特異モデルにおけるベイズ検定と変化点発見への応用

藤原香織*　渡辺澄夫**
*日本IBM　東京基礎研究所
**東京工業大学　精密工学研究所

仮説検定

特異モデルでは、統計的正則モデルで広く用いられている
検定方法は使えない。

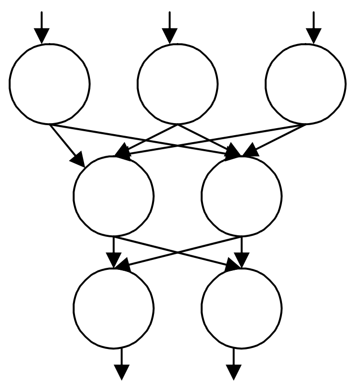
- 仮説検定とは
  - ある仮説が正しいかどうか、帰無仮説H_0 と、対立仮説H_1 とを比較し、いずれかを採択することで仮説の真偽を統計的に判断する。
  - 実用上は帰無仮説が棄却されることを期待して対立仮説を設定する。

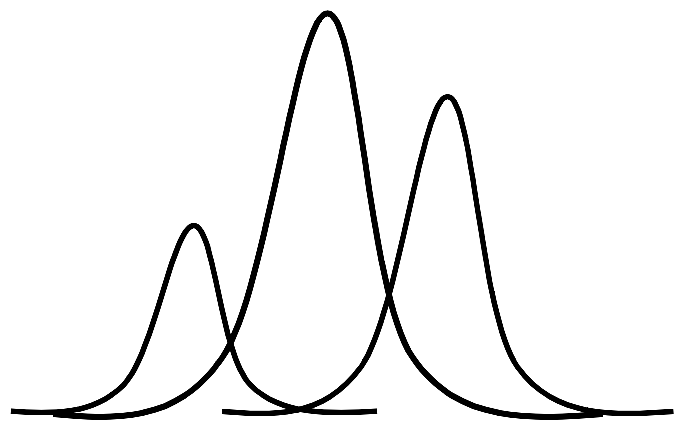
- 従来の検定方法
  - 統計的正則モデルでは、対数尤度比がカイ2乗分布に法則収束することを利用したカイ2乗検定などが広く用いられている。
  - 特異モデルでは最尤推定量は漸近正規性を持たず、ベイズ事後分布も正規分布に漸近しないため、別の方法が必要となる。

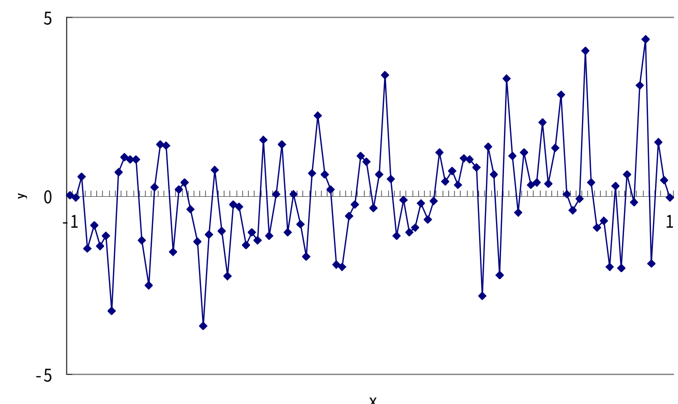
特異モデルとは

特異モデルはパターン認識、システム制御、時系列予測などの
データの構造を発見するさまざまな問題に用いられている。

例)


ニューラル
ネットワーク


混合正規分布



変化点問題

研究の目的

ベイズ尤度比を用いて特異モデルの仮説検定を行う。そのために、
特異モデルの確率的複雑さがどのような漸近挙動をするかを解明する。

<u>ベイズ尤度比(ベイズファクター)を用いた検定の概要</u>

パラメータ $w \in R^d$ によって定まる $x \in R^N$ の確率分布 $p(x|w)$ を考える。

- 帰無仮説 H0 : $\varphi_0(w)$
- 対立仮説 H1 : $\varphi_1(w)$ とすると、

ベイズ尤度比

$$L(X^n) = \frac{\int \prod_{i=1}^{n} p(X_i|w)\phi_1(w)dw}{\int \prod_{i=1}^{n} p(X_i|w)\phi_0(w)dw} \qquad (x^n = \{x_1, x_2, \ldots x_n\})$$

が、 $L(x^n) > \alpha$ → 対立仮説H1を採択 ($P(L(x^n) > \alpha \mid \varphi_0)$:危険率)

else → 帰無仮説H0を採択

$L(x^n)$ の解明は確率的複雑さ $F(x^n) = -\log L(x^n)$ の確率分布の解明と等価

ベイズ尤度比による仮説検定の変化点検出への応用

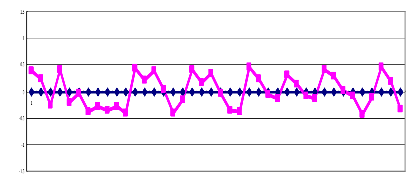
変化点検出問題の確率的複雑さの漸近挙動の理論解析を行い、
仮説検定を行った。

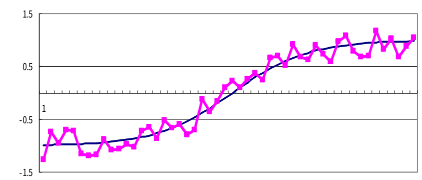
データのある点 $x_1, x_2, \ldots x_n$ が予め定められており、$y_1, y_2, \ldots y_n$ が

$y_i = f_w(x) = a\tanh(bx_i) + Z_i$ $\quad (Z_i \sim N(0, \sigma^2))$ という確率モデル

に従うとすれば、$Y = \{y_1, y_2, \ldots y_n\}$ の従う確率分布は、

$$p(Y|x, w, \sigma) = \left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^n \prod_{i=1}^{n} e^{-\frac{1}{2\sigma^2}(y_i - f_w(x))^2} \quad (w = \{a, b\})$$





- 帰無仮説　H0 : $\delta(w)$ ；データに変化がない場合
- 対立仮説　H1 : $\varphi(w)$ ；データに変化がある場合

とすれば、ベイズ尤度比は

$$L(Y) = \frac{\int e^{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i - f_w(x_i))^2}\varphi(w)dw}{e^{-\frac{1}{2\sigma^2}\sum_{i=1}^{n} y_i^2}} \quad (1)$$

主定理

$L(Y)$ は次の漸近分布に従う。

$$L(Y) \simeq \int e^{-nNa^2b^2+\sqrt{n}Mab}\varphi(a,b)dadb \qquad (2)$$

$$\left(M = \frac{1/\sqrt{n}\sum_{i=1}^{n} x_i y_i}{\sigma^2},\ N = \frac{1/n\sum_{i=1}^{n} x_i^2}{2\sigma^2}\right)$$

Φが

$$\varphi(a,b) = \begin{cases} \frac{1}{4} & (|a| \le 1, |b| \le 1, a \ne 0, b \ne 0) \\ 0 & (\text{上記以外}) \end{cases}$$

であるときは、

$$L(Y) = \int_0^{\sqrt{n}} \frac{dt}{\sqrt{n}}(-\log\frac{|t|}{\sqrt{n}})e^{-Nt^2}\cosh(Mt) \qquad (3)$$

データが帰無仮説から発生しているときには、サンプル数が増えるにつれて,
事後分布は特異点である原点の近傍に集中していく。
$L(Y)$ を原点の近傍でテイラー展開し、漸近分布に影響する部分のみ
考えると(2)が求められる。

$L(Y)$を用いた検定

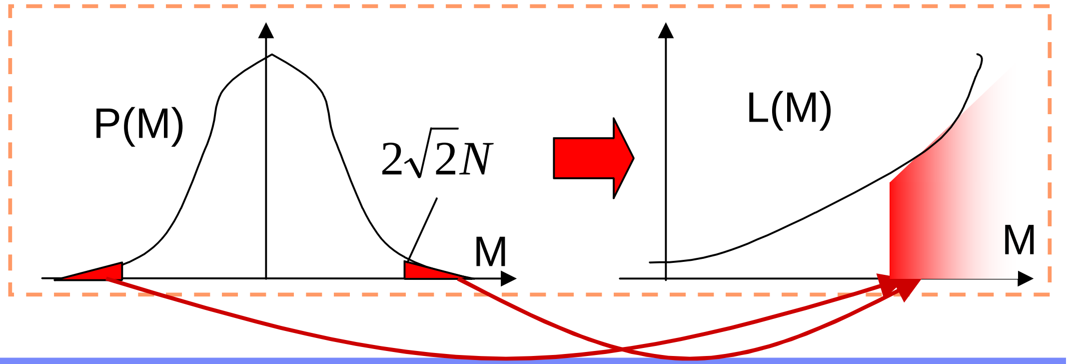
$$L(Y) = \int_0^{\sqrt{n}} \frac{dt}{\sqrt{n}} \left(-\log \frac{|t|}{\sqrt{n}}\right) e^{-Nt^2} \cosh(Mt) \quad (3)$$

$$\left(M = \frac{1/\sqrt{n}\sum_{i=1}^{n} x_i y_i}{\sigma^2},\ N = \frac{1/n \sum_{i=1}^{n} x_i^2}{2\sigma^2} \right)$$

Mは分散 $\sqrt{2N}$ の正規分布だが、式(3)はMの偶関数で単調増加関数なので、検定の際には危険率5%とすると $M = 2\sqrt{2N}$ として式(3)を計算し、確率的複雑さを求める。

この値を、データから直接求めた式(1)の値

から求められる確率的複雑さと比較することで行う。



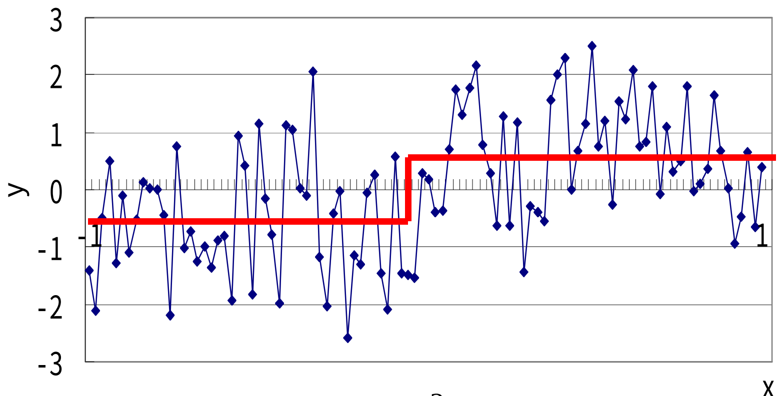
実験

実験により、ベイズ仮説検定を用いた変化点検出の有効性を検証した。

- 実験条件
  - サンプルは、$-1 \le x_i \le 1 \quad (i = 1,\ldots 100)$ なる x_i に対し y_i を次のように生成した。

$$\begin{cases} y_i = -\frac{\beta}{2} + Z_i & (x_i \le 0) \\ y_i = \frac{\beta}{2} + Z_i & (x_i > 0) \end{cases}$$

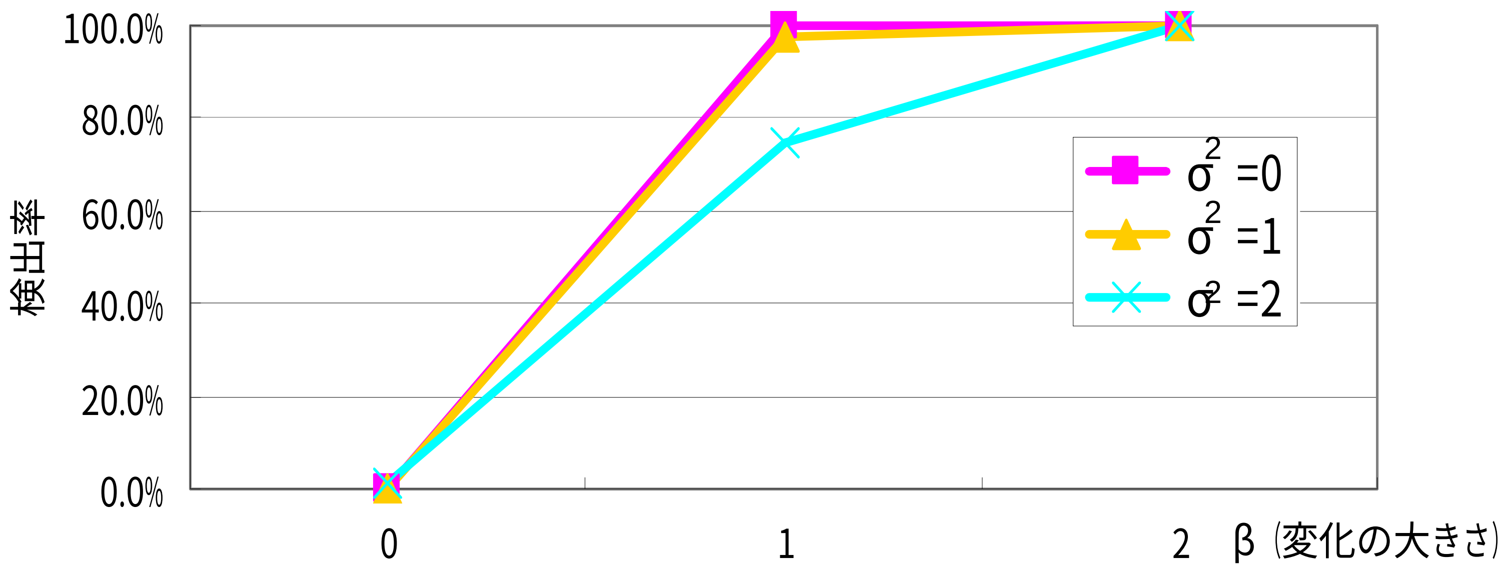
    ただし $Z_i \sim N(0, \sigma^2)$ である。
  - $\beta = 0,1,2, \quad \sigma^2 = 0,1,2$ と設定し、各条件ごとに300セットの学習例に対し変化点検出の検定を行った。



$\beta = 1, \sigma^2 = 1$ の例

結果

ノイズの分散が変化量と同程度であっても、変化点を適切に検出した。



$$\begin{cases} y_i = -\frac{\beta}{2} + Z_i & (x_i \le 0) \\ y_i = \frac{\beta}{2} + Z_i & (x_i > 0) \end{cases}$$

でノイズの大きさ σ^2・変化の大きさβ の条件ごとの変化点の検出率

まとめ

- 発表のまとめ
  - 特異モデルである変化点検出問題の、確率的複雑さの漸近挙動の理論解析を行った。
  - ベイズ対数尤度比を用いて変化点検出問題の仮説検定を行った。
  - 実験により検定の有効性を示した。
- 今後の予定
  - より柔軟に変化点を求められるよう、本変化点検出手法を拡張する。
  - 他に、応用上必要とされている特異モデルの漸近分布を解明する。

(補足)ベイズ仮説検定の危険率・検出力

ベイズ検定は、帰無仮説と対立仮説を決めたとき、
もっともよい検定法である。

- 危険率：帰無仮説が正しいときに対立仮説を選ぶ確率

$$P(L(x^n) > \alpha \mid \varphi_0) = \int \varphi_0 dw \int_{L(x^n)>a} dx_1 \cdots dx_n \prod_{i=1}^{n} p(x_i|w)$$

- 検出力：対立仮説が正しいときに対立仮説を選ぶ確率

$$P(L(x^n) > \alpha \mid \varphi_1) = \int \varphi_1 dw \int_{L(x^n)>a} dx_1 \cdots dx_n \prod_{i=1}^{n} p(x_i|w)$$

同じ危険率の元で検出力が最大である検定が最も良い検定である。対立仮説が決まれば、ベイズ検定以上に良い検定は存在しない。