

# ベイズ学習における必要最低サンプル数の推定

Estimating a minimum required sample size for Bayesian learning

徳田悟<sup>\*1</sup> 永田賢二<sup>\*1</sup> 岡田真人<sup>\*1\*2</sup>  
Satoru Tokuda Kenji Nagata Masato Okada

<sup>\*1</sup>東京大学 大学院新領域創成科学研究科  
Graduate School of Frontier Sciences, The University of Tokyo

<sup>\*2</sup>理化学研究所 脳科学総合研究センター  
RIKEN Brain Science Institute

A probability density gradually becomes identified as the sample size increases. To the best of our knowledge, there is a fundamental question that has not been elucidated: at least how many samples are needed to find out the density structure? In this study, we show that a minimum required sample size is related to the transition points just like the temperature that the ice melts into water. This is based on the fact that Bayesian learning is mathematically equivalent to statistical mechanics, which explains phase transition phenomena. We calculate an order parameter which we call Bayesian specific heat, and show a drastic change in the density structure around the transition points to validate our theory. We make some experiments in radial basis function networks as an example, by using the exchange Monte Carlo method since it is difficult to obtain Bayesian specific heat analytically.

## 1. 序論

近年の計算論的神経科学での研究成果から、我々は事前知識と入手した情報を統合するベイズ学習を行うことで、外界を認識していることが示唆されている [Doya 07]. しかし、どれ程の情報を得れば、正しい推論を行う、すなわち「気付く」ことになるのかについて、我々の知る限り、明瞭な枠組みでの議論がなされていない。ここで、氷が水に融解するというような相転移現象を扱う統計力学とベイズ学習の数理的対応に注目したい。ベイズ学習におけるサンプル数は、統計力学における逆温度に関係する。サンプル数が無限大であると仮定する漸近理論が盛んに研究されてきた [渡辺 12]. 一方、有限のサンプル数での議論は、解析の困難性からか、十分に行われていない。有限のサンプル数での定式化を深めることで、統計力学が扱う相転移現象として、前述の課題を議論できることを我々は着想した。本研究では、上記の着想を裏付ける、ベイズ比熱という新たな統計量を導入する。有限のサンプル数における解析の困難性を回避するため、交換モンテカルロ法 [Hukushima 96] を用いた数値計算を行い、ベイズ比熱の有効性を示す。

## 2. 理論

正規過程  $N(0, \gamma^{-1})$  に従うデータセット  $D = \{x_i, y_i\}_{i=1}^n$  をパラメータセット  $w$  で定義される関数  $g(x; w)$  によって回帰するとき、各データ点  $\{x, y\}$  の尤度関数は次のように書ける。

$$p(y | x; w) = \sqrt{\frac{\gamma}{2\pi}} \exp\left(-\frac{\gamma}{2}(y - g(x; w))^2\right) \quad (1)$$

本研究では、次の動径基底関数 (RBF) ネットワークを扱う。

$$g(x; w) = \sum_{k=1}^K a_k \exp\left(-\frac{b_k}{2}(x - \mu_k)^2\right) \quad (2)$$

ここで、 $w = \{a_k, b_k, \mu_k\}_{k=1}^K$  である。次のように逆温度  $\beta$  の事後分布 [渡辺 12] を定義する。

$$p(w | D, \beta) = \frac{1}{Z_n(\beta)} \prod_{i=1}^n p(y_i | x_i; w)^\beta p(w) \quad (3)$$

ここで、 $p(w)$  はパラメータセット  $w$  の事前分布、 $Z_n(\beta)$  は規格化定数を表す。通常は  $\beta = 1$  として扱うが、解析の都合上、一般化して議論する。 $Z_n(1)$  が周辺尤度に対応する。次のようにベイズ自由エネルギー [渡辺 12] を定義する。

$$F_n(\beta) = -\frac{1}{\beta} \log Z_n(\beta) \quad (4)$$

$$= nL_n(w_0) + \frac{\lambda}{\beta} \log n + \frac{1}{\beta} O_p(\log \log n) \quad (5)$$

ここで、 $L_n(w) = -(\sum_{i=1}^n \log p(y_i | x_i; w))/n$  であり、 $w_0$  は真の分布と事後分布のカルバック距離を最小化するパラメータセット、 $\lambda$  は双有理不変量 (RLCT) と呼ばれる正の有理数を表す。本研究では、次の関数  $C_n(\beta)$  を新たに定義する。

$$C_n(\beta) = \frac{\partial(nL_n(w))_\beta}{\partial\beta^{-1}} = \lambda + O_p\left(\frac{1}{\sqrt{\log n}}\right) \quad (6)$$

ここで、 $\langle \cdot \rangle_\beta$  は逆温度  $\beta$  の事後分布での期待値を表す。漸近挙動は Watanabe が示した系 [Watanabe 13] に基づいて導かれる。統計力学との対応から、 $L_n(w)$  が系のハミルトニアン、 $\langle nL_n(w) \rangle_\beta$  が内部エネルギーに当たるため、以降、 $C_n(\beta)$  をベイズ比熱と呼ぶ。

## 3. 実験

ベイズ自由エネルギーやベイズ比熱を解析的に得ることは、一般に困難である。本研究では、交換モンテカルロ法により、同時分布  $p(\{w_l\} | D, \{\beta_l\}, \theta) = \prod_l p(w_l | D, \beta_l, \theta)$  から  $\{w_l\}$  をサンプリングすることで、上記の計算を数値的に実行した。ここで、 $\beta_l$  および  $w_l$  はそれぞれ各レプリカでの逆温度とパラメータセットを表す。真のユニット数  $K_0 = 2$  で定義される真の分布  $p(y | x; w_0)$  から、 $n = 101$ 、 $\gamma = 10^6$  の人工データを 100 セット生成し、学習データとして用いた。

結果を図 1 に示す。図 1(A) に示すユニット数  $K = 2$  のモデルで学習した結果を見ると、点 (b) および点 (e) 付近においてベイズ比熱  $C_n(\beta)$  が急激に変化している。図 1(a - f) を見ると、(a - b) 間、(d - e) 間でそれぞれ、量的には密度のスケールが、質的には分布の幾何構造が、劇的な変化を示し

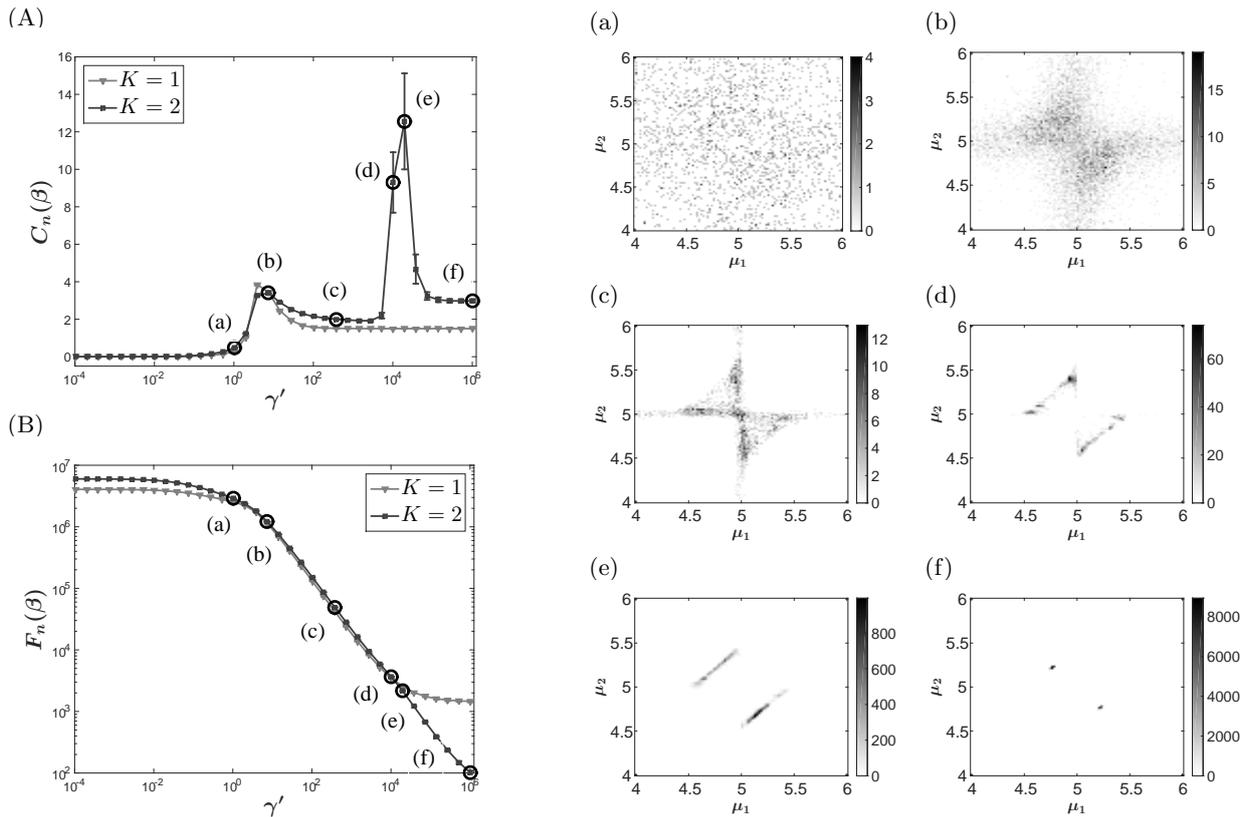


図 1: (A) および (B) は各秩序変数の変化を示している. 横軸は実効的なデータ量  $\gamma' = n\beta\gamma$  を対数表示している.  $K = 1$ ,  $K = 2$  での各値をそれぞれ正方形, 下向き三角形で表している. エラーバーは学習データについてのばらつきを標準偏差の二倍で表している. (A) 縦軸はベイズ比熱  $C_n(\beta)$  を表している. (B) 縦軸はベイズ自由エネルギー  $F_n(\beta)$  を対数表示している. (a - f) 図 1(A) および (B) 中の黒丸で囲った各点における事後確率密度を示している.  $\mu_1\mu_2$  平面でのサンプリング数をヒートマップで表している.

ている. 統計力学との対応から, これらの変化は相転移として解釈できる. 図 1(A) において, ベイズ比熱  $C_n(\beta)$  がそれぞれ一定値を示していることから, (a), (c), (f) は物質の三態に対応するような異なる相として解釈できる. これは式 (6) より, ベイズ比熱の漸近値が RLCT であり, RLCT が真のユニット数と学習モデルのユニット数によって一意に定まること [Tokuda 13] に依拠している. 実際に図 1 を見ると分布構造から, (a) はデータ量が全く足りず推論が不定な状況, (c) はデータ量が不十分なため, 本来ユニット数が  $K_0 = 2$  であるところを  $K_0 = 1$  であると推論している状況, (f) はデータ量が十分であり,  $K_0 = 2$  であると推論している状況, として特徴付けられる. 特に, (c) の分布構造はユニット数  $K_0 = 1$  の構造を  $K = 2$  という冗長なユニット数で表現することで生じる学習モデルの特異性が反映されている. 従って, (c) が属する相や (d - e) 間における相転移の存在は, 特異モデルを含むモデルセット特有の現象であると言える. 実際に図 1(A) を見ると, 正則モデルであるユニット数  $K = 1$  での学習結果は, 事前分布に由来した点 (b) 付近の相転移しか示していない. 図 1(B) を見ると, ユニット数  $K = 1$  と  $K = 2$  の学習モデルそれぞれのベイズ自由エネルギー  $F_n(\beta)$  について, 点 (e) 付近でその大小が逆転している. 式 (5) より, これは図 1(A) におけるベイズ比熱の挙動と対応していることがわかる. 従って今回の場合, 本来ユニット数が  $K_0 = 2$  の構造をユニット数  $K = 2$  であると適切にモデル選択する上で, 最低限必要となる実効的なデータ量は転移点に対応する  $\gamma' = 10^4$  程度であると言える.

#### 4. 結論

我々はベイズ学習の枠組みにおいて, ベイズ比熱という統計量を新たに定義した. RBF ネットワークにおいてベイズ比熱の性質を検証する実験を行い, データ量の増減による学習の質的变化を捉える上で, ベイズ比熱が有効であることを示した. 統計力学との対応からその変化は相転移と解釈でき, 事前分布に由来するものとモデルの特異性に起因するものに大別できることを示した. ベイズ比熱が示す転移点が適切な学習に最低限必要となる実効的なデータ量, すなわちサンプルの数や精度に対応し, それらが推定できることを示した.

#### 参考文献

[Doya 07] Doya, K. (ed.): Bayesian brain: Probabilistic approaches to neural coding, MIT press (2007).  
 [渡辺 12] 渡辺澄夫:ベイズ統計の理論と方法, コロナ社 (2012).  
 [Hukushima 96] Hukushima, K. and Nemoto K., *J. Phys. Soc. Jpn.*, Vol.65, No.6, 1604-1608 (1996).  
 [Watanabe 13] Watanabe, S., *J. Mach. Learn. Res.*, Vol.14, No.1, 867-897 (2013).  
 [Tokuda 13] Tokuda, S., Nagata, K. and Okada, M., *IPSJ Transactions on Mathematical Modeling and its Applications*, Vol.6, No.3, 117-123 (2013).