

変化点検出問題におけるベイズ検定統計量の導出と 検出力の実験的考察

大橋 耕也^{1,a)} 渡辺 澄夫^{1,b)}

概要：変化点検出問題とは、ある対象の変化の有無をサンプルから統計的に推測する問題であり、対数周辺尤度比がカイ二乗分布へ法則収束しない非正則モデルであることが知られている。非正則モデルでは、従来の正則性を仮定した仮説検定法は適用できず、個々の問題に対して対数周辺尤度比の確率的挙動を導出する必要がある。本研究では、サンプルサイズに比例して変化部分が明瞭になるような対立仮説を設計することで、対数周辺尤度比をカイ二乗分布の関数で表す方法を提案し、また、変化部分が不明な場合における提案手法の検出力の挙動を数値的に明らかにする。

キーワード：変化点検出, ベイズ検定, カイ二乗検定

A Bayesian test statistic for change-point detection and numerical study of its power

KOYA OHASHI^{1,a)} SUMIO WATANABE^{1,b)}

Abstract: A change-point detection is a problem to statistically estimate whether certain data have change-point or not and it is known that a statistical model of a change-point detection is a nonregular model whose log-marginal likelihood ratio does not converge to a chi-square distribution in law. Hypothesis tests of regular models can not be applied to nonregular ones, hence we have to derive stochastic behavior of log-marginal likelihood ratio for each problem. In this paper, we propose that a Bayesian test for change-point detection can be conducted by chi-square test by setting an alternative hypothesis that a change-point becomes clear in proportion to sample size. Moreover, we show numerical experiments of our test's power even in the case change-point is unclear in advance.

Keywords: Change-point detection, Bayesian hypothesis test, Chi-square test

1. はじめに

変化点検出問題とは、ある対象の変化の有無をノイズを含んだサンプルから統計的に推論する問題であり、多くの分野への応用を持つ問題である。例えば、ネットワークセキュリティにおける不正アクセスやマルウェア検知、医薬品開発における有効性検証や毒性の用量反応曲線といったものが例として挙げられる。

変化点検出は、変化点がない場合に対応するパラメータ

がモデルに対して冗長であり、モデルが非正則になることが知られている [3]。このため、モデルの正則性を仮定したカイ二乗検定は適用できず、個々の問題で検定統計量の確率分布を導出する必要がある。

変化の様子を双曲線正接によりモデル化をした場合の仮説検定方法が提案されているが、急激な変化に対するモデルとしては不適であり、かつ変化点が固定であるとの制約があった [4]。

本研究では、変化の様子を階段関数によりモデル化し、また対立仮説にサンプルサイズに比例して変化部分が明瞭になるという性質を課すことで、検定統計量が漸近的にカ

¹ 東京工業大学 情報理工学院 数理・計算科学系

^{a)} ohashi.k.aj@m.titech.ac.jp

^{b)} swatanab@c.titech.ac.jp

イ二乗分布に従うことを示す。

本論文の構成は以下の通りである。2節でベイズ検定に関する基礎理論の説明を与える。3節では、変化点検出問題の定式化を行う。4節で主結果を述べ、5節でその証明を与える。最後に6節では、検出力に関する実験的考察を行う。

2. ベイズ検定の基礎理論

本節ではベイズ検定を概説し、対数周辺尤度比が最強力検定を与えることを説明する。

2.1 統計的仮説検定

仮説検定とは、サンプルから仮説の尤もらしさを、別の仮説との対立により検証する手法である。尤もらしさを検証する仮説を帰無仮説 (Null Hypothesis, 以下 N.H.), 比較対象となる仮説を対立仮説 (Alternative Hypothesis, 以下 A.H.) とよぶ。

仮説検定では、サンプル $X^n = \{X_1, \dots, X_n\}$ が N.H. から発生していると仮定したときに検定統計量 $S(X^n)$ が従う確率分布から、そのサンプルが N.H. に従う “不自然さ” を測る。つまり、仮説検定とは検定統計量 $S(X^n)$ と定数 a の組 (S, a) によって定まる以下のアルゴリズムのことである：

$$S(X^n) \leq a \Rightarrow \text{N.H.を採択}$$

$$S(X^n) > a \Rightarrow \text{N.H.を棄却}$$

(S, a) に対して2つの量が定義される。

定義 2.1 (有意水準)

$$\text{Level}(S, a) \equiv P(S(X^n) > a \mid \text{N.H.})$$

定義 2.2 (検出力)

$$\text{Power}(S, a) \equiv P(S(X^n) > a \mid \text{A.H.})$$

ただし、 $P(\cdot \mid \text{N.H.})$ はサンプルが N.H. に従う条件付き確率を表す。2つの検定 S_1 と S_2 について

$$\text{Level}(S_1, a_1) = \text{Level}(S_2, a_2)$$

$$\Rightarrow \text{Power}(S_1, a_1) \geq \text{Power}(S_2, a_2) \quad \text{for } \forall a_1, a_2$$

が成り立つとき、 S_1 は S_2 よりも強力な検定であるといい、任意の検定よりも強力な検定を最強力検定という。

2.2 ベイズ検定

ベイズ検定の枠組みを説明する。

確率密度関数 $p(x|w)$ を定めるパラメータ $w \in \mathbb{R}^d$ の確率分布に関する仮説検定をベイズ検定という。つまり、

$$\text{N.H. : } w \sim \varphi_0(w)$$

$$\text{A.H. : } w \sim \varphi_1(w)$$

という仮説を検定することになる。このとき、次の定理が成り立つ。

定理 2.1 $\forall a, b \in \mathbb{R}$ と任意の検定 S について、統計量

$$L(X^n) \equiv \frac{\int \varphi_1(w) \prod_{i=1}^n p(X_i|w) dw}{\int \varphi_0(w) \prod_{i=1}^n p(X_i|w) dw}$$

は次を満たす。

$$\text{Level}(L, b) = \text{Level}(S, a)$$

$$\Rightarrow \text{Power}(L, b) \geq \text{Power}(S, a)$$

(証明) 定数 a, b は

$$P(L(X^n) > b \mid \text{N.H.}) = P(S(X^n) > a \mid \text{N.H.})$$

を満たすものとする。このとき次の不等式を示せば良い：

$$P^* \equiv P(L(X^n) > b \mid \text{A.H.}) - P(S(X^n) > a \mid \text{A.H.}) \geq 0.$$

それぞれの検定の棄却域 A, B を定義する：

$$A = \{x^n; S(x^n) > a\}$$

$$B = \{x^n; L(x^n) > b\}.$$

このとき、

$$\begin{aligned} P^* &= \int \varphi_1(w) dw \left[\int_B - \int_A \right] \prod_{i=1}^n p(x_i|w) dx_i \\ &= \int \varphi_1(w) dw \left[\int_{B \setminus A} - \int_{A \setminus B} \right] \prod_{i=1}^n p(x_i|w) dx_i \end{aligned}$$

集合 $B \setminus A$, $A \setminus B$ でそれぞれ $L(x^n) > b$, $L(x^n) \leq b$ が成り立つから、

$$\begin{aligned} P^* &\geq b \int \varphi_0(w) dw \left[\int_{B \setminus A} - \int_{A \setminus B} \right] \prod_{i=1}^n p(x_i|w) dx_i \\ &= b \int \varphi_0(w) dw \left[\int_B - \int_A \right] \prod_{i=1}^n p(x_i|w) dx_i \\ &= 0 \quad (\because \text{有意水準が等しいとの仮定}) \end{aligned}$$

(証明終わり)

定理 2.1 から N.H と A.H. が確率分布の形で与えられるときは、周辺尤度比が最強力検定を与えることが明らかになった。

周辺尤度をそれぞれ $Z_0(X^n)$, $Z_1(X^n)$ と表記する。周辺尤度比の確率的挙動を調べることは、その対数

$$\log \frac{Z_1(X^n)}{Z_0(X^n)} = \log Z_1(X^n) - \log Z_0(X^n)$$

つまり、自由エネルギーの差分の確率的挙動を解明することと等価である。

N.H. が 1 点でのみ台を持つ場合: $\delta(w - w_0)$ を考える. 確率密度関数が正則である場合, 対数周辺尤度比がカイ二乗分布へ法則収束することを利用したカイ二乗検定が広く使われている. 一方で, 確率密度関数が非正則である場合, 自由エネルギー F はある定数 λ, m と確率変数 R を用いて

$$F = - \sum_{i=1}^n \log p(x_i | w_0) + \lambda \log n - m \log \log n + R$$

という挙動を持つことが知られている [1]. しかし, 確率変数 R の挙動は一般に解明されておらず個別の検定ごとに導出する必要がある.

以上から, 非正則モデルに対する最強力検定を作るためには, 対数周辺尤度比が従う確率分布を導出すれば良いことがわかった.

3. 変化点検出問題

本節では, 階段関数により変化の様子をモデル化した場合のベイズ検定の定式化を行う.

統計モデルは次で与える:

$$p(y|x, a, b) \propto \exp \left(-\frac{1}{2\sigma^2} (y - a\Theta(x - b))^2 \right)$$

ただし,

$$\Theta(x) = \begin{cases} 0 & (x \leq 0) \\ 1 & (x > 0) \end{cases}$$

であり, パラメータ $w = (a, b)$ と統計モデルが一対一対応ではないのでこのモデルは特異性をもっている. パラメータ a は変化量の大きさを, b は変化点をそれぞれ表す.

サンプル $Y^n = \{Y_1, \dots, Y_n\}$ はこの統計モデルに従う独立な確率変数とする. つまり,

$$Y_i = a\Theta(x_i - b) + N(0, \sigma^2) \quad i = 1, \dots, n.$$

ただし, 分散 σ^2 はあらかじめ固定する.

また本研究では, $\{x_1, \dots, x_n\}$ は区間 $[0, 1]$ を n 等分した各点を昇順に割り振った固定点とする:

$$x_i = \frac{i-1}{n} \quad i = 1, \dots, n.$$

このとき, 仮説を次のように設定する:

$$\text{N.H.: } \varphi_0(w) = \delta(a)\delta(b)$$

$$\text{A.H.: } \varphi_1(w) = N(a | 0, 1^2) \times U(b | c - n^{-\alpha}, c + n^{-\alpha}),$$

ただし, N は正規分布を, U は一様分布をそれぞれ表す.

N.H. は, サンプル Y^n が標準正規分布からそれぞれ独立に発生していることを表しており, 変化点がない場合に対応する. A.H. はサンプルサイズ n が大きくなるにつれて変化点 c が明瞭になるという性質をパラメータ b の分布に課している.

ここで, 変化は区間 $[0, 1]$ 内で起こるものと想定するた

め, 明瞭になる速さを表す α は

$$0 \leq c - n^{-\alpha}, c + n^{-\alpha} \leq 1 \Leftrightarrow \alpha \geq -\frac{\log \min(c, 1-c)}{\log n}$$

を満たす定数とする.

本研究では次節で述べる各定理を導出した.

4. 主結果

本研究の主結果として, 2つの定理と2つの系を述べる.

主定理 1

$$p(y|x, a, b) \propto \exp \left(-\frac{1}{2\sigma^2} (y - a\Theta(x - b))^2 \right)$$

$$\text{N.H.: } \varphi_0(a, b) = \delta(a)\delta(b)$$

$$\text{A.H.: } \varphi_1(a, b) = N(a | 0, 1^2) \times U(b | c - n^{-\alpha}, c + n^{-\alpha})$$

このとき, 対数周辺尤度比は漸近的に自由度 1 のカイ二乗分布に従う:

$$\begin{aligned} \log L(Y^n) &= \log \frac{\int \varphi_1(a, b) \prod_{i=1}^n p(Y_i | x_i, a, b) da db}{\int \varphi_0(a, b) \prod_{i=1}^n p(Y_i | x_i, a, b) da db} \\ &\rightarrow \frac{1}{2\sigma^2} \chi_1^2 + (\text{deterministic term}) \end{aligned}$$

主定理 1 は, 第 3 節で定義した仮説検定がカイ二乗検定に帰着されることを示している.

また, 検出力に関して次の定理が得られる.

主定理 2

$$a_0 \neq 0, b_0 \in (0, 1) \text{ に対して}$$

$$p(y|x, a_0, b_0) \propto \exp \left(-\frac{1}{2\sigma^2} (y - a_0\Theta(x - b_0))^2 \right)$$

からサンプルが生成されていると仮定する.

$b_0 \leq c$ のとき, 検出力は b_0 に依存しない.

主定理 2 は, 対立仮説の変化点 c よりも手前で実際は変化した場合, その変化点は検出力に影響しないことを示している. また, 主定理 2 の系として次の 2 つの系が得られる. ただし, 仮定は主定理 2 と同等のものを置いている.

系 1

$b_0 \leq c$ のとき, 変化量 $|a_0| \rightarrow \infty$ で, 検出力は 1 に近づく.

系 2

$b_0 \leq c$ のとき, サンプルサイズ $n \rightarrow \infty$ で, 検出力は 1 に近づく.

5. 主結果の証明

(主定理 1 の証明)

周辺尤度比 $L(Y^n)$ は,

$$L(Y^n) = \frac{\int \varphi_1(a, b) \prod_{i=1}^n \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(Y_i - a\Theta(x_i - b))^2} da db}{\prod_{i=1}^n \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}Y_i^2}}$$

であり, 具体的には

$$L(Y^n) = \frac{1}{\sqrt{2\pi}} \frac{1}{2n^{-\alpha}} \int \exp\left(-\frac{1}{2\sigma^2}(Pa^2 - 2Qa)\right) da db \quad (1)$$

ただし,

$$P = \sigma^2 + \sum_{i=1}^n \Theta(x_i - b)$$

$$Q = \sum_{i=1}^n Y_i \Theta(x_i - b).$$

ここで, $Pa^2 - 2Qa$ を a について変形すると,

$$Pa^2 - 2Qa = P \left(a - \frac{Q}{P}\right)^2 - \frac{Q^2}{P}$$

であるから, 式 (1) は a に関して積分が実行でき

$$L(Y^n) = \frac{1}{2n^{-\alpha}} \int_{c-n^{-\alpha}}^{c+n^{-\alpha}} \frac{1}{\sqrt{n}} \sqrt{\frac{\sigma^2}{1-b}} \exp\left(\frac{1}{2\sigma^2}\psi_n(b)^2\right) db$$

を得る. ただし,

$$\psi_n(b) \equiv \frac{1}{\sqrt{n}} \frac{1}{\sqrt{1-b}} \sum_{i=1}^n Y_i \Theta(x_i - b)$$

と置き, 計算過程で

$$\frac{P}{n} \rightarrow 1 - b \quad \text{as } n \rightarrow \infty$$

を利用した.

サンプルサイズ n が十分大きいとき, b の積分範囲は c の近傍に集中するので, 以下のような近似を施す:

$$\begin{aligned} L(Y^n) &\approx \frac{1}{2n^{-\alpha}} \frac{1}{\sqrt{n}} \exp\left(\frac{1}{2\sigma^2}\psi_n(c)^2\right) \int_{c-n^{-\alpha}}^{c+n^{-\alpha}} \sqrt{\frac{\sigma^2}{1-b}} db \\ &= \frac{n^{\alpha-\frac{1}{2}}}{2} \exp\left(\frac{1}{2\sigma^2}\psi_n(c)^2\right) \\ &\quad \times 2 \left(\sqrt{\sigma^2(1-c+n^{-\alpha})} - \sqrt{\sigma^2(1-c-n^{-\alpha})}\right). \end{aligned}$$

以上から, 対数周辺尤度比は

$$\log L(Y^n) = \frac{1}{2\sigma^2}\psi_n(c)^2 + (\text{deterministic term})$$

であり, n が十分大きいとき, $\psi_n(c)^2$ は自由度 1 のカイ二乗分布に従う. したがって, 検定統計量 $S(Y^n) = \psi_n(c)^2$ を用いたカイ二乗検定に帰着されることが示された.

(証明終わり)

(主定理 2 の証明)

サンプルがパラメータ (a_0, b_0) のモデルから発生したときの検出力は,

$$P(S(Y^n) > \chi_{1,\alpha}^2 \mid \text{A.H.}) = P(|\psi_n(c)| > z_\alpha \mid \text{A.H.})$$

ただし, $\chi_{1,\alpha}^2$, z_α はそれぞれ自由度 1 のカイ二乗分布の有意水準 α 点, 標準正規分布の両側検定の有意水準 α 点を表す.

$b_0 \leq c$ のとき, $Y_i \sim N(a_0, \sigma^2)$ であるから,

$$\begin{aligned} &|\psi_n(c)| > z_\alpha \\ \Leftrightarrow &\frac{1}{\sqrt{n}} \frac{1}{\sqrt{1-c}} \sum_{x_i > c} Y_i < -z_\alpha, z_\alpha < \frac{1}{\sqrt{n}} \frac{1}{\sqrt{1-c}} \sum_{x_i > c} Y_i \\ \Leftrightarrow &\frac{\sum_{x_i > c} Y_i/M - a_0}{\sqrt{\sigma^2/M}} < \sqrt{\frac{M}{\sigma^2}} \left(\frac{-z_\alpha \sqrt{n(1-c)}}{M} - a_0 \right), \\ &\sqrt{\frac{M}{\sigma^2}} \left(\frac{z_\alpha \sqrt{n(1-c)}}{M} - a_0 \right) < \frac{\sum_{x_i > c} Y_i/M - a_0}{\sqrt{\sigma^2/M}} \end{aligned}$$

ここで,

$$M \equiv \#\{x_i; x_i > c, i = 1, \dots, n\}$$

$$\frac{\sum_{x_i > c} Y_i/M - a_0}{\sqrt{\sigma^2/M}} \sim N(0, 1^2) \quad (\equiv S^*)$$

である. 以上から,

$$f_l(a_0, n, c, z_\alpha) \equiv \sqrt{\frac{M}{\sigma^2}} \left(\frac{-z_\alpha \sqrt{n(1-c)}}{M} - a_0 \right),$$

$$f_u(a_0, n, c, z_\alpha) \equiv \sqrt{\frac{M}{\sigma^2}} \left(\frac{z_\alpha \sqrt{n(1-c)}}{M} - a_0 \right)$$

とすると,

$$\begin{aligned} &P(S(Y^n) > \chi_{1,\alpha}^2 \mid \text{A.H.}) \\ &= P(S^* < f_l(a_0, n, c, z_\alpha), f_u(a_0, n, c, z_\alpha) < S^*) \end{aligned}$$

であり, 検出力は変化点 b_0 に依存しないことが示された.

(証明終わり)

(系 1 の証明)

$$f_l(a_0, n, c, z_\alpha), f_u(a_0, n, c, z_\alpha)$$

$$\rightarrow \begin{cases} -\infty & \text{as } a_0 \rightarrow +\infty \\ +\infty & \text{as } a_0 \rightarrow -\infty \end{cases}$$

であるから, 検出力は 1 に近づく.

(証明終わり)

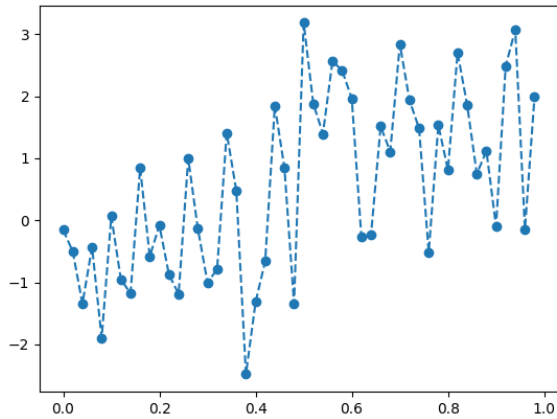


図 1 変化が検出された例
 $c = 0.5, a_0 = 1.0, b_0 = 0.5$

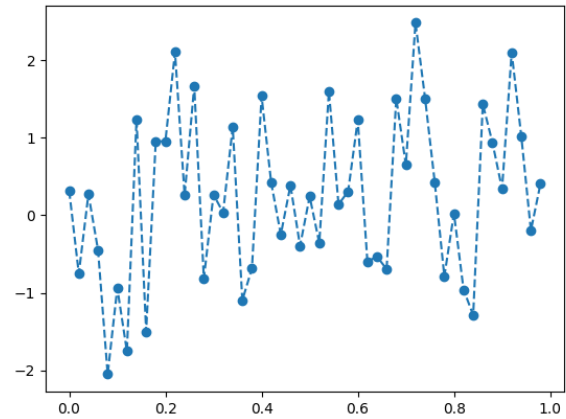


図 2 変化が検出されなかった例
 $c = 0.5, a_0 = 1.0, b_0 = 0.5$

(系 2 の証明)

$$\sqrt{\frac{M}{\sigma^2}} \left(\frac{\pm z_\alpha \sqrt{n(1-c)}}{M} - a_0 \right) = \pm z_\alpha \sqrt{\frac{1-c}{\sigma^2}} \sqrt{\frac{n}{M}} - \sqrt{\frac{M}{\sigma^2}} a_0$$

ここで,

$$\frac{n}{M} = \frac{n}{n - \lfloor nc \rfloor} = \frac{1}{1 - \frac{\lfloor nc \rfloor}{n}}$$

$$c \leq \frac{\lfloor nc \rfloor}{n} < c + \frac{1}{n}$$

であるから、はさみうちの原理より

$$\frac{n}{M} \rightarrow \frac{1}{1-c} \quad \text{as } n \rightarrow \infty$$

以上から $n \rightarrow \infty$ のとき、

$$\sqrt{\frac{M}{\sigma^2}} \left(\frac{\pm z_\alpha \sqrt{n(1-c)}}{M} - a_0 \right) \rightarrow \begin{cases} -\infty & (a_0 > 0) \\ +\infty & (a_0 < 0) \end{cases}$$

となり、検出力は 1 に近づく。

(証明終わり)

6. 検出力の実験的考察

本節では、実際に変化が生じたサンプルに対して仮説検定を行い、その検出力の挙動を実験的に考察をする。

6.1 予想した変化点と検出力の関係

本研究の提案手法では、あらかじめ変化点 c を予想し対立仮説に組み込む必要がある。そこで、予想した変化点 c と実際に変化が起きた点 b_0 がずれているときの検出力の挙動を数値実験により考察する。

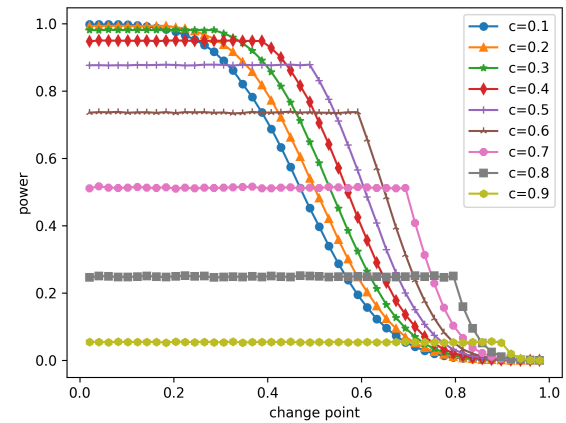


図 3 $a_0 = 1.0, n = 50$, 有意水準:0.05 での検出力の実験
主定理 2 の通り、(実際の変化点) $\leq$ (対立仮説の変化点) である場合検出力は一定であることが確認できる

実験の設定は以下の通り。

- 変化量 $a_0 = 1.0$
 - 分散 $\sigma^2 = 1.0$
 - サンプルサイズ $n = 50$
 - $b_0(\text{change point}) \in (0, 1)$
 - 各 b_0 に対して 10 万回ずつ検定を行い検出力を算出
- 得られた結果は図 3 の通りである。主定理 2 の通り、予想した変化点 c よりも手前で実際は変化が起きていた場合、検出力は一定値を取っていることが確認できる。また、各 c の周辺で実際に変化が起きた場合は、各 c での検出力が最大になっていることから、変化点があらかじめ予想できる場合はその周辺に c を設定するのが良いことも確認される。しかし、 $c < b_0$ の場合は検出力が急激に減少しているため、変化点の予想が難しい場合にどう対立仮説を設計すべきかという問題が生じてくる。

そこで、各 c における検出力の平均値を比較する。図 4

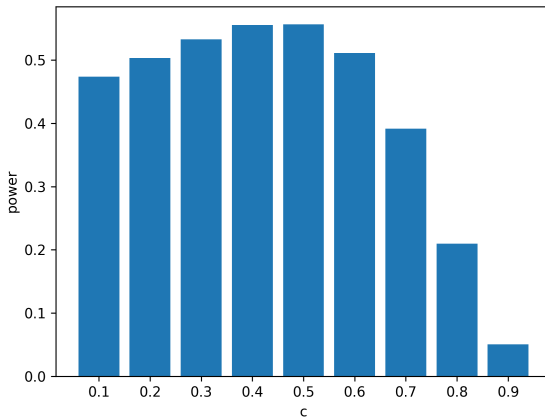


図 4 検出力の平均値による比較.
 $c = 0.5$ 付近で最大値を取ることがわかる

より, 平均的には $c = 0.5$ において検出力は最大となる. したがって, あらかじめ変化点に関する情報が無い場合, 変化は対象区間の中心付近で起きたとして対立仮説を設計するのが良いと結論付けられる.

6.2 サンプルサイズと検出力の関係

サンプルサイズ $n = 50, 100, 500, 1000$ において, 検出力の挙動を調べる. 結果は図 5 の通りである. 系 2 の通り, サンプルサイズが大きくなるにつれて $b_0 \leq c$ における検出力は 1 に近づくことが確認される.

また, サンプルサイズが大きい場合に $c < b_0$ における検出力の減少の仕方が緩やかになっている. これから, サンプルサイズが十分大きい場合, 任意の変化点に対して検出力は 1 に近づくことと予想される.

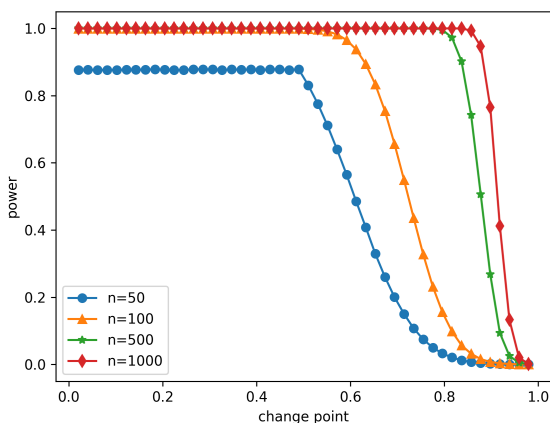


図 5 サンプルサイズと検出力の関係 ($a_0 = 1.0, c = 0.5$)
系 2 の通り, サンプルサイズが大きくなるにつれて検出力は 1 に近づくことが確認できる

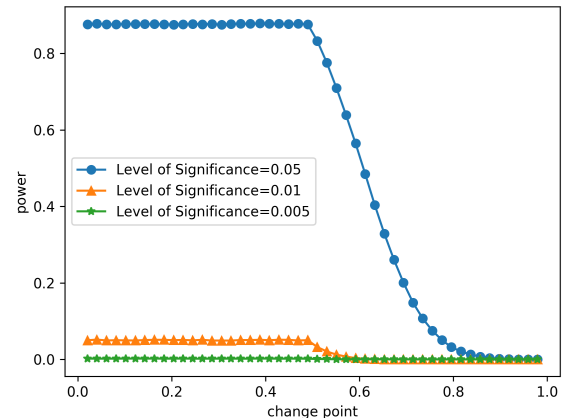


図 6 有意水準と検出力の関係 ($a_0 = 1.0, c = 0.5, n = 50$)
有意水準を小さくすると, 大幅に検出力が小さくなっている

6.3 有意水準と検出力の関係

有意水準として, 0.05, 0.01, 0.005 がよく用いられる. それぞれについて検出力との関係も確認する.

有意水準と検出力はトレードオフの関係にあるから, 有意水準を小さくしたら検出力も小さくなることはよく知られた事実である. しかし, 図 6 から, 有意水準に対して急激に検出力が減少していることが確認できる.

7. 結論

本論文では, 階段関数によりモデル化した変化点検出問題に対する, 対数周辺尤度比を用いたベイズ検定を提案した. 対立仮説にサンプルサイズに比例して変化点が明瞭になるという性質を課すことで, 検定統計量がカイ二乗分布に従うことを示し, その性質が検出力に与える影響を数値実験により考察を行った.

参考文献

- [1] 渡辺澄夫: ベイズ統計の理論と方法. コロナ社 (2012).
- [2] 渡辺澄夫: 代数幾何と学習理論. 森北出版株式会社 (2006).
- [3] 福水健次, 栗木哲, 竹内啓, 赤平昌文: 特異モデルの統計学. 岩波書店 (2004).
- [4] 藤原香織, 渡辺澄夫: 特異モデルにおけるベイズ検定と時系列解析への応用. 電子情報通信学会論文誌 D Vol. J91-D No. 4 pp.889-896 (2008).