ベイズ事後分布の相転移について

渡辺澄夫
東京工業大学

相転移とは何ですか

学生のみなさまから「学習理論における相転移とは具体的に何か」という質問をいただきましたので、具体的な例について紹介します。

相転移とは

データの数 n やハイパーパラメータ α を変化させたとき、ある値を境界としてその前後で事後分布が急激に形を変えることにより、ベイズ自由エネルギーや汎化損失の値がそれらのなめらかな関数でなくなることがあります。そのような変化を相転移と呼びます。

具体例として、ここでは次の問題を考えます。

(1) データの数 n が小さいうちは真の分布の構造を詳しく知ることはできないが
データの数が増えるにつれて真の分布の構造が見えてくる(構造の発見)。

(2) 事前分布を決めているハイパーパラメータを変えると事後分布が急激に
変化する点がある (「事前分布は何でもよい」ではないのです)。

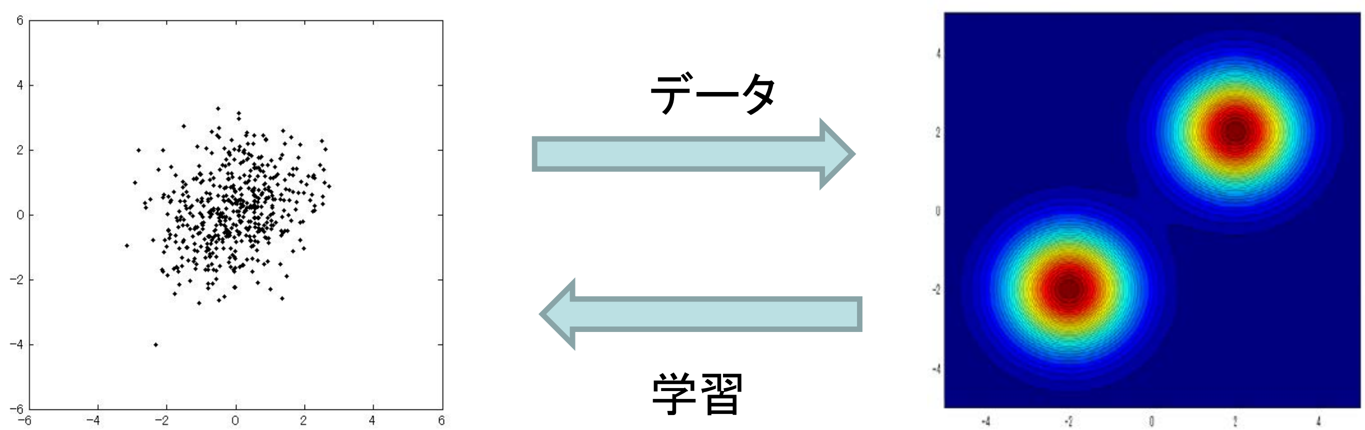
混合正規分布の例で考えます。

混合正規分布の例を考えてみましょう。「１山か２山か」の見分けがつきにくいデータに対して、２山の統計モデルを使って推測してみます。

混合正規分布:　$x \in \mathbf{R}^2$ の確率密度関数で $w=(a,b,c)$ をパラメータに持つもの

$$p(x|w) = (1-a)\, N(x-b) + a\, N(x-c)$$

パラメータ w は $(0 \leqq a \leqq 1,\ \ b, c \in \mathbf{R}^2)$.　$N(x)$ は $x \in \mathbf{R}^2$ の正規分布とします.



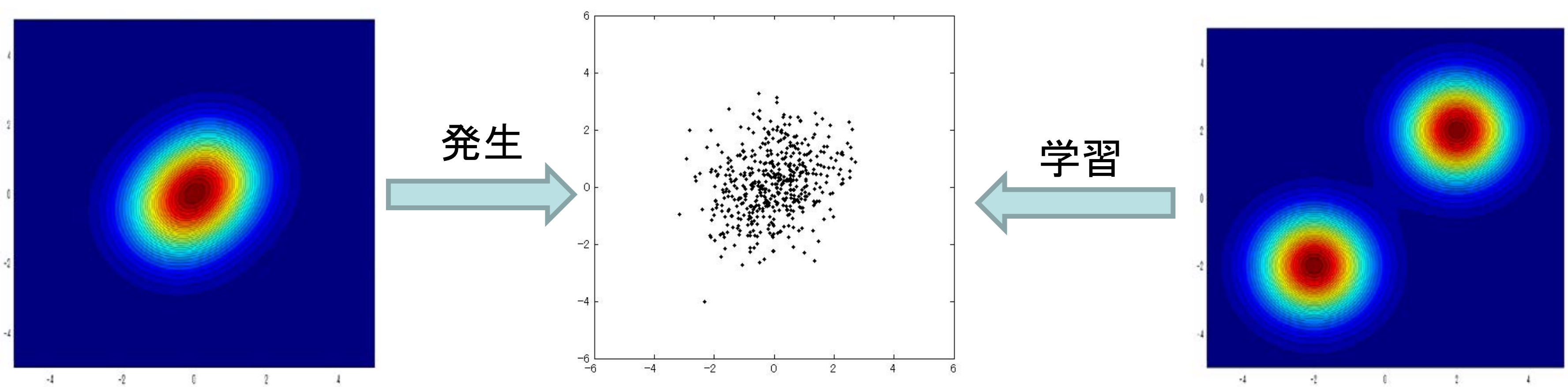
データ　　　　　　　　　　　　　　　　　　モデル

真の分布は「1山か2山か区別しにくいもの」にします。

真の分布は、「2山であるが1山と区別がつきくいもの」を想定します。
（統計的検定やモデル選択を行ないたくなるケースです）。

真の分布: データの数は n で p(x|0.5,(0.5,0.5),(-0.5,-0.5)) から独立に発生。

（注意）モデル p(x|w) と真の分布がぴったりと一致するパラメータは
a= 0.5 かつ「 b = -c = (0.5,0.5) または b = -c = (-0.5,-0.5) 」に限ります。



真の分布
(2山だが区別しにくい)

データ

モデル

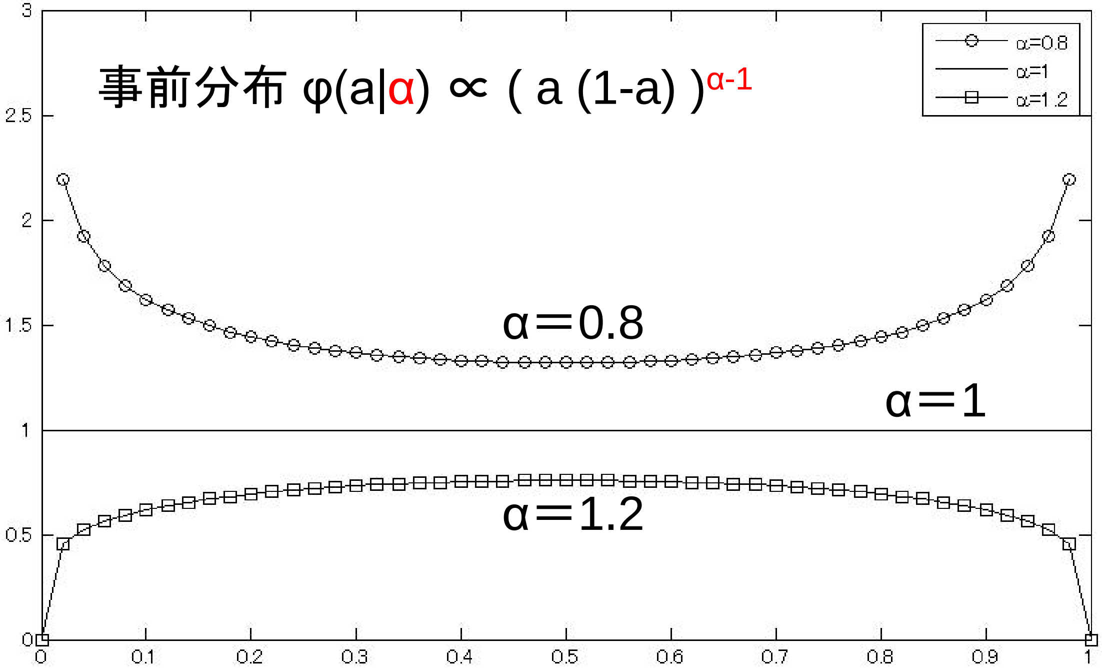
事前分布はディリクレ分布を使います。

事前分布にはディリクレ分布を使います。

事前分布 (α,hをハイパーパラメータに持つ)（αの影響を調べる. h=0.0001は固定)

事前分布 $\varphi(a|\alpha) \propto (\, a\,(1-a)\,)^{\alpha-1}$　　ディリクレ分布（α>0)

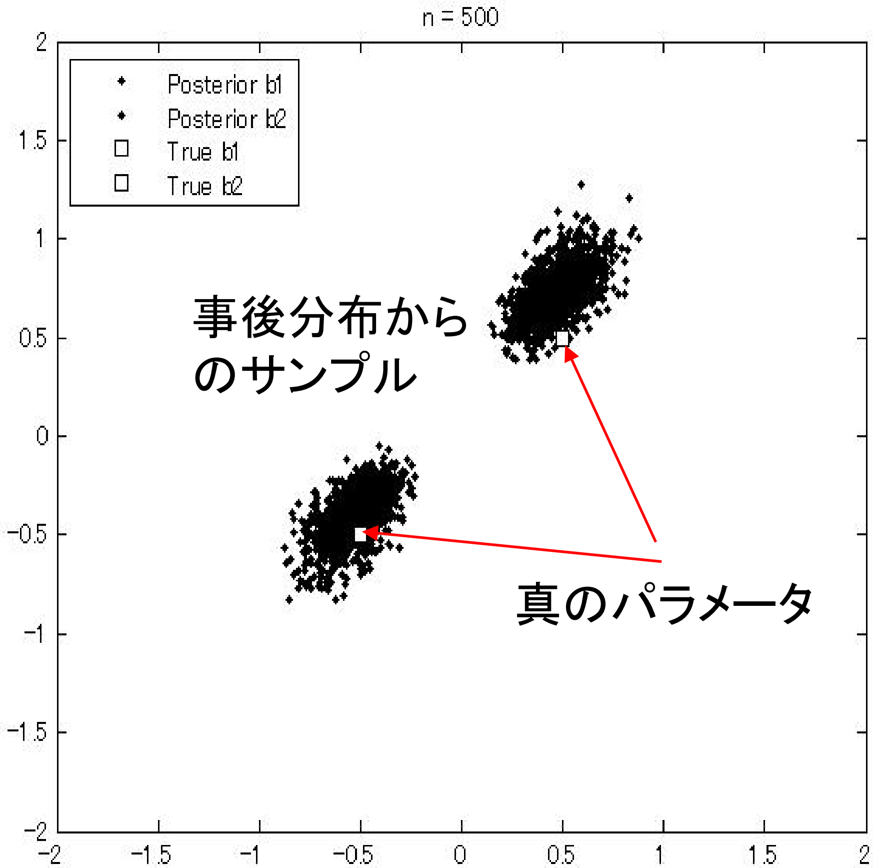
事前分布 $\varphi(b,c|h) \propto \exp\{ -(h/2)\,(||b||^2 + ||c||^2)\,\}$



（注意）α が大きいほど
事前分布の a の値は
0.5 の近くに集まります。

事後分布を目でみるために表示します。

事後分布はギブスサンプラーで生成します。事後分布が n や α の変化とともに どのように変わるかを目でみて確かめてみましょう。



事後分布は (a,b_1,b_2,c_1,c_2) で表される5次元の空間の上の確率分布です。5次元空間を描くことはできないので、以下では、次のように表示しています。

(1) a については描画しない。

(2) (b_1,b_2) と (c_1,c_2) を同じ図の上に重ねてプロットする。

・ 事後分布のサンプル点

□ 真のパラメータ（２個）

（注）a についても描画するとさらに面白いのですが、わかりにくくなるのでやめました。

ギブスサンプラーの作り方

データを $\{x_i \in \mathbf{R}^2 ; i=1,2,\ldots,n\}$ とします。

隠れ変数 $\{y_{ik} ; i=1,2,\ldots,n, k=0,1\}$ を用いてギブスサンプラーを作りましょう。

(1) 各 $\{y_{ik}\}$ の初期値を、各 i ごとに $y_{ik}=0$ または $y_{ik}=1$ と決める（任意でよい）.

(2) $n_k= \Sigma\, y_{ik}$ (k=0,1) とする. ただし（ Σ は i=1,2,…,n の和）。

(3) $B_0=\{ \Sigma(1-y_i)x_i \} /(h+n_0)$, $B_1= \{ \Sigma y_i x_i \} /(h+n_1)$ を算出する。

(4) $s_k=1/(h+n_k)$ (k=0,1) とする。

(5) ディリクレ分布 $\propto (a_0)^{\alpha-1+n_0} (a_1)^{\alpha-1+n_1}$ から(a_0,a_1)を発生。

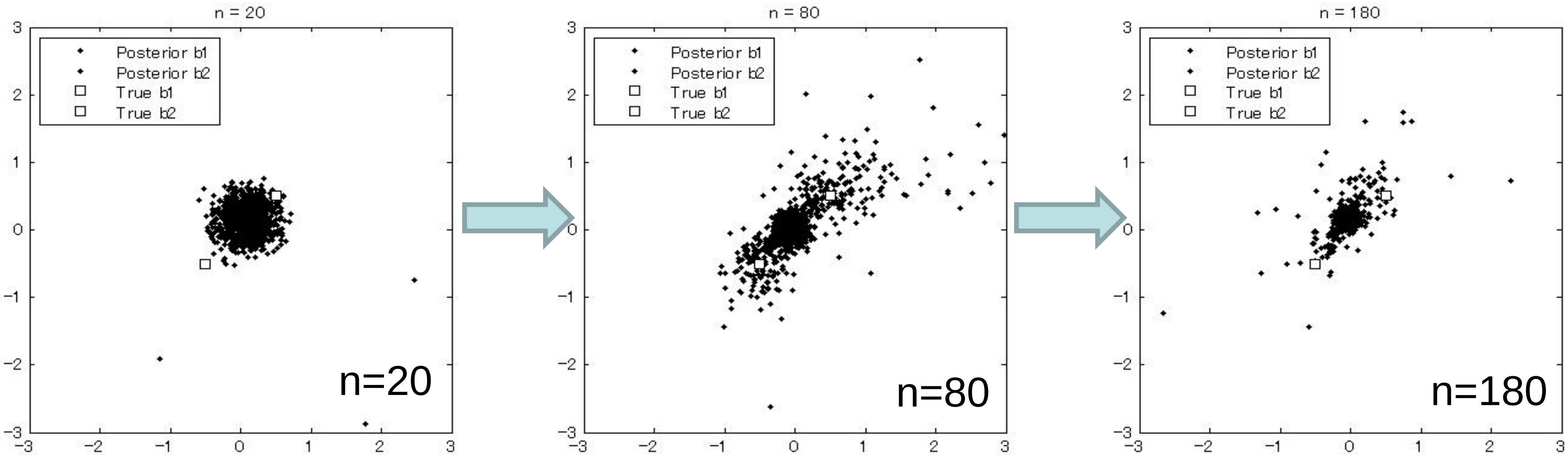
(6)「平均 B_k 共分散 $s_k I$ の正規分布」から b_k を発生 (k=0,1. I は2×2の単位行列)

(7) $L_k= a_k \exp(-(1/2)||x_i-b_k ||^2)$ (k=0,1) を用いて確率 $L_k/(L_0+L_1)$ で $y_{ik}=1$ とする。

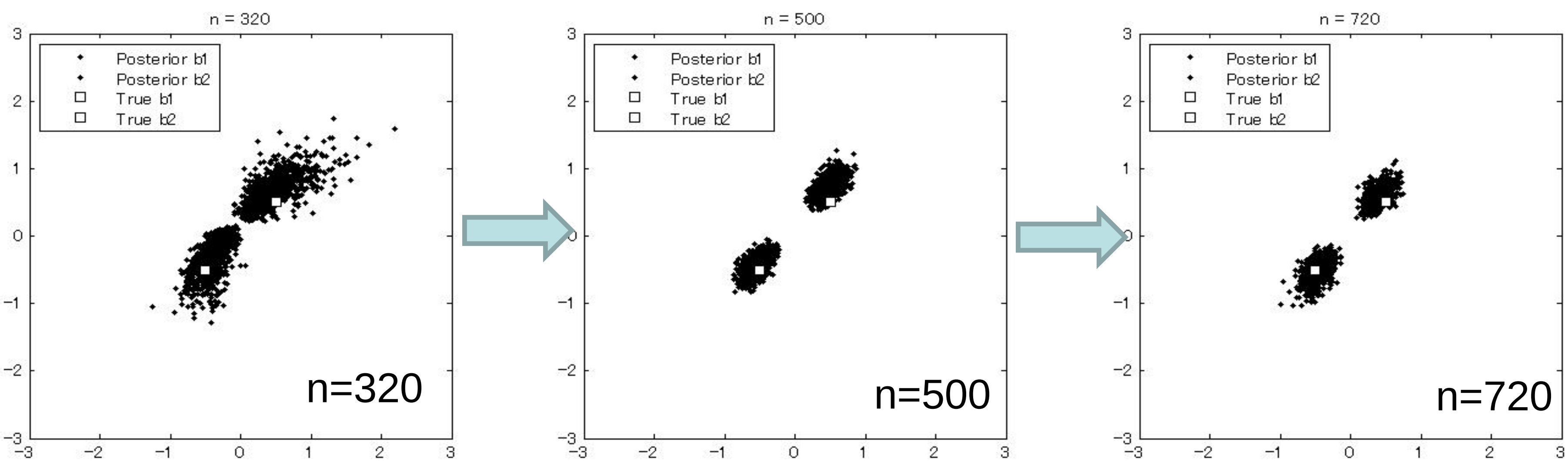
(8) (2)に戻って繰り返す。

初期値の影響が無くなってから $\{(a_1,b_0,b_1)\}$ を取り出して (a,b,c) の事後分布からのサンプリングとする。

事後分布のデータの数 n による相転移の様子 (α=1)

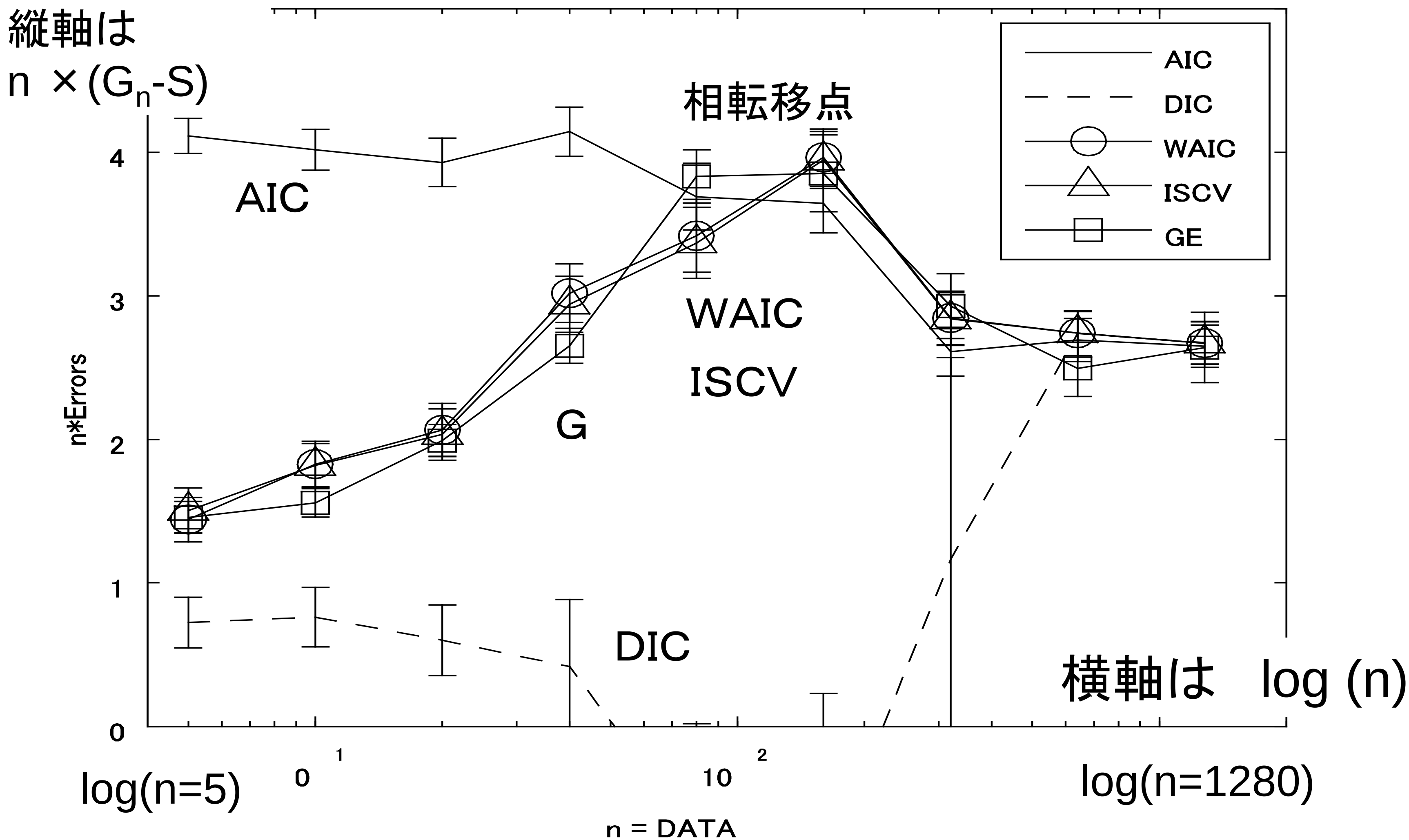


データの数 n が増えると事後分布が1山から2山に分かれます(構造の発見)。



汎化誤差の n に対する変化の様子

相転移が汎化誤差に与える影響を数値実験で調べてみましょう。
(n ×汎化誤差)　がデータの数 n に対してどう変化するかをプロットしました。相転移点上では汎化誤差が大きくなっています。ISCVとWAICは相転移点上でも汎化誤差を推定できます。



データ数 n により事後分布がどのように変化するか

わかっていることを述べます。

(0) 真の分布は2個の正規分布の混合ですが区別がつきにくい場合を考えています。

(1) データが少ないうちは「１個の正規分布から出たデータに対する学習」が
行なわれています。これをひとつの「相」と呼びます。

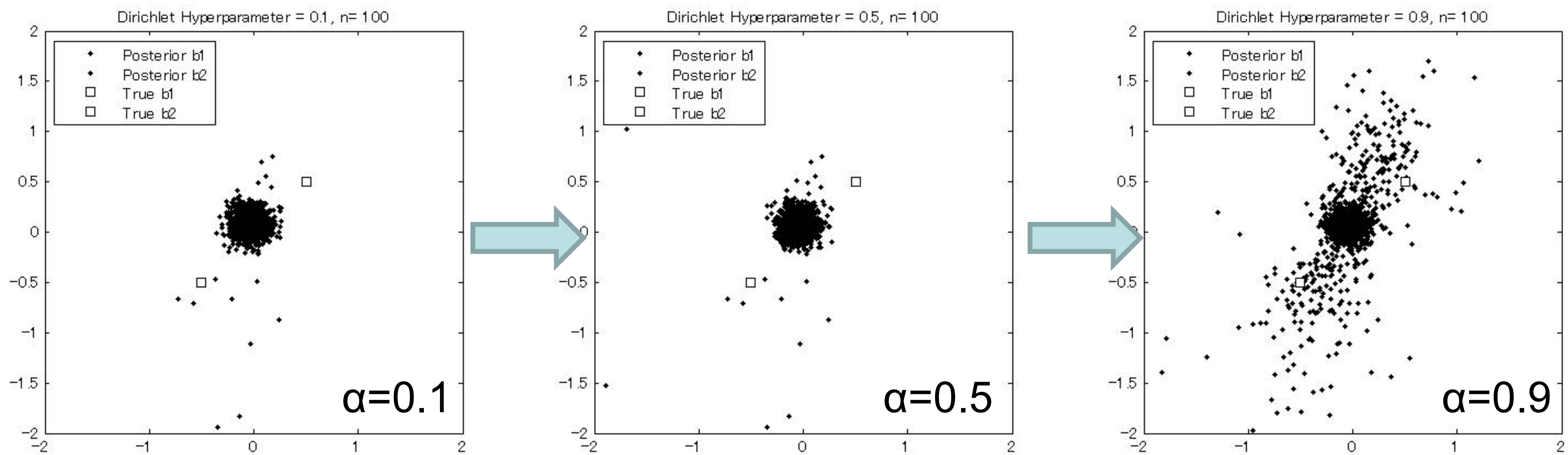
(2) データが多くなると「２個の正規分布から出たデータに対する学習」に転移
します。これもひとつの「相」と呼びます。

(3) 相と相の間に不安定な状態が存在します（相転移点）。

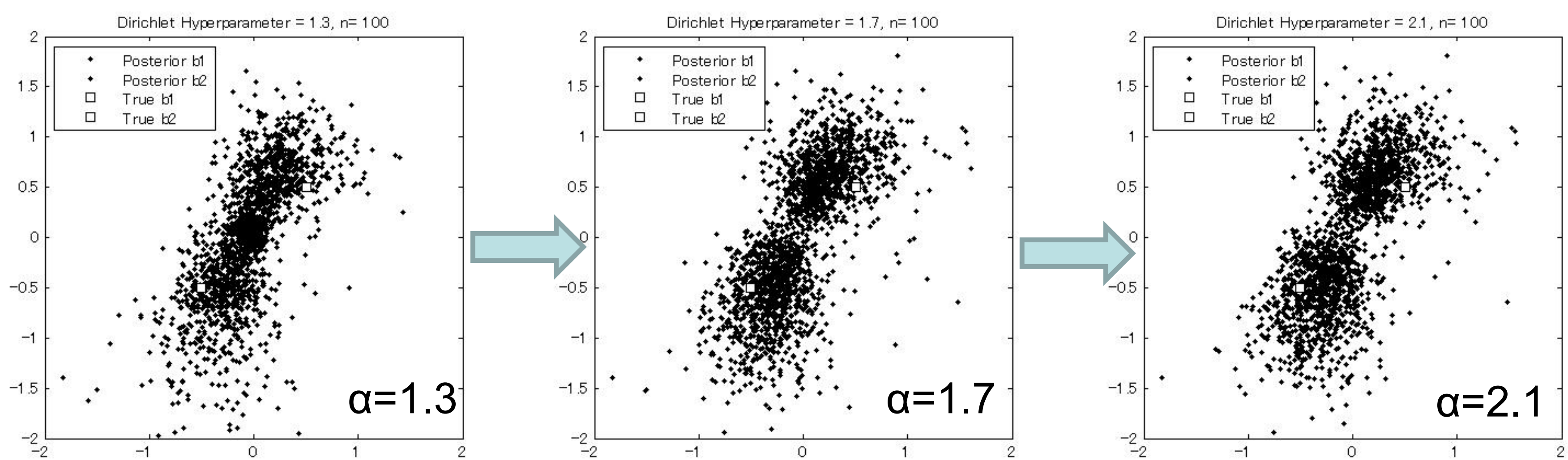
(4) どの n で相転移が起こるかは真の分布によって異なります。また真の分布が
同じでもデータの出方によるばらつきにも影響を受けます。

(5) どちらの相でも相転移点上でも CV, WAIC で汎化損失の推測ができます。

事後分布のハイパーパラメータ α による相転移 (n=100)



同じデータに対しても α を変えると事後分布が1山から2山に分かれます。



ハイパーパラメータ α により事後分布がどのように変化するか

前ページの図からわかることを述べます。

(0) 真の分布は2個の正規分布の混合ですが区別がつきにくい場合を考えています。

(1) α が小さいうちは「１個の正規分布から出たデータに対するような学習」が
行なわれています。これをひとつの「相」と呼びます・

(2) α が大きくなると「２個の積分布から出たデータに対するような学習」に転移
します。これもひとつの「相」と呼びます。

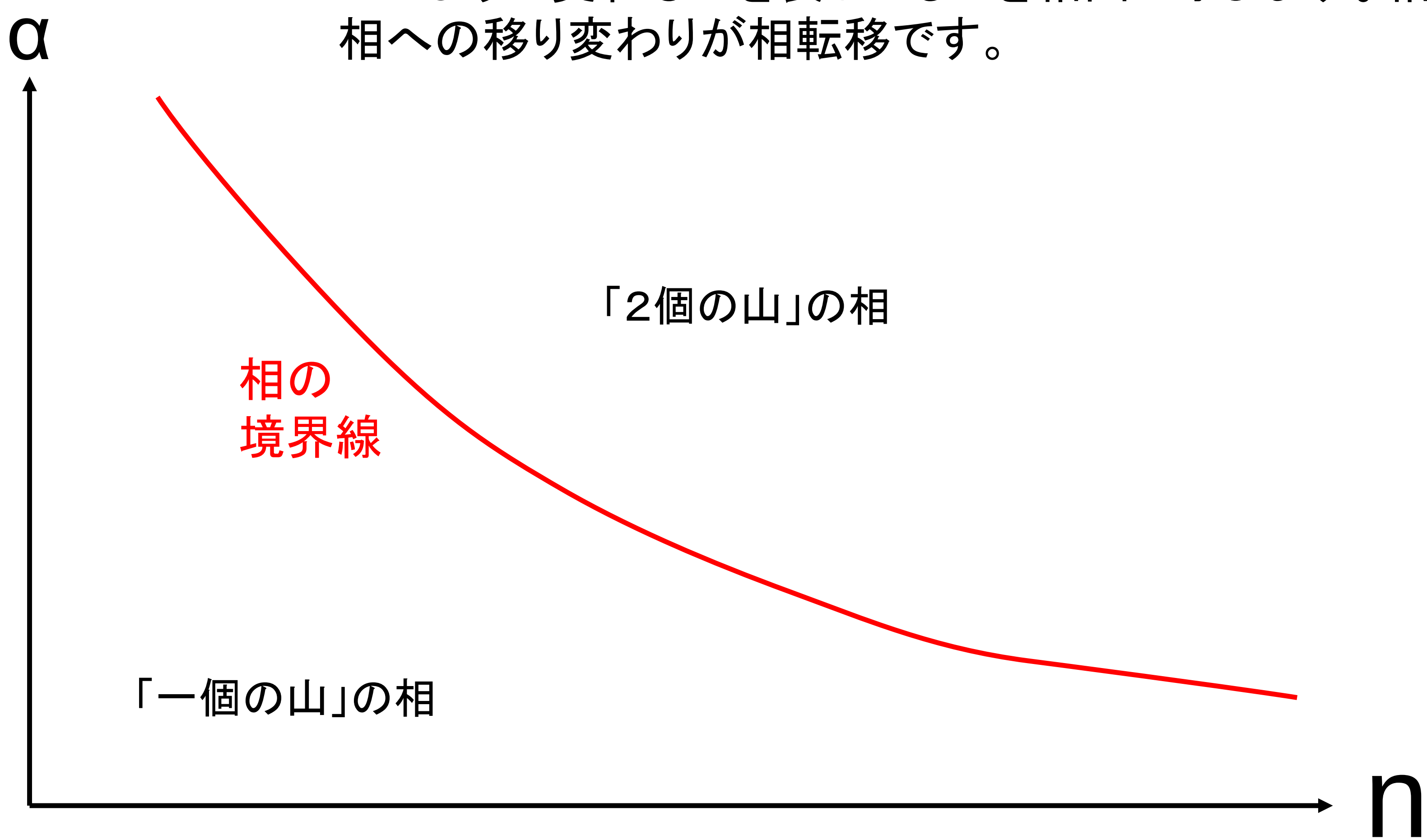
(3) 相と相の間に不安定な状態が存在します（相転移点）。

(4) どの α で相転移が起こるかは真の分布によって異なります。また真の分布が
同じでもデータの出方によるばらつきにも影響を受けます。

(5) どちらの相でも相転移点上でも CV, WAIC で汎化損失の推測ができます。

相図とは何ですか

二つの制御変数 (n, α) の値によって事後分布の様子がどのように変わるかを表したものを相図と呼びます。相から相への移り変わりが相転移です。



（注) 一般には相の個数は2個とは限りません。多数あることのほうが普通です。

実問題は相転移の近くにあることが多いです。

混合正規分布や神経回路網のように階層構造を持つモデルについて考えます。

(1) 現実の問題では、真の分布も学習モデルも2個の正規分布からできているわけではありません。

(2) 現実の問題では真の分布は不明です。モデリングされた確率分布が真の分布に対して複雑すぎるのか単純すぎるのかを考えるのがデータ解析者の仕事になります。

(3) 学習モデルがぴったりと真の分布に等しいことはほとんどありえませんが、一方で学習モデルがまったく真の分布から離れていることもありません。

(4) 混合正規分布や神経回路網のような構造を持つ統計モデルは、上記のように、真の分布に対して学習モデルが適切であるかないかを考察するという状況では相転移の近傍の状態になっています。（モデル選択・統計的検定・ハイパーパラメータの最適化では相転移の近くの状態が考察されます）。

(4) このため相転移構造の理解は、真の分布が不明であるという条件下で統計モデリングを行なう人にとって重要な基盤になります。「もしかしたら、あなたの事後分布は相転移点上にあるのでは？」。

解明されていること

(1) 渡辺一帆さんが、混合正規分布の学習に変分ベイズ法を適用する時の相転移の構造を解明されています(2006)。
(2) 中島伸一さんが、行列分解における相転移の構造を解明し汎化誤差の挙動を導出されています（2007)。
(3) 梶大介さんが、混合ベルヌーイ分布の学習に変分ベイズ法を適用するときの相転移の構造を解明されています(2011)。
(4) 中村文士さんが、混合正規分布の事前分布のハイパーパラメータを複数個にした場合の相転移構造を解明されています(2014)。
(5) 幸島匡宏さんが、非負値行列分解の相転移の構造を解明されています(2017)。

課題

統計モデルの相転移の構造を調べると統計モデリングやMCMC法の設計などについて極めて有用な情報が得られます。この理論の現段階での弱点は、「相転移という概念は必ずしもわかりやすいものではないため、その知見を統計モデリングに活用できる人材が極端に少ない」という点にあります。

「事後分布の相転移」を正しく理解して現実の事後分布の挙動を理解することを実務で使いこなせるようになると、統計モデラーとして、新しい展望を持つことができると思います。

付録：物理学における相転移

ここで参考までに物理学において現れる相転移について説明しますが、「物理学を学ばなければ統計モデリングを行なうことができない」というわけではありません。関心がない人は、このページを読む必要はありません。物理学は自然の美しさを理解するための【至高の学問】であって何かの役に立てるためのものではないことを十分にご理解の上でお読みください。

(1) 物理的な変数 $x \in \mathbf{R}^{\infty}$ に対してエネルギー関数 $H(x)$ が定義されているとき、逆温度 βの平衡状態を表す確率分布は $p(x) \propto \exp(-\beta\, H(x))$ で定義されます。考察しているシステムに外から磁場などの影響があるときには、磁場の強さを α で表すとエネルギー関数は $H(x)= H_0(x)+\alpha H_1(x)$ のような形で与えられることになります。そこで平衡状態 $p(x)$ が (β, α) に対してどのように変化するかを考えることになります。

(2) 具体的には、平衡状態の物理学的な性質を現す量（比熱や磁化など）を求めたくなります。自由エネルギー $F(\beta,\alpha)=-(1/\beta) \log \int \exp(-\beta H(x))dx$ を計算することができれば、物理的に知りたい量はすべて自由エネルギーから求めることができることが知られていますので、理論的な目標は自由エネルギーを計算することになります。

(3) 自由エネルギーが計算できると、ある点を境界として平衡状態の様子が大きく異なることがわかります。その点が相転移点です。「温度0で水が氷になる」、「磁石を熱してからさますと鉄になる」などが自然の世界の具体的な例です。

(4) 自由エネルギーが計算できれば平衡状態が解明できるわけですが、相転移を持つ系では自由エネルギーは簡単には計算できないことが多く、むしろ計算するには平衡状態についての深い洞察が必要になるため、自然現象の解明には数学的解析と物理学的洞察とが同時に必要になります。

付録:数学における相転移

ここで参考までに数学において現れる相転移について説明しますが、「数学を学ばなければ統計モデリングを行なうことができない」というわけではありません。関心がない人は、このページを読む必要はありません。数学は数学的自然の美しさを理解するための【究極の学問】であって何かの役に立てるためのものではないことを十分にご理解の上でお読みください。

(1) ノルムが定義された(非可換)環であって、ノルムから定まる位相について完備なものをC*環と呼びます。これは物理量によって定義される演算子全体の集合を数学的に定義したものと考えることができます。C*環から複素数全体の集合 C への線形連続関数を「状態」と呼びます。C*環上に状態が与えられると、あるヒルベルト空間が存在して、C*環をそのヒルベルト空間上の連続線形作用素の集合として表現することができます(GNS再構成)。

(2) 「KMS条件(物理学における平衡状態を数学的に特徴づけたもの)」を満たす状態のことをKMS状態と呼びます。KMS状態は逆温度や外場に相当する制御変数を持たせることができます。KMS状態全体の集合は制御変数に応じて変化しますが、KMS状態の集合の要素が1個から複数へと変化する点が数学的な「相転移点」の定義になります。このように特徴づけることで相転移の問題を数学として考察することが可能になります。相転移が存在することやKMS状態の作る集合がどのようになっているかなどを解明することができます。

(3) 物理学の問題を数学的に厳密に理解しようとすると、かなりしばしば、自然科学の課題は数学として最も高度で美しい問題に対応しているという経験的事実があります。例えばスピンシステムが作る代数はⅢ型フォンノイマン環という一番構造が難しいものであることが知られています(荒木-Woods)。

付録の付録：付録についての注意

数学や物理学において相転移を考察する場合には、変数の数は最初から無限であるかあるいは有限で計算しておいて無限になる極限を考えるかのいずれかであり、厳密な意味での相転移は無限次元の世界において初めて出現すると考えられています。

一方、統計的モデリングを行なう場合には、すべての変数が大きくはあっても有限であり、このためここで説明している「相転移」は、無限で初めて出現する意味での相転移ではありません。

しかしながら、確率的な現象では、しばしば、有限の世界を見る際にも「まず無限遠の点に立ってそこから有限の世界を見渡す」という考えかたが有効であることが知られています。大数の法則、中心極限定理、ウィグナーの半円則など、その例はとてもたくさんあります。

ここで説明している「相転移」もそのような意味での相転移であり、実験で示した「山が１個の相」と「山が２個の相」も、くっきりと分かれているものではなく、相転移点は少し「モヤッ」としています。

数学や物理学では相転移点上の平衡状態はそうでない点とは異なる挙動を持つ（多くの場合広い裾を持ち不安定になる）ことが知られていますが、統計モデリングにおいても同様です。
事後分布を数値的に作ることが難しくなります。統計的推測を行なう場合には、相転移点は、
避けるほうがよいのではないかと思われます。相転移が存在することを積極的に活用して新しい統計的アルゴリズムを作ることができるでしょうか。