

変化点モデルの非正則性と赤池情報量規準

二宮 嘉行 (統計数理研究所)

Yoshiyuki Ninomiya (The Institute of Statistical Mathematics)

1 変化点モデルの非正則性

最も簡単な例として, x_i ($1 \leq i \leq n$) が独立に確率密度関数

$$\begin{aligned} 1 \leq i \leq k \text{ のとき } & (2\pi)^{-1/2} \exp\{-(x - \mu^{(1)})^2/2\} \\ k+1 \leq i \leq n \text{ のとき } & (2\pi)^{-1/2} \exp\{-(x - \mu^{(2)})^2/2\} \end{aligned} \quad (1)$$

をもつ, というモデルを考える. つまり, i は時刻を表す添え字であるとし, 時刻 k までは $N(\mu^{(1)}, 1)$ に, 時刻 $k+1$ 以降は $N(\mu^{(2)}, 1)$ に x_i はしたがっているとす. $\mu^{(1)} \neq \mu^{(2)}$ であるとき, k を変化点といい, このモデルを変化点モデルという. いま, $k, \mu^{(1)}, \mu^{(2)}$ は未知のパラメータであるとする.

このモデルに対し, 対数尤度関数は

$$l_{1,x}(k, \mu^{(1)}, \mu^{(2)}) = -\frac{n}{2} \log(2\pi) - \frac{1}{2} \sum_{i=1}^k (x_i - \mu^{(1)})^2 - \frac{1}{2} \sum_{i=k+1}^n (x_i - \mu^{(2)})^2 \quad (2)$$

と書ける. これよりわかるのは, 対数尤度関数がパラメータ k で微分できないということである. 通常モデルでは対数尤度関数はパラメータで微分でき, 最尤推定量の漸近正規性はその事実を使って示されるが, 変化点モデルではこの微分不能性より k の最尤推定量は漸近正規性を満たさない.

変化点モデルには, 検定を考えた際にも通常モデルとは異なる側面がある. 変化点があるか否かを調べる検定のため,

$$1 \leq i \leq n \text{ のとき } (2\pi)^{-1/2} \exp\{-(x - \mu)^2/2\} \quad (3)$$

という変化のないモデルを考える. このモデルの分布は, (1) のモデルにおいて k の値が何であろうと $\mu^{(1)} = \mu^{(2)} = \mu$ で表現される. つまり, (1) のモデルにおいて $(k, \mu^{(1)}, \mu^{(2)}) = (1, \mu, \mu)$ と $(k, \mu^{(1)}, \mu^{(2)}) = (2, \mu, \mu), \dots, (k, \mu^{(1)}, \mu^{(2)}) = (n, \mu, \mu)$ は識別できず, どれも (3) の同じ分布になる. このことは「(1) のモデルは (3) のモデルにおいて識別不能性をもつ」と表現される. 混合分布モデルや因子分析モデルにもこの性質があり, それがゆえに尤度比検定統計量は漸近的に χ^2 分布にしたがわない.

変化点モデルはその微分不能性・識別不能性より非正則であり, 最尤推定・尤度比検定に対しては独自の漸近理論が展開されている. 次節以降それについて簡単に紹介する.

2 最尤推定量の漸近的性質

(1) のモデルにおける変化点の最尤推定量 \hat{k}_x を考える. $\bar{x}_s^t = \sum_{i=s+1}^t x_i / (t-s)$ と表現することによると, 変化点 k が与えられているときの $\mu^{(1)}$ と $\mu^{(2)}$ の最尤推定量, つまり (2) の最大点は $\hat{\mu}_{k,x}^{(1)} = \bar{x}_0^k$ と $\hat{\mu}_{k,x}^{(2)} = \bar{x}_k^n$ である. これを $l_{1,x}(k, \mu^{(1)}, \mu^{(2)})$ に代入すると

$$\begin{aligned} l_{1,x}(k, \hat{\mu}_{k,x}^{(1)}, \hat{\mu}_{k,x}^{(2)}) &= -\frac{n}{2} \log(2\pi) - \frac{1}{2} \sum_{i=1}^k (x_i - \bar{x}_0^k)^2 - \frac{1}{2} \sum_{i=k+1}^n (x_i - \bar{x}_k^n)^2 \\ &= -\frac{n}{2} \log(2\pi) - \sum_{i=1}^n x_i^2 + \frac{k}{2} (\bar{x}_0^k)^2 + \frac{n-k}{2} (\bar{x}_k^n)^2 \end{aligned} \quad (4)$$

が得られる. いま, パラメータ $(k, \mu^{(1)}, \mu^{(2)})$ の真値を $(k^*, \mu^{(1)*}, \mu^{(2)*})$ で表すことにする.

(4) から k に関する定数である $l_{1,x}(k^*, \hat{\mu}_{k^*,x}^{(1)}, \hat{\mu}_{k^*,x}^{(2)})$ をひくと

$$l_{1,x}(k, \hat{\mu}_{k,x}^{(1)}, \hat{\mu}_{k,x}^{(2)}) - l_{1,x}(k^*, \hat{\mu}_{k^*,x}^{(1)}, \hat{\mu}_{k^*,x}^{(2)}) = \frac{k}{2} (\bar{x}_0^k)^2 + \frac{n-k}{2} (\bar{x}_k^n)^2 - \frac{k^*}{2} (\bar{x}_0^{k^*})^2 + \frac{n-k^*}{2} (\bar{x}_{k^*}^n)^2$$

が得られる. これを最大にする k が最尤推定量 \hat{k}_x である. $k > k^*$ かつ $k - k^* = O(1)$ としたもとでこれが漸近的にどうなるかをみるため, 右辺を $\bar{x}_0^{k^*}$, $\bar{x}_{k^*}^k$, \bar{x}_k^n にわけて表現すると

$$\begin{aligned} & -\frac{k^*(k-k^*)}{2k} (\bar{x}_0^{k^*})^2 + \frac{k^*(k-k^*)}{k} \bar{x}_0^{k^*} \bar{x}_{k^*}^k + \frac{(k-k^*)^2}{2k} (\bar{x}_{k^*}^k)^2 \\ & + \frac{(n-k)(k-k^*)}{2(n-k^*)} (\bar{x}_k^n)^2 - \frac{(k-k^*)^2}{2(n-k^*)} (\bar{x}_{k^*}^k)^2 - \frac{(n-k)(k-k^*)}{n-k^*} \bar{x}_{k^*}^k \bar{x}_k^n \end{aligned}$$

となる. $\bar{x}_0^{k^*}$ と \bar{x}_k^n がそれぞれ $\mu^{(1)*}$ と $\mu^{(2)*}$ に収束すること, $k - k^*$ は k や $n - k^*$ と比べて無視できることを用いれば, これは

$$\begin{aligned} & -\frac{k-k^*}{2} \mu^{(1)*2} + (k-k^*) \mu^{(1)*} \bar{x}_{k^*}^k + \frac{k-k^*}{2} \mu^{(2)*2} - (k-k^*) \mu^{(2)*} \bar{x}_{k^*}^k + o_P(1) \\ & = (k-k^*) (\mu^{(1)*} - \mu^{(2)*}) (\bar{x}_{k^*}^k - \mu^{(2)*}) - \frac{k-k^*}{2} (\mu^{(1)*} - \mu^{(2)*})^2 + o_P(1) \\ & = \sum_{i=k^*+1}^k \left\{ (\mu^{(1)*} - \mu^{(2)*}) (x_i - \mu^{(2)*}) - \frac{1}{2} (\mu^{(1)*} - \mu^{(2)*})^2 \right\} + o_P(1) \end{aligned}$$

となる. $i \geq k^* + 1$ において x_i の期待値は $\mu^{(2)*}$ であるので, これの最右辺の中括弧の期待値は負である. $k < k^*$ かつ $k^* - k = O(1)$ としたもとでも同様の式が得られるので, 両側に伸びる負のドリフト付ランダムウォークを

$$Q_{k,x} = \mathbb{I}_{\{k > k^*\}} \sum_{i=k^*+1}^k \left\{ (\mu^{(1)*} - \mu^{(2)*}) (x_i - \mu^{(2)*}) - \frac{1}{2} (\mu^{(1)*} - \mu^{(2)*})^2 \right\}$$

$$+ I_{\{k < k^*\}} \sum_{i=k+1}^{k^*} \left\{ (\mu^{(2)*} - \mu^{(1)*})(x_i - \mu^{(1)*}) - \frac{1}{2}(\mu^{(1)*} - \mu^{(2)*})^2 \right\} \quad (5)$$

と定義すれば $k - k^* = O(1)$ ならば

$$l_{1,x}(k, \hat{\mu}_{k,x}^{(1)}, \hat{\mu}_{k,x}^{(2)}) - l_{1,x}(k^*, \hat{\mu}_{k^*,x}^{(1)}, \hat{\mu}_{k^*,x}^{(2)}) = Q_{k,x} + o_P(1)$$

であることがわかる。 $|k - k^*| \rightarrow \infty$ ならば $Q_{k,x} \rightarrow -\infty$ となることから想像できるように、 $|k - k^*| \rightarrow \infty$ ならば $l_{1,x}(k, \hat{\mu}_{k,x}^{(1)}, \hat{\mu}_{k,x}^{(2)}) - l_{1,x}(k^*, \hat{\mu}_{k^*,x}^{(1)}, \hat{\mu}_{k^*,x}^{(2)}) \rightarrow -\infty$ となる。これより $\hat{k}_x - k^* = O_P(1)$ であることがわかり、

$$\hat{k}_x - \underset{k}{\operatorname{argsup}} Q_{k,x} = o_P(1) \quad (6)$$

がいえる。 $\operatorname{argsup}_k Q_{k,x}$ は正規分布にしたがわないことから、 \hat{k}_x が漸近正規性を満たさないことがわかる。また、 $\hat{k}_x - k^* = O_P(1)$ からは、 $l_{1,x}(\hat{k}_x, \hat{\mu}_{\hat{k}_x}^{(1)}, \hat{\mu}_{\hat{k}_x}^{(2)}) = \sup_k l_{1,x}(k, \hat{\mu}_{k,x}^{(1)}, \hat{\mu}_{k,x}^{(2)})$ であることを使って

$$l_{1,x}(\hat{k}_x, \hat{\mu}_{\hat{k}_x}^{(1)}, \hat{\mu}_{\hat{k}_x}^{(2)}) - l_{1,x}(k^*, \hat{\mu}_{k^*,x}^{(1)}, \hat{\mu}_{k^*,x}^{(2)}) = \sup_k Q_{k,x} + o_P(1) \quad (7)$$

もいえる。

3 尤度比検定統計量の漸近的性質

(3) を帰無仮説のモデル、(1) を対立仮説のモデルとした尤度比検定を考える。帰無仮説のモデルのパラメータは μ のみであり、その対数尤度関数を $l_{0,x}(\mu)$ と書くことにする。 μ の最尤推定量は $\hat{\mu}_x = \bar{x}_0^n$ であるため

$$l_{0,x}(\hat{\mu}_x) = -\frac{n}{2} \log(2\pi) - \frac{1}{2} \sum_{i=1}^k (x_i - \bar{x}_0^n)^2 = -\frac{n}{2} \log(2\pi) - \sum_{i=1}^n x_i^2 + \frac{n}{2} (\bar{x}_0^n)^2$$

となる。(4) との差をとれば

$$l_{1,x}(k, \hat{\mu}_{k,x}^{(1)}, \hat{\mu}_{k,x}^{(2)}) - l_{0,x}(\hat{\mu}_x) = \frac{k}{2} (\bar{x}_0^k)^2 + \frac{n-k}{2} (\bar{x}_k^n)^2 - \frac{n}{2} (\bar{x}_0^n)^2$$

が得られる。これの k に関する最大値をとったものが尤度比検定統計量であり、それが漸近的にどうなるかを考える。ただし、変化点は端にない、具体的には δ を小さな正の定数として $\delta n \leq k \leq (1-\delta)n$ とする。右辺を変形して

$$\frac{1}{2} \frac{kn}{n-k} (\bar{x}_0^k - \bar{x}_0^n)^2 = \frac{1}{2} \frac{n^2}{k(n-k)} \left(\frac{1}{\sqrt{n}} \sum_{i=1}^k x_i - \frac{k}{n} \frac{1}{\sqrt{n}} \sum_{i=1}^n x_i \right)^2$$

と表現する. ここで $k = nt$ ($\delta \leq t \leq 1 - \delta$) として $n \rightarrow \infty$ とすると, 平方しているものの第一項目はあるウィナー過程 W_t に収束し, また第二項目の収束先は tW_1 と書けることがわかる. つまり, 尤度比検定統計量の収束として

$$\sup_{\delta n \leq k \leq (1-\delta)n} l_{1,\mathbf{x}}(k, \hat{\mu}_{k,\mathbf{x}}^{(1)}, \hat{\mu}_{k,\mathbf{x}}^{(2)}) - l_{0,\mathbf{x}}(\hat{\mu}_{\mathbf{x}}) \xrightarrow{d} \frac{1}{2} \sup_{\delta \leq t \leq 1-\delta} \frac{B_t^2}{t(1-t)}$$

がいえる. ここで, B_t はブラウンブリッジで, $E[B_t] = 0$ と $E[B_s B_t] = \min(s, t) - st$ を満たすものである. この右辺の裾確率については近似式が与えられており, それを用いて検定の p 値は評価される.

4 赤池情報量規準

赤池情報量規準 AIC (Akaike [1]) は漸近理論を用いて導かれるものであるので, 通常
の漸近理論を満たさない変化点モデルは AIC も独自のものとなる. 以下では, 変化点が
一個のモデルに対する AIC を導出する. 変化点が m 個のモデルに対する AIC も同様に
導出することができ, それらを比較することで変化点数の推定がおこなえる.

(1) のモデルに対する AIC をその元来の定義に基づいて考えると,

$$l_{1,\mathbf{x}}(\hat{k}_{\mathbf{x}}, \hat{\mu}_{\mathbf{x}}^{(1)}, \hat{\mu}_{\mathbf{x}}^{(2)}) - l_{1,\mathbf{x}}(\hat{k}_{\mathbf{y}}, \hat{\mu}_{\mathbf{y}}^{(1)}, \hat{\mu}_{\mathbf{y}}^{(2)}) \quad (8)$$

の弱極限を $b(k^*, \mu^{(1)*}, \mu^{(2)*})$ として, それは

$$-2l_{1,\mathbf{x}}(\hat{k}_{\mathbf{x}}, \hat{\mu}_{\mathbf{x}}^{(1)}, \hat{\mu}_{\mathbf{x}}^{(2)}) + 2E[b(k^*, \mu^{(1)*}, \mu^{(2)*})] \quad (9)$$

で与えられる. ここで, \mathbf{y} は \mathbf{x} のコピー, つまり \mathbf{x} と独立に同じ分布に従う確率ベクトルである. $E[b(k^*, \mu^{(1)*}, \mu^{(2)*})]$ は最大対数尤度の漸近バイアスと呼ばれる. それを評価するため, (8) を

$$\begin{aligned} & \{l_{1,\mathbf{x}}(\hat{k}_{\mathbf{x}}, \hat{\mu}_{\mathbf{x}}^{(1)}, \hat{\mu}_{\mathbf{x}}^{(2)}) - l_{1,\mathbf{x}}(k^*, \hat{\mu}_{k^*,\mathbf{x}}^{(1)}, \hat{\mu}_{k^*,\mathbf{x}}^{(2)})\} + \{l_{1,\mathbf{x}}(k^*, \hat{\mu}_{k^*,\mathbf{x}}^{(1)}, \hat{\mu}_{k^*,\mathbf{x}}^{(2)}) - l_{1,\mathbf{x}}(k^*, \mu^{(1)*}, \mu^{(2)*})\} \\ & + \{l_{1,\mathbf{x}}(k^*, \mu^{(1)*}, \mu^{(2)*}) - l_{1,\mathbf{x}}(\hat{k}_{\mathbf{y}}, \mu^{(1)*}, \mu^{(2)*})\} + \{l_{1,\mathbf{x}}(\hat{k}_{\mathbf{y}}, \mu^{(1)*}, \mu^{(2)*}) - l_{1,\mathbf{x}}(\hat{k}_{\mathbf{y}}, \hat{\mu}_{\mathbf{y}}^{(1)}, \hat{\mu}_{\mathbf{y}}^{(2)})\} \end{aligned}$$

と四つの項にわけて評価する. まず一つ目の項については (7) が成立している. 次に二つ目の項については, $-\sum_{i=1}^{k^*} \{(x_i - \bar{x}_0^{k^*})^2 - (x_i - \mu^{(1)*})^2\}/2 - \sum_{i=k^*+1}^n \{(x_i - \bar{x}_{k^*}^n)^2 - (x_i - \mu^{(2)*})^2\}/2 = k^*(\bar{x}_0^{k^*} - \mu^{(1)*})^2/2 + (n - k^*)(\bar{x}_{k^*}^n - \mu^{(2)*})^2/2$ と変形できるので, $\sqrt{k^*}(\bar{x}_0^{k^*} - \mu^{(1)*})$ と $\sqrt{n - k^*}(\bar{x}_{k^*}^n - \mu^{(2)*})$ をそれぞれ $N_{\mathbf{x}}^{(1)}$ と $N_{\mathbf{x}}^{(2)}$ と書いて

$$l_{1,\mathbf{x}}(k^*, \hat{\mu}_{k^*,\mathbf{x}}^{(1)}, \hat{\mu}_{k^*,\mathbf{x}}^{(2)}) - l_{1,\mathbf{x}}(k^*, \mu^{(1)*}, \mu^{(2)*}) = N_{\mathbf{x}}^{(1)2}/2 + N_{\mathbf{x}}^{(2)2}/2 \quad (10)$$

がいえ。三つ目の項については、 $\hat{k}_y > k^*$ のとき $\sum_{i=k^*+1}^{\hat{k}_y} \{(x_i - \mu^{(1)*})^2 - (x_i - \mu^{(2)*})^2\}/2 = \sum_{i=k^*+1}^{\hat{k}_y} \{(\mu^{(1)*} - \mu^{(2)*})(x_i - \mu^{(2)*}) - (\mu^{(1)*} - \mu^{(2)*})^2/2\}$ と変形できることからわかるように、それは (5) のランダムウォークを用いて $Q_{\hat{k}_y, x}$ と表現される。 (6) より $\hat{k}_y - \text{argsup}_k Q_{k, y} = o_P(1)$ であるから、

$$l_{1, x}(k^*, \mu^{(1)*}, \mu^{(2)*}) - l_{1, x}(\hat{k}_y, \mu^{(1)*}, \mu^{(2)*}) = Q_{\text{argsup}_k Q_{k, y, x}} + o_P(1) \quad (11)$$

がいえ。最後に四つ目の項について、まずそれは

$$\begin{aligned} & -\frac{1}{2} \sum_{i=1}^{\hat{k}_y} \{(x_i - \mu^{(1)*})^2 - (x_i - \bar{y}_0^{\hat{k}_y})^2\} - \frac{1}{2} \sum_{i=\hat{k}_y+1}^n \{(x_i - \mu^{(2)*})^2 - (x_i - \bar{y}_{\hat{k}_y}^n)^2\} \\ &= -\hat{k}_y (\bar{x}_0^{\hat{k}_y} - \mu^{(1)*})(\bar{y}_0^{\hat{k}_y} - \mu^{(1)*}) + \frac{\hat{k}_y}{2} (\bar{y}_0^{\hat{k}_y} - \mu^{(1)*})^2 \\ & \quad - (n - \hat{k}_y) (\bar{x}_{\hat{k}_y}^n - \mu^{(2)*})(\bar{y}_{\hat{k}_y}^n - \mu^{(2)*}) + \frac{n - \hat{k}_y}{2} (\bar{y}_{\hat{k}_y}^n - \mu^{(2)*})^2 \end{aligned}$$

と変形できることがわかる。ここで、 $\hat{k}_y - k^* = o_P(1)$ であること、またそれより $\bar{x}_0^{\hat{k}_y} - \mu^{(1)*} = \bar{x}_0^{k^*} - \mu^{(1)*} + o_P(1/n) = o_P(1)$ かつ $\bar{x}_{\hat{k}_y}^n - \mu^{(2)*} = \bar{x}_{k^*}^n - \mu^{(2)*} + o_P(1/n) = o_P(1)$, $\bar{y}_0^{\hat{k}_y} - \mu^{(1)*} = \bar{y}_0^{k^*} - \mu^{(1)*} + o_P(1/n) = o_P(1)$, $\bar{y}_{\hat{k}_y}^n - \mu^{(2)*} = \bar{y}_{k^*}^n - \mu^{(2)*} + o_P(1/n) = o_P(1)$ であることを用いると、これは \hat{k}_y を k^* で置き換えたものと漸近同値であることがわかる。よって

$$\begin{aligned} & l_{1, x}(\hat{k}_y, \mu^{(1)*}, \mu^{(2)*}) - l_{1, x}(\hat{k}_y, \hat{\mu}_y^{(1)}, \hat{\mu}_y^{(2)}) \\ &= -N_x^{(1)} N_y^{(1)} + N_y^{(1)2}/2 - N_x^{(2)} N_y^{(2)} + N_y^{(2)2}/2 + o_P(1) \end{aligned} \quad (12)$$

が得られる。ここで、 $N_y^{(1)}$ と $N_y^{(2)}$ はそれぞれ $\sqrt{k^*}(\bar{y}_0^{k^*} - \mu^{(1)*})$ と $\sqrt{n - k^*}(\bar{y}_{k^*}^n - \mu^{(2)*})$ である。 (7), (10), (11), (12) より (8) は

$$\begin{aligned} & \sup_k Q_{k, x} + Q_{\text{argsup}_k Q_{k, y, x}} \\ &+ N_x^{(1)2}/2 + N_x^{(2)2}/2 + N_y^{(1)2}/2 + N_y^{(2)2}/2 - N_x^{(1)} N_y^{(1)} - N_x^{(2)} N_y^{(2)} + o_P(1) \end{aligned}$$

となることがわかる。 $N_x^{(1)}$, $N_x^{(2)}$, $N_y^{(1)}$, $N_y^{(2)}$ がそれぞれ独立に標準正規分布にしたがうことを用いれば、(9) の漸近バイアス項はある程度評価でき、AIC は

$$-2l_{1, x}(\hat{k}_x, \hat{\mu}_x^{(1)}, \hat{\mu}_x^{(2)}) + 2E \left[\sup_k Q_{k, x} + Q_{\text{argsup}_k Q_{k, y, x}} \right] + 4 \quad (13)$$

で与えられる。 $2E[\sup_k Q_{k, x} + Q_{\text{argsup}_k Q_{k, y, x}}]$ が変化点パラメータ k に対する罰則項であり、4 がパラメータ $\mu^{(1)}$, $\mu^{(2)}$ に対する罰則項である。

5 罰則項の評価

変化点パラメータに対する罰則項は数値的に与えられるものではあるが、それでは罰則項が容易なものであるという情報量規準の特長の一つが消えてしまう。そこで、罰則項を容易な形で評価するため、ランダムウォークを平均と分散が同じウィーナー過程で近似する、という逐次解析でよく用いられる方法を考える (Siegmund [16])。まずは、 \boldsymbol{x} と \boldsymbol{y} が同じ分布にしたがうことを用い、(13) の変化点パラメータに対する罰則項を

$$2\mathbb{E}\left[\sup_k Q_{k,\boldsymbol{x}} + \frac{1}{2}\left|\operatorname{argsup}_k Q_{k,\boldsymbol{x}} - k^*\right|(\mu^{(1)*} - \mu^{(2)*})^2\right]$$

と書き直しておく。そして、 $\{W_t\}_{t \in \mathbb{R}}$ を $\mathbb{E}[W_t] = 0$ かつ $\operatorname{Var}[W_t] = |t|$ を満たすような両側に独立に伸びるウィーナー過程とし、これの $Q_{k,\boldsymbol{x}}$ を

$$C_t(\mu^{(1)*}, \mu^{(2)*}) = (\mu^{(1)*} - \mu^{(2)*})W_t - \frac{1}{2}(\mu^{(1)*} - \mu^{(2)*})^2 t$$

で置き換える。ここで $t = k - k^*$ である。実はこの近似は、変化量である $|\mu^{(1)*} - \mu^{(2)*}|$ が 0 に近づくという漸近論で保証されるものである。

この近似を用いると変化点パラメータの罰則項は

$$2\mathbb{E}\left[\sup_t C_t(\mu^{(1)*}, \mu^{(2)*}) + \frac{1}{2}\left|\operatorname{argsup}_t C_t(\mu^{(1)*}, \mu^{(2)*})\right|(\mu^{(1)*} - \mu^{(2)*})^2\right]$$

と書くことができる。そこで、負のドリフト付きウィーナー過程の最大値と最大点の分布に関する Bhattacharya & Brockwell [4] や Shepp [14] の結果を用いる。まず、 $\sup_t C_t(\mu^{(1)*}, \mu^{(2)*})$ の確率密度関数が $2 \exp(-x) - 2 \exp(-2x)$ ($x \geq 0$), 0 ($x < 0$) で与えられることより、

$$\mathbb{E}\left[\sup_t C_t(\mu^{(1)*}, \mu^{(2)*})\right] = \int_0^\infty x \{2 \exp(-x) - 2 \exp(-2x)\} dx = \frac{3}{2} \quad (14)$$

が得られる。また、 $g(t; a) = 6a^2 \exp(4a^2 t) \Phi(-3at^{1/2}) - 2a^2 \Phi(-at^{1/2})$ ($t \geq 0$) とすると $\operatorname{argsup}_t C_t(\mu^{(1)*}, \mu^{(2)*})$ の確率密度関数は $g(|t|; |\mu^{(1)*} - \mu^{(2)*}|/2)$ で与えられる。このことと、Stryhn [17] より $\int_0^\infty tg(t; a) dt = 3/(8a^2)$ であることを用いれば、

$$\begin{aligned} & \mathbb{E}\left[\frac{1}{2}\left|\operatorname{argsup}_t C_t(\mu^{(1)*}, \mu^{(2)*})\right|(\mu^{(1)*} - \mu^{(2)*})^2\right] \\ &= (\mu^{(1)*} - \mu^{(2)*})^2 \int_0^\infty tg\left(t; \frac{1}{2}|\mu^{(1)*} - \mu^{(2)*}|\right) dt = \frac{3}{2} \end{aligned} \quad (15)$$

が得られる。(14) と (15) より、変化点パラメータの罰則項は 6 と評価される。つまり、変化点パラメータは通常のパラメータの三倍の罰則をもつことになる。

6 一般化

ここまで最も単純な変化点モデルに対する漸近的性質を扱ってきたが、実際には一般化された形で論文や本が出版されている。例えば変化点の最尤推定量の漸近的性質は、一般の分布の枠組みで、通常のパラメータを既知とするケースにおいては Yao [19] が、通常のパラメータを未知とするケースにおいては Csörgő & Horváth [6] の 1.6 節が導いている。

変化点の有無を調べる尤度比検定統計量の漸近的性質については、Yao & Davis [21] が正規系列のケースを扱っており、一般のケースでは Csörgő & Horváth [6] の 1.1, 1.2, 1.3 節が詳細な議論を展開している。本稿では簡単のため変化点は端にないという制約をおいたが、それらでは制約なしの設定で漸近論を導いている。複数の変化点に関する尤度比検定は Bai & Perron [3] や Bai [2] で扱われている。後者では m 変化点と $m + 1$ 変化点を比較する尤度比検定統計量の漸近分布が導かれており、これを利用すれば多重検定方式による変化点数の推定がおこなえる。

AIC については、正規系列のケースを Ninomiya [11] が、一般のケースを Ninomiya [12] が扱っている。それらでは当然 m 変化点モデルのもとで AIC を導いており、やはり変化点パラメータの罰則は通常のパラメータの罰則の三倍となることを示している。つまり、通常のパラメータ数を p_m としたとき、AIC の罰則項は $6m + 2p_m$ になるということである。なお、BIC (Schwarz [13]) については AIC より早く Yao [20] でナイーブなものが提案されており、モデル選択の一致性をもつことが示されている。その後 Siegmund [15], Zhang & Siegmund [22], Hannart & Naveau [7] によって、変化点モデルの非正則性を考慮した BIC がベイズファクターより導かれている。

7 データ解析

変化点モデルに対する AIC としては、その非正則性を考慮せず、変化点パラメータにも通常のパラメータと同じ罰則を与えることも考えられる。これを AIC_{naive} と区別して書くことにする。以下では、AIC と AIC_{naive} は変化点解析のベンチマークデータにおいて異なる結論を導くことをみる。

図 1 は、Maguire et al. [10] で与えられて Jarrett [9] で修正された、英国鉱山事故のインターバルの時系列を表している。Worsley [18] はデータを独立に指数分布にしたがう系列とみて変化点解析をおこない、尤度比検定で変化点の一つあると結論付けた。導出した AIC はその結果を支持している一方で、 AIC_{naive} は 3 変化点モデルを選択している。

図 2 は、July 1, 1971 から August 2, 1974 までのダウ平均株価の週ごとのリターンレートの時系列を表している。Hsu [8] はデータを独立に共通の平均をもつ正規分布にしたが

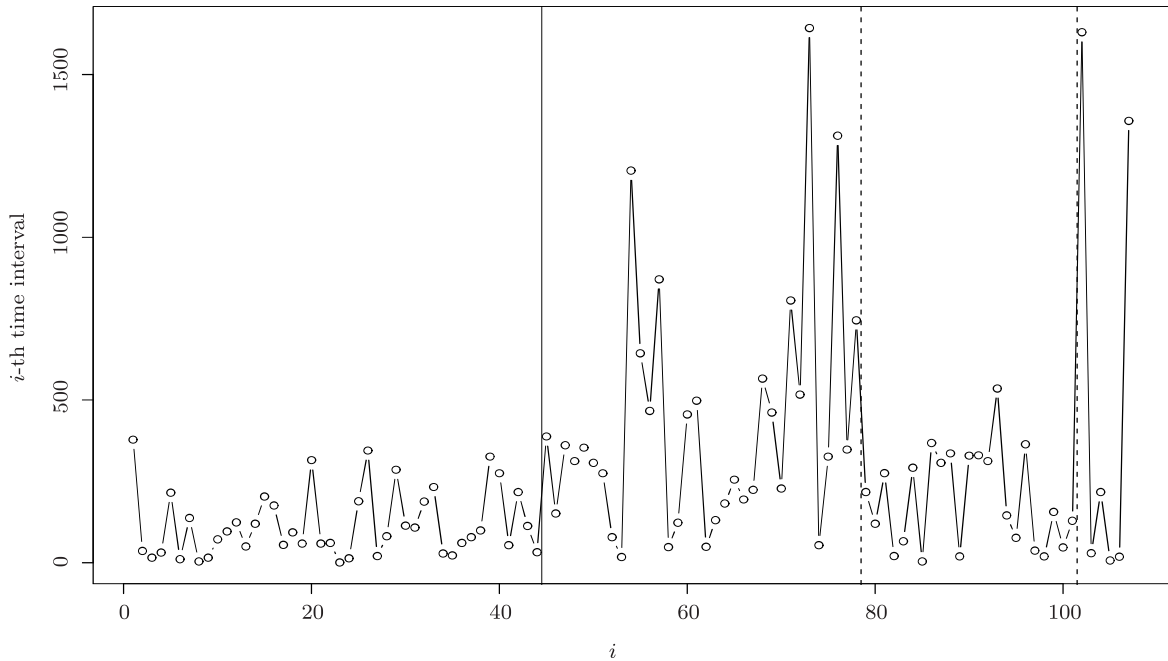


図 1: 鉾山事故データ. 垂直に引かれた実線は導出された AIC が選んだ変化点, 二つの点線は AIC_{naive} が新たに選んだ変化点を表す.

う系列とみて検定をおこない, 分散に変化点が一つあると結論付け, それはウォーターゲートに関するものであると推測した. 同じモデルで, Chen and Gupta [5] は変化点パラメータの罰則を無視した情報量規準を適用し, 新たに二つの変化点を検出した. 導出した AIC は 1 変化点モデルを選択している一方で, AIC_{naive} は Chen and Gupta [5] の結果を支持している.

参考文献

- [1] Akaike, H. (1973), “Information Theory and an Extension of the Maximum Likelihood Principle,” In *2nd International Symposium on Information Theory*, eds. B. N. Petrov and F. Csaki, Budapest: Akademiai Kiado, 716–723.
- [2] Bai, J. (1999), “Likelihood Ratio Tests for Multiple Structural Changes,” *Journal of Econometrics*, 91, 299–323.
- [3] Bai, J. and Perron, P. (1998), “Estimating and Testing Linear Models with Multiple Structural Changes,” *Econometrica*, 66, 47–78.

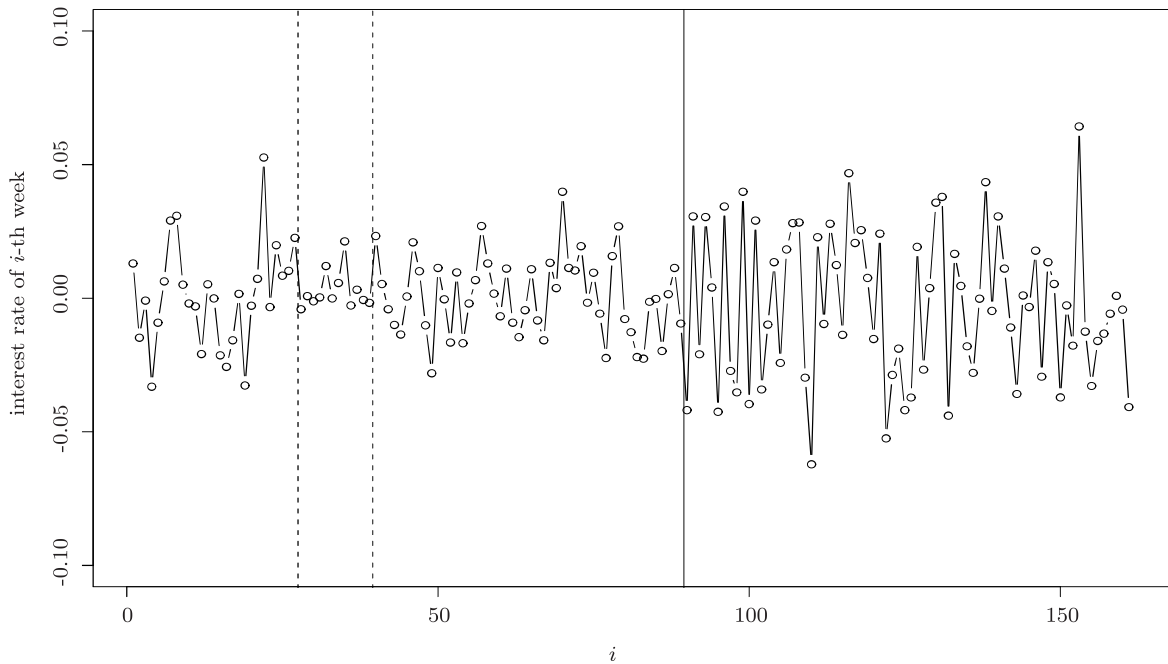


図 2: 米国株価データ. 垂直に引かれた実線は導出された AIC が選んだ変化点, 二つの点線は AIC_{naive} が新たに選んだ変化点を表す.

- [4] Bhattacharya, P. K. and Brockwell, P. J. (1976), “The Minimum of an Additive Process with Applications to Signal Estimation and Storage Theory,” *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, 37, 51–75.
- [5] Chen, J. and Gupta, A. K. (1997), “Testing and Locating Variance Changepoints with Application to Stock Prices,” *Journal of the American Statistical Association*, 92, 739–747.
- [6] Csörgő, M. and Horváth, L. (1997), *Limit Theorems in Change-Point Analysis*, John Wiley and Sons.
- [7] Hannart, A. and Naveau, P. (2012), “An Improved Bayesian Information Criterion for Multiple Change-Point Models,” *Technometrics*, 54, 256–268.
- [8] Hsu, D. A. (1979), “Detecting Shifts of Parameters in Gamma Sequences with Applications to Stock Price and Air Traffic Flow Analysis,” *Journal of the American Statistical Association*, 74, 31–40.
- [9] Jarrett, R. G. (1979), “A Note on the Intervals Between Coal-Mining Disasters,” *Biometrika*, 66, 191–193.

- [10] Maguire, B. A., Pearson, E. S. and Wynn, A. H. A. (1952), “The Time Intervals Between Industrial Accidents,” *Biometrika*, 39, 168–180.
- [11] Ninomiya, Y. (2005), “Information Criterion for Gaussian Change-Point Model,” *Statistics and Probability Letters*, 72, 237–247.
- [12] Ninomiya, Y. (2015), “Change-Point Model Selection via AIC,” *Annals of the Institute of Statistical Mathematics*, 67, 943–961.
- [13] Schwarz, G. (1978), “Estimating the Dimension of a Model,” *The Annals of Statistics*, 6, 461–464.
- [14] Shepp, L. A. (1979), “The Joint Density of the Maximum and its Location for a Wiener Process with Drift,” *Journal of Applied Probability*, 16, 423–427.
- [15] Siegmund, D. O. (2004), “Model Selection in Irregular Problems: Applications to Mapping Quantitative Trait Loci,” *Biometrika*, 91, 785–800.
- [16] Siegmund, D. O. (2013), *Sequential Analysis: Tests and Confidence Intervals*, Springer Science & Business Media.
- [17] Stryhn, H. (1996), “The Location of the Maximum of Asymmetric Two-Sided Brownian Motion with Triangular Drift,” *Statistics and Probability Letters*, 29, 279–284.
- [18] Worsley, K. J. (1986), “Confidence Regions and Tests for a Change-Point in a Sequence of Exponential Family Random Variables,” *Biometrika*, 73, 91–104.
- [19] Yao, Y. C. (1987), “Approximating the Distribution of the Maximum Likelihood Estimate of the Change-Point in a Sequence of Independent Random Variables,” *The Annals of Statistics*, 15, 1321–1328.
- [20] Yao, Y. C. (1988), “Estimating the Number of Change-Points via Schwarz’s Criterion,” *Statistics and Probability Letters*, 6, 181–189.
- [21] Yao, Y. C. and Davis, R. A. (1986), “The Asymptotic Behavior of the Likelihood Ratio Statistic for Testing a Shift in Mean in a Sequence of Independent Normal Variates,” *Sankhya*, 48, 339–353.
- [22] Zhang, N. H. and Siegmund, D. O. (2007), “A Modified Bayes Information Criterion with Applications to the Analysis of Comparative Genomic Hybridization Data,” *Biometrics*, 63, 22–32.