MIZUHO

One MIZUHO

電子情報通信学会　情報理論研究会

特異モデルに対するベイズ仮説検定

~均一性の検定を題材に～

2023.08.03

みずほリサーチ&テクノロジーズ

サイエンスソリューション部

仮屋　夏樹

One MIZUHO
未来へ。お客さまとともに

謝辞

今回このような光栄な機会を頂き、電子情報通信学会　情報理論研究会　幹事の皆様に、改めてお礼を申し上げます。

本講演のきっかけとなった研究[NK+Watanabe2020] をはじめ、本稿で紹介した研究のうち筆者による研究は筆者が東京工業大学大学院在学時代の渡辺澄夫教授との共同研究に基づきます。

ただし言うまでもなく本稿に誤りがあった場合は、それは筆者に帰するものです。

一連の研究を通じ、継続的かつ親身な指導を下さった渡辺教授にこの場を借りてお礼を申し上げます。

みずほリサーチ＆テクノロジーズ　サイエンスソリューション部のご紹介

アウトライン

1. 問題設定・先行研究・研究の目的
2. ベイズ仮説検定の枠組み・特異学習理論の観点からの問題意識
3. 周辺尤度比の漸近解析による混合正規分布の均一性の検定
4. 変分ベイズ法を用いた均一性の検定
5. まとめと今後の課題

イントロダクション：混合分布

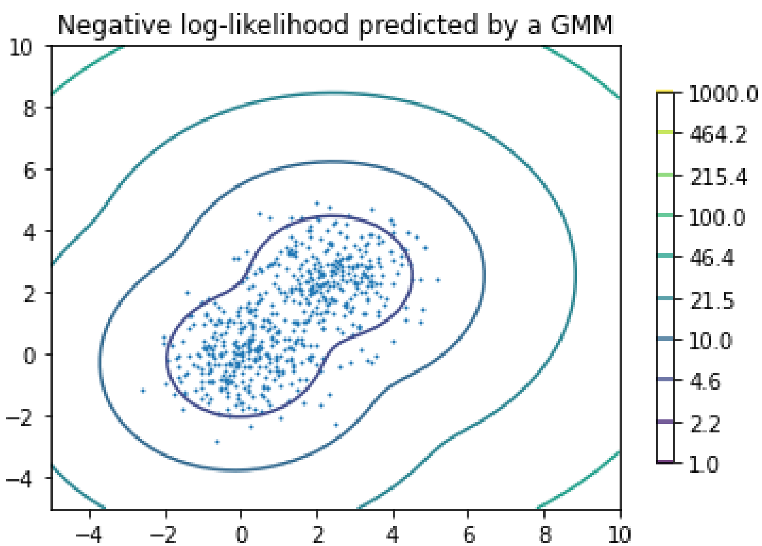
■ 異なる要因の影響が混在する多峰のデータを記述するためのモデルとして、
混合分布は様々な分野で活用されている。

混合分布の例：混合正規分布

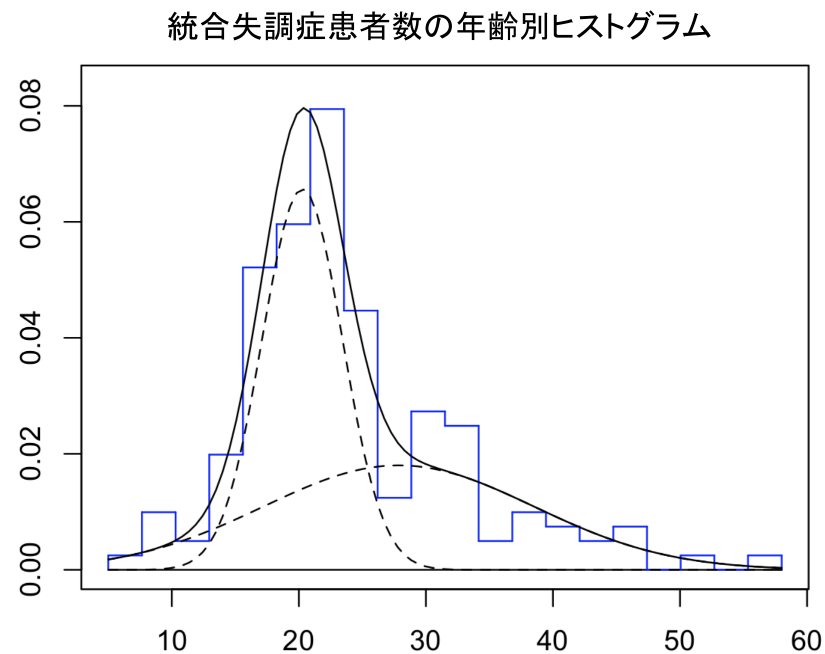
コンポーネント数2の混合正規分布

$$p_0(x|w) = (1-a)\mathcal{N}(0,1^2) + a\mathcal{N}(b,1^2),$$

$\boldsymbol{x} \in \mathbb{R}^d$



実データのフィッティング例



出典：Levine[1981]のデータをR Package “MixureInf”でフィッティング
http://cran.r-project.org/package=MixtureInf

問題設定：均一性の検定

- 混合分布の利用において、データを良く記述するコンポーネント数の決定は非常に重要である。
- このための有力なアプローチの一つとして仮説検定が用いられ、「均一性の検定」として知られている。

例：（2コンポーネント）混合正規分布における均一性の検定

1つの正規分布からデータが生成されたのか、2つの正規分布の混合からか？

確率モデル

$$p_0(x|w) = (1-a)\mathcal{N}(0, 1^2) + a\mathcal{N}(b, 1^2),$$

帰無仮説

$$H_0 : a = 0 \vee b = 0$$

（原点中心の単一の正規分布）

対立仮説

$$H_1 : otherwise$$

（複数の正規分布の混合）

問題設定：特異モデル

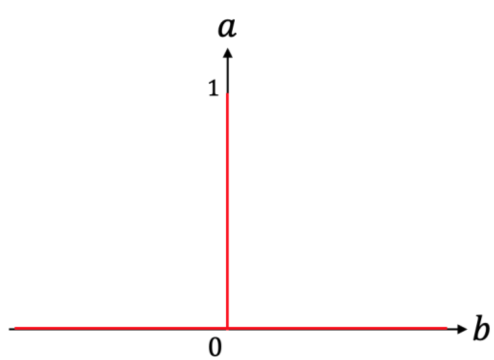
■ 統計的推論や仮説検定において、複雑な確率モデルを用いる場合には対象とする分布がモデルのパラメータと1対1対応しないことが生じうる。（特異モデル）

■ 結果として、よく知られている最尤推定やカイ2乗検定などが妥当性を失う。

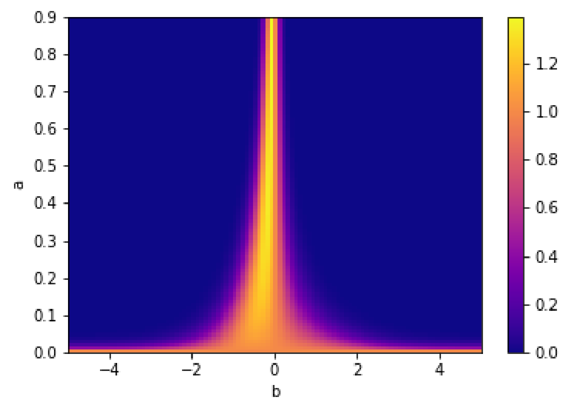
例：混合正規分布は特異モデル

確率モデル	$p_0(x\|w) = (1-a)\mathcal{N}(0,1^2) + a\mathcal{N}(b,1^2),$
帰無仮説	$H_0 : a = 0 \vee b = 0$ （原点中心の単一の正規分布）

パラメータ(a,b)空間での帰無仮説の分布



パラメータ(a,b)空間での対数尤度比



先行研究：特異モデルの均一性の検定

- 特異モデルの検定の難しさ（2コンポーネント混合正規分布でも研究多数）
  - 検定統計量（尤度比）が発散する（Hartigan[1985]）
  - 2つのコンポーネントが識別可能なケースでは、漸近形の導出が可能（Ghosh+Sen[1985]）
  - 識別可能でないケースの最尤法での扱いは難しく、理論研究が継続的になされてきている。（Garel[2001]（Gaussian Processの最大値で記述）, Liu+Shao[2003][2004]など）

- 一方で、正則化などの手法で、より実践的に検定を構築する取り組みもある。
  - 修正尤度比検定（Chen+Chen+Kalbfleisch,[2001]）
    検定統計量である対数尤度比（LLR）に補正項を追加（MLLR）。

$$MLLR = LLR + C \log\left[4a(1-a)\right]$$

  - EM test (Chen+Li[2009])
    EMアルゴリズムでMLRを最大化し、閾値に対する大小で採択仮説を決定。
  - D test (Charnigo+Sun[2004])
    データを帰無仮説/対立仮説に当てはめたときのL2距離で検定。

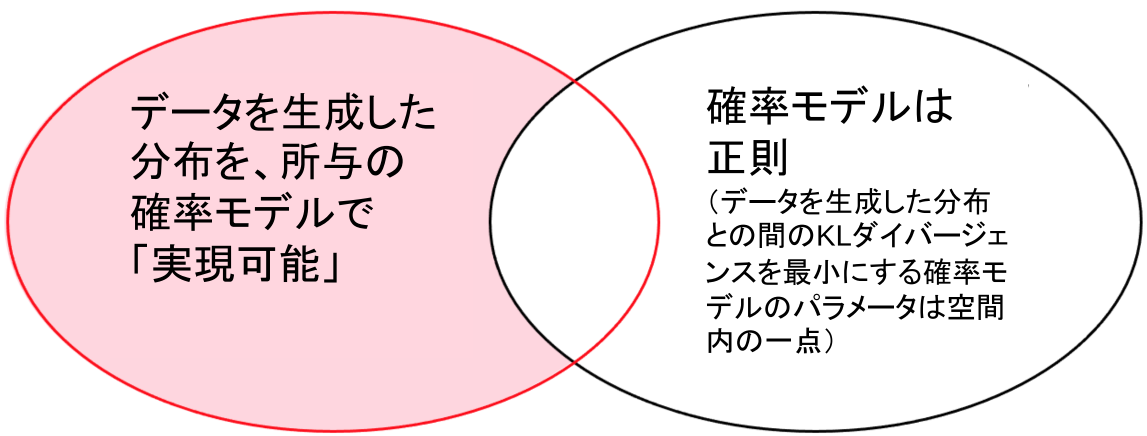
- これらの研究状況を踏まえると、理論に立脚したシンプルな仮説検定が構築できれば有意義ではないか、と考えられる。

研究の目的：特異モデルのベイズ学習理論

- 特異モデルを扱うことができる一般的な理論（特異学習理論）がベイズ統計の枠組みで、構築されつつある。（Watanabe[2018]）
- 現在まで統計的推論を中心にその有効性が立証されてきたが、検定の研究はごく僅か。
- 本研究では「特異モデルの仮説検定」を「特異学習理論」に基づき理論的に議論する。そのためのミニマルなモデルとして、「混合正規分布における均一性の検定」を扱う。

研究の対象：「特異モデル」かつ「実現可能」

解析のターゲット：正則でない（＝特異モデル）で、かつ実現可能（下の図の赤領域）
（本研究では帰無仮説に相当する確率分布を確率モデルで表現可能なことを前提とする）



アウトライン

1. 問題設定・先行研究・研究の目的
2. ベイズ仮説検定の枠組み・特異学習理論の観点からの問題意識
3. 周辺尤度比の漸近解析による混合正規分布の均一性の検定
4. 変分ベイズ法を用いた均一性の検定
5. まとめと今後の課題

イントロダクション：仮説検定

■ 仮説検定の枠組み

1. サンプルが帰無仮説（H0）、または対立仮説（H1）のもとで生成されたと仮定する。

帰無仮説	$H_0 : \{X^n\} \overset{iid}{\sim} P_0(X)$
対立仮説	$H_1 : \{X^n\} \overset{iid}{\sim} P_1(X)$

2. 1の仮定の下、サンプルの関数である「検定統計量」（T）の値を計算する。

3. 2で得られたTと予め設定したηを比較して、その大小に応じて採択する仮説を決定する。

$$\boxed{\begin{aligned} T(\{X^n\}) \leq \eta &\Rightarrow choose\ \ H_0 \\ T(\{X^n\}) > \eta &\Rightarrow choose\ \ H_1 \end{aligned}}$$

仮説検定の「ルール」は、
検定統計量Tと、採択の
基準を与える しきい値η
の組で定められる。

■ 「仮説検定をつくる」とは、検定統計量Tの確率的な挙動を明らかにした上で、必要となる検定の精度（どの程度過誤を許容するか）に応じηを定めることに他ならない。

ベイズ仮説検定の枠組み

- 帰無仮説と対立仮説はそれぞれパラメータの事前分布で与えられる。

- 混合正規分布の均一性検定のため、以降では以下の仮説を設定。

帰無仮説	（原点中心の単一の正規分布）
対立仮説	（複数の正規分布の混合）

- 検定統計量として周辺尤度比$L(X^n)$を選ぶとき、仮説検定は最強力検定となる。（Neyman-Pearson）

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

解析の対象

- 有意水準/検出力は帰無/対立仮説の下で$L(X^n)$が閾値ϵを上回る確率として定義される。

$$Level(L(\boldsymbol{X}^n, \epsilon)) = P(L(\boldsymbol{X}^n) > \epsilon|\varphi_0)$$
$$Power(L(\boldsymbol{X}^n, \epsilon)) = P(L(\boldsymbol{X}^n) > \epsilon|\varphi_1)$$

特異学習理論の枠組み

■ 特異学習理論で重要な役割を果たす量と計算の流れを下に示す。

特異学習理論で重要な量と計算の流れ

サンプル（所与）

$$\{X_1, X_2, \ldots X_n\} \in \mathbb{R}^N \overset{iid}{\sim} q(x).$$

モデル（人が決定）

確率モデル $p(x|w)$　　事前分布 $\varphi(w)$

事後分布

$$p(w|X^n) \equiv \frac{1}{Z}\varphi(w)\prod_i p(X_i|w)$$

周辺尤度

$$Z \equiv \int_W \varphi(w)\prod_i p(X_i|w)dw$$

自由エネルギー
（対数周辺尤度の符号反転）

$$F = -\log Z$$

仮説検定・統計的推定双方で重要な量

予測分布

$$p(x|X^n) \equiv \int_W dw p(x|w)p(w|X^n)$$

汎化損失

$$G \equiv -\int q(x)\log p(x|X^n)dx$$

統計的推定で重要な量

MIZUHO

特異学習理論で解明されている内容

- **推論・検定双方で重要な量である自由エネルギーの一般的な漸近形が導かれている。**
  （Watanabe[2001]）
  - ➢ ただし、具体的な漸近形はモデルごとに導出が必要である。

- **現在まで様々なモデルで汎化損失を定める「実対数閾値」を求める研究が行われてきた。**
  - ➢ 混合正規分布（山崎+渡辺[2003]）、縮小ランク回帰（青柳+渡辺[2005]）、NMF（林+渡辺[2017]等

自由エネルギー（対数周辺尤度の符号反転）の漸近形

$$F = -\log Z$$

$$\underbrace{F}_{\text{対数周辺尤度の符号反転}} = -\underbrace{\sum_i \log p(X_i|w_0)}_{\text{データ}X_i\text{を生成した真の分布}p(X_i|w_0)\text{による対数尤度}} + \lambda \log n - (m-1)\log\log n + \Theta(\xi_n) + o_p(1)$$

実対数閾値 λ（正則モデルではd/2）

多重度 m（特異モデル特有の項）

確率変数 $\Theta(\xi_n)$（正則モデルではカイ二乗分布）

汎化損失の漸近形（予測性能の指標であり、自由エネルギーがわかれば計算可能）

$$\underbrace{G}_{\text{汎化損失}} \equiv \underbrace{L_n(w_0)}_{\text{対数損失関数}} + \frac{\overset{\text{実対数閾値}}{\lambda}}{n} + o\left(\frac{1}{n}\right)$$

研究の問題意識:特異学習理論の観点から

- 一方、仮説検定の構築には自由エネルギー（対数周辺尤度の符号反転）のうち確率的に変動する項の把握が必要になる。これに関する研究は極めて限られる。
  - cf. 藤原＋渡辺[2008]、大橋＋渡辺[2017]

- 検定統計量として周辺尤度比を用いた仮説検定は、「最強力検定」となることが知られており、周辺尤度比を検定統計量とした仮説検定は特に重要である。

- 本研究では、特異モデルである「混合正規分布」を対象に、「自由エネルギー」（または実質的に等価な周辺尤度比）を理論的に解析し、その確率的挙動を明らかにする。

ベイズ仮説検定での検定統計量と自由エネルギーの関係

$$F - F_0 = -\log L(X^n) = -\log \frac{\int \varphi_1(w) \prod_i p(X_i|w)}{\int \varphi_0(w) \prod_i p(X_i|w)}$$

自由エネルギーの一般形と仮説検定で必要な項

$$F - F_0 = \lambda \log n - (m-1) \log\log n + \boxed{\Theta(\xi_n)} + o_p(1)$$

仮説検定に必要な項
（これを把握するのが本研究の解析のねらい）

アウトライン

1. 問題設定・先行研究・研究の目的
2. ベイズ仮説検定の枠組み・特異学習理論の観点からの問題意識
3. 周辺尤度比の漸近解析による混合正規分布の均一性の検定
4. 変分ベイズ法を用いた均一性の検定
5. まとめと今後の課題

問題設定

- ここまで見た通り、特異モデルを扱えるようなベイズ仮説検定構築のため、検定統計量の漸近挙動を明らかにする必要がある。

- 特異モデルの「混合正規分布」を対象に最強力検定の検定統計量である周辺尤度比を、理論的に解析してその確率的挙動を明らかにする。（NK+Watanabe[2022]）

- 利用する確率モデルとしては、（均一性の検定の理論的な研究でしばしば扱われる）コンポーネント数=2の混合正規分布を採用する。

- （均一性の検定の一般的な設定として）帰無仮説は単一の正規分布を表すものとし、対立仮説としては複数の正規分布の混合に相当するものを扱う。

- 具体的な仮説の設定については、次頁で述べる。

解析の対象

■ 複数パターンの仮説に対して帰無仮説の下周辺尤度比の漸近形を解析的に導出。

■ なお、対立仮説を与える各パラメータの事前分布として有限領域の一様分布を仮定。

解析対象のまとめ

確率モデル

$$p(x|w) = (1-a)\mathcal{N}(0,1^2) + a\mathcal{N}(b, \frac{1}{c}),$$

帰無仮説・対立仮説

Case	帰無仮説(N.H.)、対立仮説(A.H.)
Case 1 混合比が不定	$\mathrm{N.H.}: \varphi_0(w) = \delta(a)\delta(b)\delta(c-1),$ $\mathrm{A.H.}: \varphi_1(w) = U_a(0,1)\delta(b-\beta)\delta(c-1),$
Case 2 混合比・中心が不定	$\mathrm{N.H.}: \varphi_0(w) = \delta(a)\delta(b)\delta(c-1),$ $\mathrm{A.H.}: \varphi_1(w) = U_a(0,1)U_b(0,B)\delta(c-1),$
Case 3 混合比・中心・分散が不定	$\mathrm{N.H.}: \varphi_0(a,b,c) = \delta(a)\delta(b)\delta(c-1),$ $\mathrm{A.H.}: \varphi_1(a,b,c) = U_a(0,1)U(b,c),$ $U(b,c)$ は楕円板D上一様分布。$D = \left\{(b,c)\ ;\ b^2 + \frac{(c-1)^2}{2} \le \frac{R_0^2}{n}\right\}$

以降で取り上げる内容

解析手法:「スケーリング」則の利用

- 解析では特異点からの、Kullback-Leibler divergenceに基づくスケーリング則を利用。（Watanabe+Amari[2003]）

- これは正則モデルでの最尤推定に基づくカイ二乗検定の導出の一般化と考えても良い。

- 正則モデルでは対数尤度比を帰無仮説に相当するパラメータ（一点）からのTaylor展開＋中心極限定理で、尤度比がカイ二乗分布を従うことを導いた。（次ページ参照）

- 特異点解消定理に基づく適切な座標変換の下、パラメータ空間の特異点の周りでTaylor展開を考えることで、これと同様の手続きが特異モデルの場合にも可能である。

参考：正則モデルでのカイ二乗検定の導出のアウトライン

尤度比と帰無仮説・対立仮説

尤度比、
対数尤度比

$$\boxed{L(X^n) = \frac{\prod_i p_0(X_i|w^*)dw}{\prod_i p_0(X_i|w_0)dw}. \qquad l_n(w^*) \equiv \log L(X^n)}$$

$$\begin{aligned} \text{N.H.} &: \ w \sim \delta(w - w_0), \\ \text{A.H.} &: \ w \sim \delta(w - w^*), w_0 \neq w^*, \end{aligned}$$

対数尤度比のTaylor展開

対数尤度比を
帰無仮説のパ
ラメータ周りで
Taylor展開

スケーリング$n^{-1/2}$

$$\boxed{l_n(w_0 + \frac{\gamma}{\sqrt{n}})} = l_n(w_0) + \frac{1}{\sqrt{n}}\sum_{j=1}^{d}\frac{\partial l_n(w_0)}{\partial w_j^*}\gamma_j + \frac{1}{2n}\sum_{j,k}^{d}\frac{\partial^2 l_n(w_0)}{\partial w_j^*\partial w_k^*}\gamma_j\gamma_k + o(1)$$

$$= \sum_{j=1}^{d}\boxed{W_j}\gamma_j - \frac{1}{2}\sum_{j,k}^{d} I_{jk}\gamma_j\gamma_k + o(1)$$

n十分大で中心極限
定理から分散共分散
行列がFisher情報行
列となる多変量正規
分布に収束

n十分大でFisher情報
行列に収束

注）
このスライドでは一般的な
ケースを記述している。
パラメータ空間としてd次元
空間を考え、w_0, γとして多
次元ベクトルを考えている。

上を最大化する最尤推定解を代入した対数尤度比が従う分布

カイ2乗分布

対数尤度比を最大化するような解を求めると $2\max l_n(w^*) \sim \boxed{\chi_d^2}$

学習理論の観点からのコメント

- スケーリング則の具体例として、Case 2での扱いを見る。

Case 2での周辺尤度比の表式

$$L = \int_0^1 da \int_0^{\underset{\text{スケーリング}}{\boxed{B_0/\sqrt{n}}}} \frac{\sqrt{n}\,db}{B_0} \exp\Big(-\overset{\text{KL情報量のn倍}}{\boxed{\frac{n}{2}a^2b^2}} + \overset{\text{KL情報量の平方根（の定数倍）}}{\boxed{\sqrt{n}ab}}\,\xi_n + o_p(1)\Big)$$

特異点ab=0周りでabの2次式でかけている。

$$= \int_0^1 da \int_0^{B_0} \frac{db}{B_0} \exp\Big(-\frac{1}{2}a^2b^2 + ab\xi_n + o_p(1)\Big).$$

- 特異点”ab=0”周りでは被積分関数の対数は”ab”の2次式で書けている。(「標準形」)
  そのため”ab”の2次式とみなし、サンプルサイズに対するスケーリングを適用することで正則モデルのとき同様の議論ができた。

- ただし被積分関数の主要項はパラメータ空間内の1点での値で近似（＝鞍点法）はできず、”ab=0”という集合からの積分への寄与を考慮する必要がある。
  これが、特異モデル特有の挙動の背景である。

- モデルによらず、一般論として特異モデルの周辺尤度比の被積分関数が標準形で記述することが可能なことが知られているが（Watanabe[2018]）、特異点と、その周辺でのKL情報量の挙動はモデルごとに把握する必要がある。

結果（Case 3）

定理（周辺尤度比の漸近形）

右の表に示す条件で、帰無仮説を正とすると周辺尤度比は漸近的に以下の分布に分布収束する。
（この漸近形を用いることで、最強力検定を構成できる。）

$$L_\infty(\Xi_n) = \frac{1}{2R_0^2}\int_0^{R_0^2}\left(\frac{R_0}{\sqrt{t}}-1\right)\exp\left(-\frac{t}{2}\right) I_0(\sqrt{t}\Xi_n)\,dt.$$

$$\Xi_n = \sqrt{\xi_n^2+\eta_n^2},$$

$$\boxed{\begin{aligned}\xi_n &= \frac{1}{\sqrt{n}}\sum_{i=1}^n X_i,\\ \eta_n &= \frac{1}{\sqrt{2n}}\sum_{i=1}^n (X_i^2-1).\end{aligned}}$$

いずれも正規分布に収束する確率変数であり、これらよりLの確率的挙動は記述ができる。

Case 3	値
帰無仮説 対立仮説	$\mathrm{N.H.}: \varphi_0(a,b,c) = \delta(a)\delta(b)\delta(c-1),$ $\mathrm{A.H.}: \varphi_1(a,b,c) = U_a(0,1)U(b,c),$
スケーリング	$b^2 + \frac{(c-1)^2}{2} = R_0^2 \times n^{-1}$

数値計算による有意水準（Level）の検証

漸近形から導出したthreshold を用いて有限サンプルに対し有意水準を計算。漸近論はよく機能する。

level	10%	5%	1%
rejection region $L > r$	r=0.550	r=0.581	r=0.659
numerically calculated level(n=100)	9.53%	4.81%	1.21%
numerically calculated level(n=200)	9.72%	4.64%	0.88%
numerically calculated level(n=400)	10.04%	4.91%	0.79%
numerically calculated level(n=800)	10.33%	5.09%	1.03%

参考：ベイズファクターに基づく「仮説検定」へのコメント

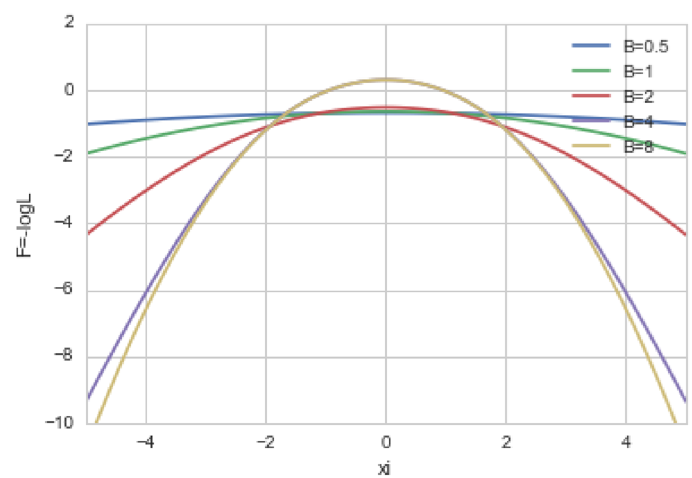
- 仮説の選択にあたり、周辺尤度比の確率分布ではなく周辺尤度比の値そのものの大小で仮説を選択するアプローチ（ベイズファクターによる「仮説検定」）もよく利用されている。
- 今回のCase 2を例に、対数周辺尤度比をプロットしたのが右図。
- Bが小さい例（＝帰無仮説と対立仮説が近いケース）では対数周辺尤度比の符号は負のまま。
- 帰無仮説と対立仮説が近いケースでの判別には、周辺尤度の確率分布に注目する方が好ましいと考えられる。

ベイズファクターによる「仮説検定」

サンプルから周辺尤度比を計算して、その大小で採択する仮説を決定する。
（例：Lが1より大きければ対立仮説を採択、そうでなければ帰無仮説を採択など）

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

確率変数（Xi）の値に対する対数周辺尤度比のプロット



ここまでのまとめ

- 特異モデルである混合正規分布を対象に、均一性の検定の理論をベイズ統計の枠組みに基づいて議論した。

- 特異点近傍で適切なスケーリング則することで、具体的な対立仮説に対して、周辺尤度比の漸近分布を系統的に導出できる。

- 以上の結果は以下の論文で報告されている。（NK+Watanabe[2022] ）
  Natsuki Kariya and Sumio Watanabe. “Asymptotic analysis of singular likelihood ratio of normal mixture by Bayesian learning theory for testing homogeneity”. Communications in Statistics-Theory and Methods 51.17 (2022): 5873-5888

アウトライン

1. 問題設定・先行研究・研究の目的
2. ベイズ仮説検定の枠組み・特異学習理論の観点からの問題意識
3. 周辺尤度比の漸近解析による混合正規分布の均一性の検定
4. 変分ベイズ法を用いた均一性の検定
5. まとめと今後の課題

問題意識

- 前章で見た通りベイズ仮説検定では事後分布の積分計算が必要であるが、一般的にはこれは計算コストを要する手続きである。

- この問題はベイズ統計に共通する課題だが、近似手段として変分ベイズ法が有力であることが知られている。

- しかし検定統計量に対して変分近似を行い、その確率分布を用いた仮説検定を構築した例はこれまでに知られていない。

- 学習理論の観点からは、次の点でこの問題に興味が持たれる。過去変分ベイズ法を用いた混合正規分布を対象にした、変分ベイズ法による自由エネルギーの導出は研究されてきた。

（Watanabe+Watanabe[2006][2007]）

- 上の先行研究ではO(log n)の導出結果が得られているものの、O(1)の確率変動する項の挙動は明らかではない。したがって、高次の項を含めた解析は理論的にも興味が持たれるものである。

- 以上の背景から、混合正規分布を対象に、変分ベイズを用いた均一性の検定について解析した。（NK+Watanabe[2020]）

解析対象

- 解析対象として、引き続き2コンポーネントの混合正規分布を扱う。
- 対立仮説の設定にあたっては、変分ベイズ法で標準的な共役事前分布として、混合比についてはディリクレ分布、中心については正規分布（分散は固定）を採用する。

確率モデル

$$p_0(x|w) = (1-a)\mathcal{N}(0, 1^2) + a\mathcal{N}(b, 1^2),$$

帰無仮説・対立仮説（事前分布）の設定

$$\varphi_0(a,b) = \delta(a)\delta(b),$$

$$\varphi_1(a,b) = Dir(a|\phi) \times \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2}b^2\right).$$

$$Dir(a|\phi) \equiv \frac{\Gamma\left(\sum_k \phi\right)}{\prod_k \Gamma(\phi)} \prod_k a_k^{\phi-1}.$$

解析手法：変分ベイズ

- 潜在変数yを導入して事後分布を書き直し、その上で潜在変数の確率分布とその他のパラメータの従う確率分布が独立であるような確率分布を用いて（KL情報量最小化の意味で）事後分布を近似する。

事後分布（潜在変数による記述）

y_{ik}: i番目のサンプルがk番目のクラスタに属していれば1、そうでなければ0を取る潜在変数。

$$\boxed{p(w, \{y_{ik}\} \,|X^n) \equiv \frac{1}{Z_n} \prod_{i,k} \left\{ a_k \frac{1}{\sqrt{2\pi}} e^{-(X_i - b_k)^2/2} \right\}^{y_{ik}} \varphi_1(w)}$$

潜在変数を用いた事後分布の記述

変分ベイズ

事後分布の近似として、潜在変数の確率分布とパラメータの確率分布の積となる関数形を仮定する。

その上で、以下のKullback-Leibler divergenceを最小にするような分布 $[q(\{y_{ik}\})r(w)]$
を近似解として求める。

$$D(qr||p) = \int dw \sum_{\{y_{ik}\}} q(\{y_{ik}\})r(w) \log \frac{q(\{y_{ik}\})r(w)}{p(w, \{y_{ik}\} \,|X^n)},$$

結果：変分自由エネルギーの表式

- 事前分布のハイパーパラメータφと潜在変数yの関数として、自由エネルギーの表式が得られた。
- これを最小にする潜在変数の配位が求めるものだが、漸近解析を行うと変分解および変分自由エネルギーの主要項がハイパーパラメータによって変化する相転移の存在が明らかになった。

混合正規分布の均一性の検定における変分自由エネルギーの表式

$$F - F_0 = \sum_i \sum_k \hat{y_{ik}} \log \hat{y_{ik}} + \log \frac{\Gamma(\sum_k \sum_i \hat{y_{ik}} + 2\phi)}{\prod \Gamma(\sum_i \hat{y_{ik}} + \phi)}$$
$$+ \frac{1}{2} \log \left(1 + \sigma^2 \sum_i \hat{y_{i1}} \right) - \frac{1}{2} \frac{(\sum X_j \hat{y_{j1}})^2}{\sum_i \hat{y_{i1}} + \frac{1}{\sigma^2}}$$
$$- \log \frac{\Gamma(2\phi)}{\prod(\Gamma(\phi))}$$

変分自由エネルギーの主要項とハイパーパラメータの関係

$$F - F_0 = \begin{cases} \phi \log n + o(\log n) & (\phi < 1) \\ \log n + o(\log n) & (otherwise) \end{cases}$$

主要項の挙動はハイパーパラメータに応じ変化
変分自由エネルギーはハイパーパラメータに応じて相転移を示す。

参考:導出のアウトライン(変分自由エネルギー)

■ KL情報量最小の条件から得られる変分条件

$$q(\{y_{ik}\}) \propto \exp\left[E_r\left\{\log p(w, \{y_{ik}\} \,|X^n)\right\}\right],$$

$$r(w) \propto \exp\left[E_q\left\{\log p(w, \{y_{ik}\} \,|X^n)\right\}\right].$$

から、潜在変数の分布とパラメータの分布が潜在変数・パラメータの平均値(変分解)で表せる。

$$\begin{aligned} q(y_{ik}) &\propto \exp\left[\sum_i \sum_k y_{ik}\left\{\langle \log a_k \rangle - \frac{1}{2}\langle (X_i - \delta_{k1} b)^2 \rangle\right\}\right] \\ &= \prod_i \prod_k \left\{\exp\left[\langle \log a_k \rangle - \frac{1}{2}\langle (X_i - \delta_{k1} b)^2 \rangle\right]\right\}^{y_{ik}} \end{aligned}$$

$$\begin{aligned} r(w) \propto \prod_k \prod_i a_k^{\hat{y_{ik}}} \exp\left[-\frac{1}{2}\left(\frac{1}{\sigma^2} + \sum_i \hat{y_{ik}}\right)\right] \\ \times \exp\left(b - \frac{\sum_i X_i \hat{y_{ik}}}{\frac{1}{\sigma^2} + \sum_i \hat{y_{ik}}}\right)^2 \end{aligned}$$

これが、もともと仮定した分布の関数系とコンシステントになる条件(セルフコンシステント方程式)より変分解を潜在変数の変分解のみで表すことが可能になる。

$$\langle \log a_k \rangle = \psi\left(\sum_i \hat{y_{ik}} + \phi\right) - \psi\left(n + 2\phi\right),$$

$$\langle b \rangle = \frac{\sum_i X_i \hat{y_{ik}}}{\sum_i \hat{y_{ik}} + \frac{1}{\sigma^2}},$$

$$\langle b^2 \rangle = \langle b \rangle^2 + \frac{1}{\sum_i \hat{y_{ik}} + \frac{1}{\sigma^2}}.$$

$$\begin{aligned} \hat{y_{i0}} &\propto \exp\left[\psi\left(\sum_i \hat{y_{i0}} + \phi\right) - \psi(n + 2\phi) - \frac{1}{2} X_i^2\right] \\ \hat{y_{i1}} &\propto \exp\left[\psi\left(\sum_i \hat{y_{i1}} + \phi\right) - \psi(n + 2\phi) - \frac{1}{2}\langle (X_i - b)^2 \rangle\right] \\ &= \exp\left[\psi\left(\sum_i \hat{y_{i1}} + \phi\right) - \psi(n + 2\phi)\right] \\ &\quad \times \exp\left[-\frac{1}{2}\left\{(X_i - \langle b \rangle)^2 + \frac{1}{\sum_i \hat{y_{i1}} + \frac{1}{\sigma^2}}\right\}\right] \end{aligned}$$

結果：変分ベイズ検定

- ハイパーパラメータφが1より大きいケースで、定数項オーダーまでの変分自由エネルギーの漸近分布を導出。
- ハイパーパラメータを転移点φ=1よりも大きく取り変分ベイズを適用すれば、変分自由エネルギーの漸近形はカイ2乗分布に従うことがわかった。

混合正規分布の均一性の検定における変分自由エネルギーの表式

$$
\begin{aligned}
F - F_0 = \log n &+ (\phi - 1)\log(\phi - 1) - \left(\phi - \frac{1}{2}\right)\log\left(\phi - \frac{1}{2}\right) \\
&+ \left(2\phi - \frac{3}{2}\right)\log\left(2\phi - \frac{3}{2}\right) + \frac{1}{2}\log\sigma^2 - \boxed{\frac{1}{2}\xi^2} \\
&- \frac{1}{2}\log 2\pi - \log\frac{\Gamma(2\phi)}{(\Gamma(\phi))^2} + o_p(1)
\end{aligned}
$$

$$\xi \sim \mathcal{N}(0, 1^2)$$

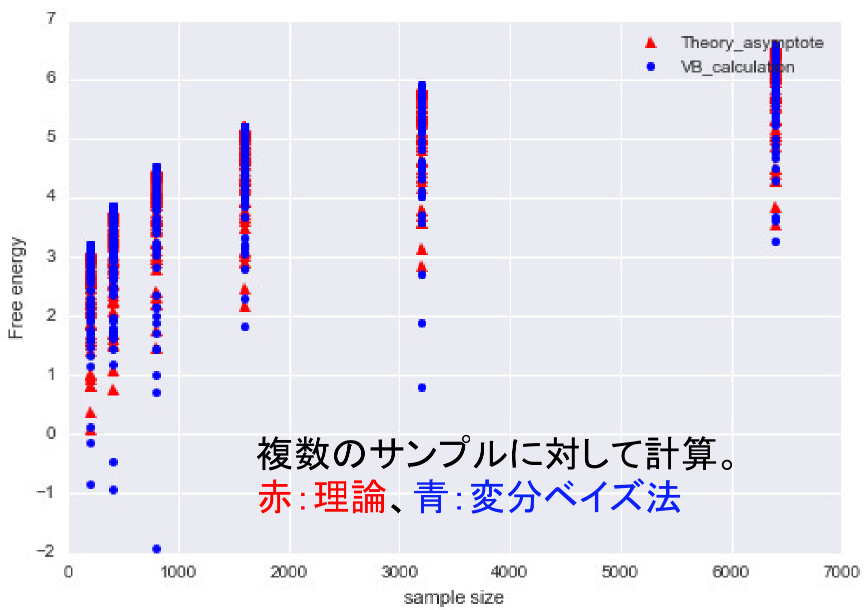
変分自由エネルギーの確率的に変動する項
⇒これを用いて仮説検定が可能

MIZUHO

結果：数値計算による変分ベイズ検定の検証

- 以上より、変分自由エネルギーを（繰り返し法で）計算することで仮説検定が構築可能であることがわかった。
- 人工的につくったサンプルに対しての数値実験を通じ、繰り返し法はよく漸近論と一致すること、また、変分ベイズ検定に基づく仮説検定が機能することがわかった。

変分自由エネルギーの繰り返し法での計算と漸近論比較



複数のサンプルに対して計算。
赤：理論、青：変分ベイズ法

変分ベイズ検定のアルゴリズムと数値検証

変分ベイズ検定のアルゴリズム

1. 変分ベイズ法（VB-EM）で変分自由エネルギーを計算。
2. 今回導出した変分ベイズ自由エネルギーの漸近分布から、求める有意水準に相当する閾値を決め、1で計算した値と比較する。

有意水準の数値検証

sample size	rejection rates		
n	10%	5%	1%
100	7.1%	3.4%	0.5%
200	8.4%	4.1%	0.8%
400	9.5%	4.5%	0.7%
800	9.6%	5.1%	1.1%

ここまでのまとめ

- 特異モデルである混合正規分布を対象に、変分ベイズ統計の枠組みに基づく仮説検定について議論した。
- 共役事前分布として広く用いられるディリクレ分布を混合比の事前分布として採用した場合、ハイパーパラメータの値に応じて相転移が発生する。
- ハイパーパラメータφ>1のケースで変分自由エネルギーの漸近分布が導出された。この結果をもとに均一性の検定（変分ベイズ検定）を行うことができる。
- 以上の結果は以下の論文で報告されている。（NK+Watanabe[2020]）
Natsuki Kariya and Sumio Watanabe. Testing homogeneity for normal mixture models: Variational bayes approach. IEICE TRANSACTIONS on Fundamentals of Electronics, Communications and Computer Sciences, 103 (11):1274–1282.

アウトライン

1. 問題設定・先行研究・研究の目的
2. ベイズ仮説検定の枠組み・特異学習理論の観点からの問題意識
3. 周辺尤度比の漸近解析による混合正規分布の均一性の検定
4. 変分ベイズ法を用いた均一性の検定
5. まとめと今後の課題

まとめと今後の課題

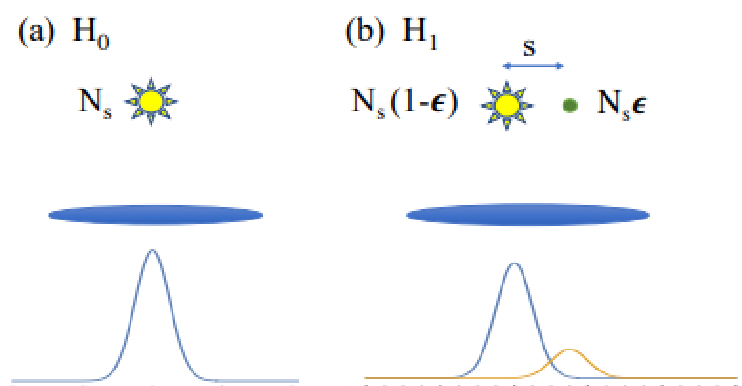
まとめ

- 特異モデルを扱う場合、仮説検定においても従来の枠組みを超えた理論が必要。
- 今回は混合正規分布を対象に、特異学習理論に基づいて特異モデルを扱う仮説検定について議論した。
- 漸近論の導出には、特異点周りでのスケーリング則を用いることで正則モデルでの議論の拡張としての導出が行える。
- 変分ベイズ自由エネルギーの漸近挙動を明らかにすることで、高次項の挙動を用いて仮説検定を行うこともできる。

今後の課題

- 今後の課題としては、
  - より多様な特異モデルに対する理論の構築
  - 現在正則理論を前提に議論している他分野(e.g. 理論物理学)へのアプローチ

  などが考えられる。



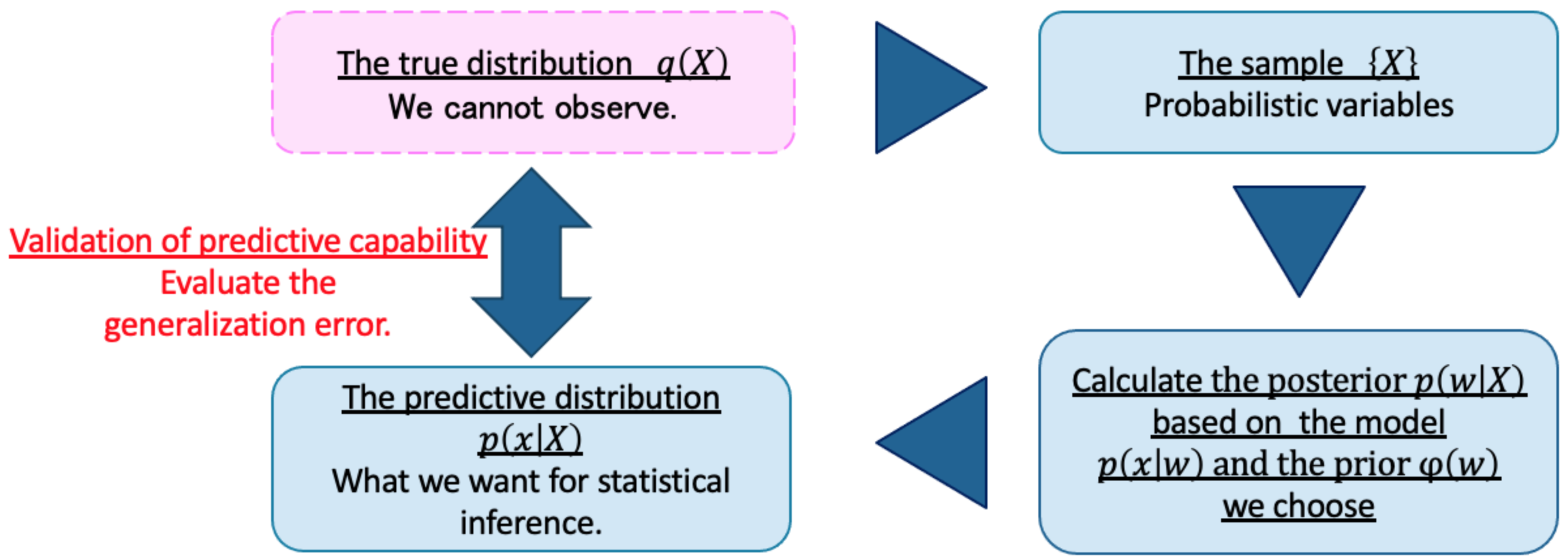
Phys. Rev. Lett. 127, 130502 (2021)

以下、参考資料

先行研究：ベイズ統計に基づく統計的推論の枠組み

- ベイズ統計に基づく統計的推論の枠組みの概念図を以下に示す。
- 所与のデータに対して、事前分布と確率モデルを設計し、事後分布を計算。
- 得られた事後分布からモデルのデータへの適合の指標（自由エネルギー）や、予測精度の指標（汎化損失）を計算することで、推論の妥当性を評価することができる。

$$\{X_1, X_2, \dots X_n\} \in \mathbb{R}^N \overset{iid}{\sim} q(x).$$



導出のアウトライン（相転移）

- 各項の評価・証明は煩雑なため、論文本体に譲る。
- ハイパーパラメータの値に応じて、自由エネルギーにおける主要項の出所が変わる。

$$F - F_0 = \sum_i \sum_k \hat{y_{ik}} \log \hat{y_{ik}} + \log \frac{\Gamma(\sum_k \sum_i \hat{y_{ik}} + 2\phi)}{\prod \Gamma(\sum_i \hat{y_{ik}} + \phi)}$$
$$+ \frac{1}{2} \log \left(1 + \sigma^2 \sum_i \hat{y_{i1}}\right) - \frac{1}{2} \frac{(\sum X_j \hat{y_{j1}})^2}{\sum_i \hat{y_{i1}} + \frac{1}{\sigma^2}}$$
$$- \log \frac{\Gamma(2\phi)}{\prod(\Gamma(\phi))}$$

青囲み
ハイパーパラメータφが１以下のときの主要項を含む箇所（ φ log n）

赤囲み
ハイパーパラメータφが１以上のときの主要項を含む箇所（1/2* log nがそれぞれ出る）