

一般化最小二乗法による合成空間モンテカルロ積分法

関本快士[†]山形大学大学院理工学研究科[†]安田宗樹[‡]山形大学大学院理工学研究科[‡]

1. はじめに

ボルツマンマシン (Boltzmann Machine (BM)) は機械学習をはじめとする様々な分野において重要な役割を担う確率モデルである [1]. BM 上の統計量計算は学習や推論において不可欠であるが, これは組合せ爆発の問題を含んでいるため, 実用的には近似的評価が必要となる. 空間モンテカルロ積分法 (spatial Monte Carlo integration (SMCI)) はモンテカルロ積分法 (Monte Carlo integration (MCI)) を空間的に拡張した手法であり, 高精度に統計量を近似できる手法として知られている [2]. SMCI では, 空間の取り方を任意に選べるため, 同一の統計量を評価する複数の推定量が存在し得る. 本研究では, 一般化最小二乗法 (Generalized Least Squares (GLS)) に基づき, 複数の SMCI の推定量を合成して新たな推定量を得る手法を提案する [3]. GLS は, SMCI の推定量が満たすある仮定の下で分散最小の不偏推定量を与えるため, これを組み込むことで, 更なる高精度近似を達成できる [3].

2. ボルツマンマシン

BM は, 無向グラフ $G(\mathcal{V}, \mathcal{E})$ 上に確率分布を定義するマルコフ確率場の特殊形である [1]. ここで, $\mathcal{V} := \{1, 2, \dots, n\}$ はノードの集合であり, \mathcal{E} は頂点 i, j を繋ぐリンク (i, j) の集合である. ノードに対応させた確率変数 $\mathbf{x} := \{x_i \in \{-1, +1\} | i \in \mathcal{V}\}$ を用いて, BM は次の確率分布で定義される.

$$P(\mathbf{x}; \theta) := \frac{1}{Z(\theta)} \exp \left(\sum_{i \in \mathcal{V}} h_i x_i + \sum_{(i,j) \in \mathcal{E}} J_{i,j} x_i x_j \right) \quad (1)$$

ここで, $Z(\theta)$ は規格化定数であり, パラメータ $h_i, J_{i,j}$ をまとめて θ と表記する.

確率変数 $\mathbf{x}_{\mathcal{T}} := \{x_i | i \in \mathcal{T} \subseteq \mathcal{V}\}$ を対象としたある関数 $f(\mathbf{x}_{\mathcal{T}})$ の BM 上での期待値は次の形で表される.

$$\mathbb{E}_{\mathcal{P}}[f(\mathbf{x}_{\mathcal{T}})] := \sum_{\mathbf{x}} f(\mathbf{x}_{\mathcal{T}}) P(\mathbf{x}; \theta) \quad (2)$$

上式に含まれる多重和 $\sum_{\mathbf{x}}$ の計算量は n の増加に伴って指数的に増加するため, 期待値は近似的に計算する必要がある.

3. 空間モンテカルロ積分法

SMCI は標本平均で期待値を近似する MCI を空間的に拡張して部分的に和をとる手法であり, MCI より平均的に高い近似性能をもつ [2]. SMCI では, \mathcal{T} を含む部分集合 \mathcal{U} と, \mathcal{U} の最近接集合 $\partial\mathcal{U}$ を考慮して近似する. $\mathcal{T}, \mathcal{U}, \partial\mathcal{U}$ は, それぞ

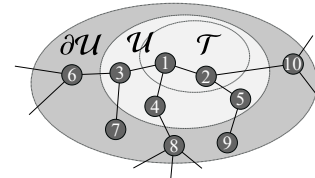


図1 SMCI で考慮される各領域の概念図. 目標領域 $\mathcal{T} = \{1, 2\}$ に対して和領域を $\mathcal{U} = \{1, 2, 3, 4, 5\}$ と選ぶと, サンプル領域は $\partial\mathcal{U} = \{6, 7, 8, 9, 10\}$ となる.

れ目標領域, 和領域, サンプル領域と呼ばれる. 図1に各領域の関係を示す.

$P(\mathbf{x}; \theta)$ から独立同分布で生成された N 個のサンプル点の集合 $\mathbb{S} := \{\mathbf{s}^{(\mu)} | \mu = 1, 2, \dots, N\}$ を用いて, 式 (2) の期待値の推定量は次の形で表される.

$$m_{\mathcal{T}}(\mathcal{U}; \mathbb{S}) := \frac{1}{N} \sum_{\mu=1}^N \sum_{\mathbf{x}_{\mathcal{U}}} f(\mathbf{x}_{\mathcal{T}}) P(\mathbf{x}_{\mathcal{U}} | \mathbf{s}_{\partial\mathcal{U}}^{(\mu)}) \quad (3)$$

ここで, $\mathbf{s}_{\partial\mathcal{U}}^{(\mu)}$ はサンプル領域の確率変数 $\mathbf{x}_{\partial\mathcal{U}}$ に対応するサンプル点である. 領域 $\mathcal{T} \subseteq \mathcal{U}_1 \subseteq \mathcal{U}_2$ において, $m_{\mathcal{T}}(\mathcal{U}_2; \mathbb{S})$ は $m_{\mathcal{T}}(\mathcal{U}_1; \mathbb{S})$ より統計的に高精度である [2].

SMCI では, \mathcal{U} の取り方は任意のため, 同一の統計量を評価する複数の推定量が存在し得る. 次節で述べる GLS の枠組みを適用すると, 複数の SMCI 推定量を合成的に扱い, より高精度な推定量を得ることができる.

4. 一般化最小二乗法

GLS は, 誤差の間に相関がある線形回帰モデルに対して, 通常の最小二乗法よりも有効的に未知の回帰係数を推定する手法である [3]. GLS では, 次の線形回帰モデルを考える.

$$\mathbf{y} = \Xi\boldsymbol{\beta} + \boldsymbol{\varepsilon}, \mathbb{E}[\boldsymbol{\varepsilon}] = \mathbf{0}, \text{Var}[\boldsymbol{\varepsilon}] = \boldsymbol{\Omega} \quad (4)$$

ここで, $\mathbf{y} \in \mathbb{R}^M$ は目的変数ベクトルで, $\Xi \in \mathbb{R}^{M \times D}$ は説明変数行列, $\boldsymbol{\beta} \in \mathbb{R}^D$ は未知の回帰係数ベクトル, $\boldsymbol{\varepsilon} \in \mathbb{R}^M$ は回帰誤差ベクトルである. また, $\boldsymbol{\Omega} \in \mathbb{R}^{M \times M}$ は既知の分散共分散行列である. $\boldsymbol{\Omega}$ は正則で, Ξ は列フルランクであると仮定すると, GLS の推定量 (GLS estimator (GLSE)) は次の形で得られる.

$$\boldsymbol{\beta}_{\text{GLS}} := (\Xi^T \boldsymbol{\Omega}^{-1} \Xi)^{-1} \Xi^T \boldsymbol{\Omega}^{-1} \mathbf{y} \quad (5)$$

ガウス・マルコフの定理より, GLSE は $\boldsymbol{\beta}$ の最良線形不偏推定量である [3]. 上述の仮定に加えて, $\boldsymbol{\varepsilon}$ が多次元ガウス分布に従うと仮定すると, GLSE は $\boldsymbol{\beta}$ の最良不偏推定量 (Best Unbiased Estimator (BUE)) となる [3].

5. 合成空間モンテカルロ積分法

本節では, GLS に基づき, 複数の SMCI 推定量を合成的に用いて新たな推定量を得る手法を提案する. 目標領域 \mathcal{T} を包含する M 個の和領域 $\mathcal{U}_1, \mathcal{U}_2, \dots, \mathcal{U}_M$ を考える. \mathbb{S} を用いた推定

Composite Spatial Monte Carlo Integration Based on Generalized Least Squares

[†] Kaiji Sekimoto; Graduate School of Science and Engineering, Yamagata University

[‡] Muneki Yasuda; Graduate School of Science and Engineering, Yamagata University

量のベクトル $\mathbf{m}_T := (m_T(\mathcal{U}_1; \mathbb{S}) \ m_T(\mathcal{U}_2; \mathbb{S}) \ \dots \ m_T(\mathcal{U}_M; \mathbb{S}))^T$ と真の期待値 $\mathbb{E}_p[f(\mathbf{x}_T)]$ の間には次の関係が成り立つ。

$$\mathbf{m}_T = \mathbf{1} \mathbb{E}_p[f(\mathbf{x}_T)] + \boldsymbol{\varepsilon} \quad (6)$$

ここで、 $\mathbf{1}$ は全ての要素が1の M 次元ベクトルであり、 $\boldsymbol{\varepsilon} \in \mathbb{R}^M$ は近似誤差ベクトルである。 \mathbb{S} は複雑な相互作用を持つBMから生成されるため、それをを用いた推定量は互いに独立ではない。よって、

$$\mathbb{E}[\boldsymbol{\varepsilon}] = \mathbf{0}, \text{Var}[\boldsymbol{\varepsilon}] = \boldsymbol{\Omega} \quad (7)$$

と表せるが、 $\boldsymbol{\Omega}$ は未知のため、提案法では標本分散共分散行列により近似する。式(6),(7)は式(4)の特殊形であり、このモデルに対してGLSを適用することで $\mathbb{E}_p[f(\mathbf{x}_T)]$ を推定する。つまり、提案法では $\mathbb{E}_p[f(\mathbf{x}_T)]$ を次のように近似する。

$$\mathbb{E}_p[f(\mathbf{x}_T)] \approx (\mathbf{1}^T \boldsymbol{\Omega}^{-1} \mathbf{1})^{-1} \mathbf{1}^T \boldsymbol{\Omega}^{-1} \mathbf{m}_T \quad (8)$$

中心極限定理より、 $\boldsymbol{\varepsilon}$ は多次元ガウス分布に従うため、提案法の推定量はBUEであることが期待される。また、 M 個の推定量に加えて別の推定量が得られた際、新たな推定量を加える前後の推定量の分散をそれぞれ $\mathbb{V}_{\text{before}}, \mathbb{V}_{\text{after}}$ とする。加える前後の厳密な分散共分散行列が双方とも正定値であると仮定すると、次の不等式が成り立つ。

$$\mathbb{V}_{\text{after}} \leq \mathbb{V}_{\text{before}} \quad (9)$$

この不等式から、提案法の近似精度は合成数の増加に伴って単調に向上することが期待される。

6. 数値実験

本節では5節で述べた提案法の理論的な有効性を数値実験を通して検証する。本実験ではある期待値に対して3つの和領域を考え、それらを”SMCI I/II/III”と表記する。図2に示すように、IとIIの和領域はIIIのそれを包含するが、IとIIの間に包含関係は成り立たない。提案法は”SMCI I”と”SMCI II”を合成



図2 SMCI I/II/III の和領域の包含関係。

した場合、全てを合成した場合の2タイプを考え、それぞれ”I + II”と”I + II + III”と表記する。

まず、期待値の近似精度を評価する。周期境界条件をもつ 4×5 の正方形格子にBMを定義し、パラメータ $h_i, J_{i,j}$ を区間 $[-0.1, +0.1]$ の一様分布からランダムに生成した。 \mathbb{S} は50ステップのギブスサンプリングで生成した[1]。期待値の近似精度は $\mathbb{E}_p[x_i]$ と $\mathbb{E}_p[x_i x_j]$ の平均絶対誤差 (mean absolute error (MAE)) で評価した。図3に N に対する結果を示す。この結果は、合成、及び合成数を増やすことにより近似精度が向上するという理論的な期待と一致している。

次に、応用としてBM学習に適用する。BM学習では $\mathbb{E}_p[x_i]$ と $\mathbb{E}_p[x_i x_j]$ を反復的に計算する必要があり、SMCIを用いた学習法は高い学習性能をもつことが知られている[4]。本実験では、パラメータを区間 $[-0.1, +0.1]$ の一様分布からランダムに生成した生成BMから1000個のデータ点の集合を生成し、そのデータを用いて、パラメータを0で初期化した学習BMを学

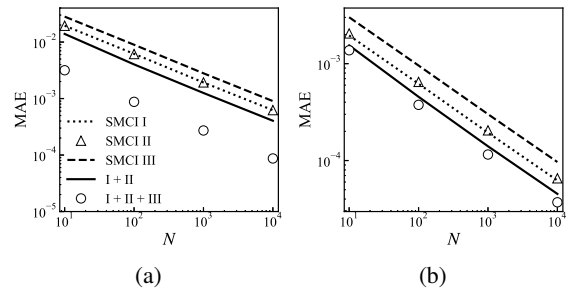


図3 N に対する期待値 (a) $\mathbb{E}_p[x_i]$, (b) $\mathbb{E}_p[x_i x_j]$ のMAE。各プロットは1000回の実験の平均値である。

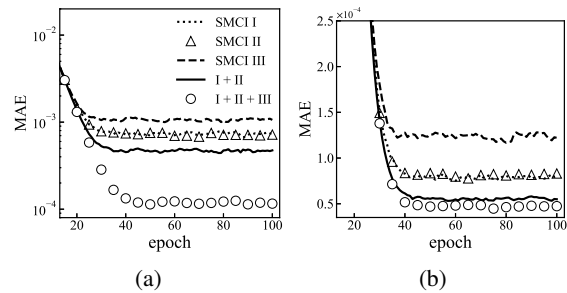


図4 エポックに対するパラメータ (a) h_i , (b) $J_{i,j}$ のMAE。各プロットは100回の実験の平均値である。

習させる。2つのBMのグラフ構造は上記の実験で用いたBMと同様である。学習で用いる \mathbb{S} は、1ステップの継続的コントラスティブ・ダイバージェンス法に基づいて逐次生成した[5]。データ点の集合は50ステップのギブスサンプリングで生成し、学習率は0.02に固定した[1]。近似学習の性能は、真の学習結果 $h_i^{\text{true}}, J_{i,j}^{\text{true}}$ とエポック e で得られた近似学習結果 $h_i^{(e)}, J_{i,j}^{(e)}$ のMAEで評価した。BM学習では $h_i^{\text{true}}, J_{i,j}^{\text{true}}$ が一意に求まるため、MAEが低くなるほど高性能であるといえる[1, 4]。実験の結果を図4に示す。図4より、合成、及び合成数を増やすことで真の学習結果により近い値に収束していることが分かる。

7. まとめ

本研究では、SMCIの推定量をGLSに基づいて合成し、新たな推定量を得る手法を提案した。合成や合成数の増加により、更なる高精度化を達成することが理論的かつ数値的に示された。また、提案法を用いたBM学習は高い近似性能をもつことが実証された。

謝辞

本研究は科研費 (18K11459, 18H03303, 21K11778) 及び JST CREST (JP-MJCR1402) の助成を受けたものである。

文献

- [1] 麻生英樹, 安田宗樹, 前田新一, 岡野原大輔, 岡谷貴之, 久保田陽太郎, ボカレラダシカ, 深層学習, 人工知能学会 (監修), 神高敏弘 (編), 近代科学社, 2015.
- [2] M. Yasuda and K. Uchizawa, A Generalization of Spatial Monte Carlo Integration, *Neural Computation*, vol.4, no.33, pp.1037-1062, 2021.
- [3] R. H. Myers, Classical and modern regression with applications, second edition, Boston, MA : PWS-KENT, 1990.
- [4] M. Yasuda, Learning algorithm of boltzmann machine based on spatial monte carlo integration method, *Algorithms*, vol.4, no.11, 42, 2018.
- [5] T. Tieleman, Training restricted boltzmann machines using approximations to the likelihood gradient, In *Proceedings of the 25th international conference on Machine Learning*, pp.1064-1071, 2008.