{x}$ の同時確率関数とする．モデル選択は $f(\boldsymbol{y}|\boldsymbol{k}^*,\boldsymbol{\theta}^*)$ と $f(\boldsymbol{y}|\hat{\boldsymbol{k}}_{\boldsymbol{x}},\hat{\boldsymbol{\theta}}_{\boldsymbol{x}})$ の間の Kullback-Leibler ダイバージェンスの二倍

$$2\mathrm{KL}\{f(\boldsymbol{y}|\boldsymbol{k}^*,\boldsymbol{\theta}^*),f(\boldsymbol{y}|\hat{\boldsymbol{k}}_{\boldsymbol{x}},\hat{\boldsymbol{\theta}}_{\boldsymbol{x}})\}=2\mathrm{E}_{\boldsymbol{y}}\{\log f(\boldsymbol{y}|\boldsymbol{k}^*,\boldsymbol{\theta}^*)\}-2\mathrm{E}_{\boldsymbol{y}}\{\log f(\boldsymbol{y}|\hat{\boldsymbol{k}}_{\boldsymbol{x}},\hat{\boldsymbol{\theta}}_{\boldsymbol{x}})\},$$

を小さくしようとすること行うことができる．ここで，$\boldsymbol{y}$ は $\boldsymbol{x}$ のコピーであり，$\mathrm{E}_{\boldsymbol{y}}$ はその $\boldsymbol{y}$

に関する期待値を表すとする．右辺の第一項はモデルに依らないため，第二項を考えさえすればよい．そのシンプルな推定量は $-2\log f(\boldsymbol{x}|\hat{\boldsymbol{k}}_{\boldsymbol{x}},\hat{\boldsymbol{\theta}}_{\boldsymbol{x}})$ であるが，これは過小評価する．そこで，AIC 型の情報量規準では，それをバイアス補正した

$$\begin{aligned}&-2\log f(\boldsymbol{x}|\hat{\boldsymbol{k}}_{\boldsymbol{x}},\hat{\boldsymbol{\theta}}_{\boldsymbol{x}})+2\mathrm{E}_{\boldsymbol{x}}[\log f(\boldsymbol{x}|\hat{\boldsymbol{k}}_{\boldsymbol{x}},\hat{\boldsymbol{\theta}}_{\boldsymbol{x}})-\mathrm{E}_{\boldsymbol{y}}\{\log f(\boldsymbol{y}|\hat{\boldsymbol{k}}_{\boldsymbol{x}},\hat{\boldsymbol{\theta}}_{\boldsymbol{x}})\}]\\ &=-2\log f(\boldsymbol{x}|\hat{\boldsymbol{k}}_{\boldsymbol{x}},\hat{\boldsymbol{\theta}}_{\boldsymbol{x}})+2\mathrm{E}\Big[\sup_{(\boldsymbol{k},\boldsymbol{\theta})}L_{\boldsymbol{x}}(\boldsymbol{k},\boldsymbol{\theta})-L_{\boldsymbol{x}}\Big\{\operatorname*{argsup}_{(\boldsymbol{k},\boldsymbol{\theta})}L_{\boldsymbol{y}}(\boldsymbol{k},\boldsymbol{\theta})\Big\}\Big],\end{aligned}\tag{2.1}$$

を考える．ここで，E は $\boldsymbol{x}$ と $\boldsymbol{y}$ の両方での期待値を表し，また $L_{\boldsymbol{x}}(\boldsymbol{k},\boldsymbol{\theta})=\log f(\boldsymbol{x}|\boldsymbol{k},\boldsymbol{\theta})-\log f(\boldsymbol{x}|\boldsymbol{k}^*,\boldsymbol{\theta}^*)$ である．しかし，(2.1) における期待値は陽に求められないため，通常の AIC と同じようにその漸近評価を用いることを考える．つまり，(2.1) の代わりに

$$-2\log f(\boldsymbol{x}|\hat{\boldsymbol{k}}_{\boldsymbol{x}},\hat{\boldsymbol{\theta}}_{\boldsymbol{x}})+2\mathrm{E}\{b(\boldsymbol{k}^*,\boldsymbol{\theta}^*)\}\tag{2.2}$$

を考える．ここで，$b(\boldsymbol{k}^*,\boldsymbol{\theta}^*)$ は $\sup_{(\boldsymbol{k},\boldsymbol{\theta})}L_{\boldsymbol{x}}(\boldsymbol{k},\boldsymbol{\theta})-L_{\boldsymbol{x}}\{\operatorname{argsup}_{(\boldsymbol{k},\boldsymbol{\theta})}L_{\boldsymbol{y}}(\boldsymbol{k},\boldsymbol{\theta})\}$ の弱極限であるとする．ただし，$\sup_{(\boldsymbol{k},\boldsymbol{\theta})}$ と $\operatorname{argsup}_{(\boldsymbol{k},\boldsymbol{\theta})}$ は $L_{\boldsymbol{x}}(\boldsymbol{k},\boldsymbol{\theta})$ が $\mathrm{O_P}(1)$ あるいは正の値となるような $(\boldsymbol{k},\boldsymbol{\theta})$ の集合上でとるものとする．

$A'(\boldsymbol{\theta}^{*(j)})$ を $\partial A(\boldsymbol{\theta}^{(j)})/\partial\boldsymbol{\theta}^{(j)}|_{\boldsymbol{\theta}^{(j)}=\boldsymbol{\theta}^{*(j)}}$，$B_1^{(j)}(\boldsymbol{\theta}^*)=A(\boldsymbol{\theta}^{*(j+1)})-A(\boldsymbol{\theta}^{*(j)})-(\boldsymbol{\theta}^{*(j+1)}-\boldsymbol{\theta}^{*(j)})^{\mathrm{T}}A'(\boldsymbol{\theta}^{*(j)})$，$B_2^{(j)}(\boldsymbol{\theta}^*)=A(\boldsymbol{\theta}^{*(j)})-A(\boldsymbol{\theta}^{*(j+1)})-(\boldsymbol{\theta}^{*(j)}-\boldsymbol{\theta}^{*(j+1)})^{\mathrm{T}}A'(\boldsymbol{\theta}^{*(j+1)})$ とし，$Q_{k,\boldsymbol{x}}^{(j)}$ を

$$\begin{aligned}&I_{\{k<k^{*(j)}\}}\sum_{i=k+1}^{k^{*(j)}}[(\boldsymbol{\theta}^{*(j+1)}-\boldsymbol{\theta}^{*(j)})^{\mathrm{T}}\{\boldsymbol{T}(\boldsymbol{x}_i)-A'(\boldsymbol{\theta}^{*(j)})\}-B_1^{(j)}(\boldsymbol{\theta}^*)]\\ &+I_{\{k>k^{*(j)}\}}\sum_{i=k^{*(j)}+1}^{k}[(\boldsymbol{\theta}^{*(j)}-\boldsymbol{\theta}^{*(j+1)})^{\mathrm{T}}\{\boldsymbol{T}(\boldsymbol{x}_i)-A'(\boldsymbol{\theta}^{*(j+1)})\}-B_2^{(j)}(\boldsymbol{\theta}^*)].\end{aligned}$$

とする．すると次の定理が得られる．

定理 1．$\boldsymbol{x}$ が (1.1) にしたがい，また (1.2) が成立すると，(2.2) における漸近バイアスは

$$\mathrm{E}\{b(\boldsymbol{k}^*,\boldsymbol{\theta}^*)\}=\sum_{j=1}^{m}\mathrm{E}\Big(\sup_k Q_{k,\boldsymbol{x}}^{(j)}+Q^{(j)}_{\operatorname{argsup}_k Q^{(j)}_{k,\boldsymbol{y}},\boldsymbol{x}}\Big)+p_m,\tag{2.3}$$

で与えられる．ここで p_m は $\boldsymbol{\theta}$ における異なるパラメータの数である．

情報量規準を使うメリットの一つはその使いやすさであることを鑑み，$1\le j\le m$ に対して

$$\boldsymbol{\theta}^{*(j+1)}-\boldsymbol{\theta}^{*(j)}=\alpha_n^{-1/2}\Delta_{\boldsymbol{\theta}^*}^{(j)}\quad \text{かつ}\quad \mathrm{O}(1)\neq\alpha_n=\mathrm{o}(n)\tag{2.4}$$

なる近接条件を仮定する．ここで $\Delta_{\boldsymbol{\theta}^*}^{(j)}$ は定数ベクトルである．すると変化点推定量の漸近挙動が変わり，定理 1 の代わりに以下が得られる．

定理 2．(2.4) のもとで (2.2) における漸近バイアスは

$$\mathrm{E}\{b(\boldsymbol{k}^*,\boldsymbol{\theta}^*)\}=3m+p_m$$

で与えられる．

参 考 文 献

Ninomiya (2015). Change-point model selection via AIC. *Annals of the Institute of Statistical Mathematics* 67, 943–961.