

Bayes factorを用いたRAIアルゴリズムによる大規模ベイジアンネットワーク学習

著者	名取 和樹, 宇都 雅輝, 植野 真臣
雑誌名	電子情報通信学会論文誌. D, 情報・システム
巻	J101-D
号	5
ページ	754-768
発行年	2018-05-01
URL	http://id.nii.ac.jp/1438/00008843/

doi: 10.14923/transinfj.2017JDP7089

Bayes factor を用いた RAI アルゴリズムによる
大規模ベイジアンネットワーク学習

名取 和樹^{†a)} 宇都 雅輝^{†b)} 植野 真臣^{†c)}

Learning Huge Bayesian Networks by RAI Algorithm Using Bayes Factor

Kazuki NATORI^{†a)}, Masaki UTO^{†b)}, and Maomi UENO^{†c)}

あらまし 漸近一致性をもつベイジアンネットワークの構造学習は NP 困難である。これまで動的計画法や A^* 探索、整数計画法による探索アルゴリズムが開発されてきたが、未だに 60 ノード程度の構造学習を限界とし、大規模構造学習の実現のためには、全く異なるアプローチの開発が急務である。一方で因果モデルの研究分野では、条件付き独立性テスト (CI テスト) と方向付けによる画期的に計算量を削減した構造学習アプローチが提案されている。このアプローチは制約ベースアプローチと呼ばれ、RAI アルゴリズムが最も高精度な最先端学習法として知られている。しかし RAI アルゴリズムは、CI テストに仮説検定法または条件付き相互情報量を用いている。前者の精度は帰無仮説が正しい確率を表す p 値とユーザが設定する有意水準に依存する。 p 値はデータ数の増加により小さい値を取り、誤って帰無仮説を棄却してしまう問題が知られている。一方で、後者の精度はしきい値の設定に強く影響する。したがって、漸近的に真の構造を学習できる保証がない。本論文では、漸近一致性を有する Bayes factor を用いた CI テストを RAI アルゴリズムに組み込む。これにより、数百ノードをもつ大規模構造学習を実現する。数種類のベンチマークネットワークを用いたシミュレーション実験により、本手法の有意性を示す。

キーワード ベイジアンネットワーク、確率的グラフィカルモデル、構造学習、条件付き独立性検定

1. ま え が き

ベイジアンネットワークは、確率変数をノードで表しノード間の依存関係を非循環有向グラフ (Directed Acyclic Graph: DAG) で表現する確率的グラフィカルモデルである。ベイジアンネットワークは、確率構造に DAG を仮定することにより、同時確率分布を条件付き確率の積に分解する。ベイジアンネットワークは同時確率分布の最も仮定が少ない近似モデルとされ、これを利用した確率推論は高い予測精度をもつことから [1]、システムの故障診断や危険予測システム、医療診断システムなど様々な目的で応用されてきた [2]。

ベイジアンネットワークの構造は一般にデータから推定する必要がある。この問題をベイジアンネットワークの構造学習と呼ぶ。

ベイジアンネットワークの構造学習法としては、漸近一致性を有する学習スコアを用いて、全ての構造の候補からスコアが最も高い構造を探索する厳密解探索アプローチが従来から用いられてきた。このアプローチは、構造の探索数がノード数に対し指数的に増加する NP 困難問題 [3] である。効率的に厳密解を探索するために、動的計画法 [4]~[8]、 A^* 探索 [9]、整数計画法 [10] などの従来の人工知能アプローチによる構造学習法が提案されてきたが、未だ 60 ノード程度の構造学習が限界である。一方で、確率推論の厳密手法 [11] は現在 200 ノード程度まで実現できており、同程度のノード数のネットワーク学習は急務である。

一方、因果モデルの研究分野では、計算効率が高い構造学習を実現する手法が提案されている。この手法は制約ベースアプローチと呼ばれ、完全無向グラフに、2 ノード間の条件付き独立性検定 (Conditional

[†] 電気通信大学大学院情報理工学研究所, 調布市
Graduate School of Informatics and Engineering, The
University of Electro-Communications, 1-5-1 Chofugaoka,
Chofu-shi, 182-8585 Japan

a) E-mail: natori@ai.lab.uec.ac.jp

b) E-mail: uto@ai.lab.uec.ac.jp

c) E-mail: ueno@ai.lab.uec.ac.jp

DOI:10.14923/transinfj.2017JDP7089

Independence test: CI テスト) を適用して学習される無向グラフに対し, オリエンテーションルール [12] による辺の方向付けを行うことで DAG を学習する. 制約ベースアプローチの研究では, PC アルゴリズム [13], MMHC アルゴリズム [14], RAI アルゴリズム [15] が提案されており, RAI アルゴリズムが最も高精度であると知られている.

制約ベースアプローチの学習では, 一般に χ^2 検定, G^2 検定または条件付き相互情報量 (Conditional Mutual Information: CMI) が CI テストに用いられる. χ^2 検定, G^2 検定は 2 ノード間が独立であることを帰無仮説として, この仮説が正しい確率を表す p 値とユーザが設定する有意水準に精度依存する. p 値はデータ数の増加により小さい値を取ることがある. これにより, Gail ら [16] では, p 値が帰無仮説の棄却域である有意水準以下を取る場合が増え, 真では帰無仮説が正しいが誤って棄却してしまう第一種の過誤 (Type I error) の発生率が高まることを指摘している. したがって, これらの CI テストを用いた学習は, データ数の増加により過学習を起こすため, 漸的に真の構造学習が保証されない. 一方で, CMI を用いた CI テストは設定するしきい値に強く影響を受けるため, これを用いた学習は漸近一致性をもたない.

本論文では, 漸近一致性を有する CI テストを制約ベースアプローチに適用する. 具体的には, Bayes factor を用いた CI テストを RAI アルゴリズムに組み込む. この手法の利点として, 以下が挙げられる.

- (1) 漸近一致性をもつため, 十分なデータから真の条件付き独立構造を推定できる.
- (2) 制約ベースアルゴリズムを採用しているため, 従来の厳密解探索アプローチの手法では学習できないノード数の学習が可能である.

既に Bayesian Dirichlet equivalence uniform (BDeu) に基づく Bayes factor を用いた CI テストは Steck ら [17] により提案され, ベイジアンネットワーク学習の理論解析に用いられているが, ベイジアンネットワークの構造学習に適用されていない.

本論文では, 複数のベンチマークネットワークを用いて, 提案手法と従来手法の精度評価を行った. 結果として, 以下の有意性が示された.

- (1) 11 ノードのベンチマークネットワークの実験では, 厳密解探索アプローチの最先端手法である整数計画法を用いた構造学習法と同様にデータ数が十分に大きくなると真の構造を推定できる.

- (2) 整数計画法を用いた構造学習法では, 60 ノード以上のネットワーク構造について制限時間内に学習が終了しなかったが, 提案手法では 200 ノード以上の大規模ネットワーク構造の学習を実行できる.

2. ベイジアンネットワーク

ベイジアンネットワークは, 確率変数をノードとし, ノード間の依存関係を非循環有向グラフ (Directed Acyclic Graph: DAG) で表現し, 各ノードの条件付き確率で表現される確率的グラフィカルモデルである.

今, $\{X_1, \dots, X_N\}$ を N 個の離散確率変数集合 \mathbf{X} とし, 各変数 X_i は r_i 個の状態集合 $\{1, \dots, r_i\}$ から一つの値を取るとする. ここで, 変数 X_i が値 k を取るとき, $X_i = k$ と書く. このとき, ベイジアンネットワークの構造 g において, 各変数 X_i の親変数集合を Π_i としたときの同時確率分布 $p(X_1, \dots, X_N | g)$ は以下のように表現できる.

$$p(X_1, \dots, X_N | g) = \prod_{i=1}^N p(X_i | \Pi_i, g) \quad (1)$$

ベイジアンネットワークを統計的に扱うためには, 式 (1) のモデルを母数化する必要がある. 式 (1) では, 条件付き確率 $p(X_i | \Pi_i, g)$ がデータより推定したい対象であるのでパラメータとして扱う.

今, θ_{ijk} を, X_i の親変数集合 Π_i が j 番目のパターンを取るとき ($\Pi_i = j$ と書く) $X_i = k$ となる条件付き確率 $p(X_i = k | \Pi_i = j, g)$ を示すパラメータとする. ベイジアンネットワークの構造学習では, その推定値として, 頑健で予測精度が高い期待事後確率推定値 (Expected a Posteriori: EAP) が最も良く用いられる. 変数集合 \mathbf{X} に対する n 個のデータを $\mathbf{D} = \{D_1, \dots, D_n\}$ としたとき, EAP は事前分布にディレクレ分布を仮定すると以下となる [18].

$$\widehat{\theta}_{ijk} = \frac{\alpha_{ijk} + n_{ijk}}{\alpha_{ij} + n_{ij}} \quad (2)$$

ここで, n_{ijk} は X_i の親変数集合 Π_i が j 番目のパターンを取ったときの $X_i = k$ となる頻度を表し, n_{ij} は $n_{ij} = \sum_{k=1}^{r_i} n_{ijk}$ を表す. データ数 n は $n = \sum_{j=1}^{q_i} n_{ij}$, ($i = 1, \dots, N$) となる. ここで, q_i は X_i の親変数集合 Π_i の取りうるパターン数を表す. また, α_{ijk} はディレクレ事前分布のパラメータを表し, α_{ij} は $\alpha_{ij} = \sum_{k=1}^{r_i} \alpha_{ijk}$ を表す. ベイジアンネットワークの構造学習では, この推定値を周辺化した周

辺ゆう度スコアを最大化する構造を探索すればよい。ここで、条件付き確率パラメータ集合 $\Theta = \{\theta_{ijk}\}$, ($i = 1, \dots, N, j = 1, \dots, q_i, k = 1, \dots, r_i$) の事前分布として、以下のディレクレ分布 $p(\Theta)$ を仮定する。

$$p(\Theta) = \prod_{i=1}^N \prod_{j=1}^{q_i} \frac{\Gamma\left(\sum_{k=1}^{r_i} \alpha_{ijk}\right)}{\prod_{k=1}^{r_i} \Gamma(\alpha_{ijk})} \prod_{k=1}^{r_i} \theta_{ijk}^{\alpha_{ijk}-1} \quad (3)$$

このとき、周辺ゆう度スコアは次式で表される。

$$\begin{aligned} p(\mathbf{D} | g, \alpha) &= \int_{\Theta} p(\mathbf{D} | \Theta, g) p(\Theta) d\Theta \quad (4) \\ &= \prod_{i=1}^N \prod_{j=1}^{q_i} \frac{\Gamma(\alpha_{ij})}{\Gamma(\alpha_{ij} + n_{ij})} \prod_{k=1}^{r_i} \frac{\Gamma(\alpha_{ijk} + n_{ijk})}{\Gamma(\alpha_{ijk})} \end{aligned}$$

ここで、 $\alpha = \{\alpha_{ijk}\}$, ($i = 1, \dots, N, j = 1, \dots, q_i, k = 1, \dots, r_i$). Heckerman ら [19] は、二つのベイジアンネットワーク構造の同時確率分布が等価（マルコフ等価と呼ばれる）であるなら、それらの周辺ゆう度も同一でなければならないというゆう度等価を導入した。そして、ゆう度等価に矛盾しないディレクレ分布の条件として、以下のハイパーパラメータを提案している。

$$\alpha_{ijk} = \alpha p(X_i = k, \Pi_i = j | g^h) \quad (5)$$

ここで、 α は Equivalent Sample Size (ESS) と呼ばれる事前知識の重みを示す擬似サンプルである。 g^h はユーザが事前に考えているネットワーク構造の仮説であり、この構造を仮定したときにユーザが主観的に設定した確率 $p(X_i = k, \Pi_i = j | g^h)$ に基づいて、ESS を α_{ijk} に分配する。この指標は、Bayesian Dirichlet equivalent (BDe) と呼ばれる。更に ESS をパラメータ数で除し、 $\alpha_{ijk} = \alpha / (r_i q_i)$ としたスコアを提案している。このスコアは BDe の特殊形とみなすことができ、Bayesian Dirichlet equivalence uniform (BDeu) と呼ばれる。BDeu は、

$$\begin{aligned} p(\mathbf{D} | g, \alpha) & \quad (6) \\ &= \prod_{i=1}^N \prod_{j=1}^{q_i} \frac{\Gamma(\alpha/q_i)}{\Gamma(\alpha/q_i + n_{ij})} \prod_{k=1}^{r_i} \frac{\Gamma(\alpha/(r_i q_i) + n_{ijk})}{\Gamma(\alpha/(r_i q_i))} \end{aligned}$$

として定義される。Heckerman ら [19] や Ueno [20]~[22] の研究では、ユーザが事前知識をもたない場合、無情報事前分布を用いた BDeu が最も有用であると報

告している。BDeu は、以下の漸近一致性をもつことが知られている [23]。

[定理 2.1] $n \rightarrow \infty$ のとき、BDeu を最大化するベイジアンネットワークの同時確率分布は真の分布に近づく。

証明については Koller ら [23] を参照してほしい。

一般にこの構造学習法は、厳密解探索アプローチと呼ばれる。しかし、このアプローチによる構造学習は NP 困難であり、ノード数の増加に伴い、計算量が爆発的に増加してしまう問題がある。厳密解探索を効率的に行うために、動的計画法 [4]~[8]、 A^* 探索 [9]、整数計画法 [10] といった従来の探索手法を用いた構造学習法が提案されてきた。しかし、現在の最先端手法を用いても、60 ノード程度の構造学習が限界である。

因果モデルの研究分野では、大幅に計算量を削減できる制約ベースアプローチと呼ばれる構造学習法が提案されてきた。このアプローチの基本的なアルゴリズムは以下のとおりである。

(1) 完全無向グラフを生成する。

(2) (1) で生成された完全無向グラフに対し条件付き独立性検定 (Conditional Independence test: CI テスト) により辺を削除する。

(3) (2) で得られた無向グラフに対してオリエンテーションルール [12] を用いて方向付けを行う。

制約ベースアルゴリズムとして、PC アルゴリズム [13]、MMHC アルゴリズム [14]、RAI アルゴリズム [15] が提案されてきた。しかし、これらのアルゴリズムでは χ^2 検定、 G^2 検定、条件付き相互情報量 (Conditional Mutual Information: CMI) を CI テストに用いるため、漸近一致性をもたない。

本論文の主なアイデアは、最新の制約ベースアプローチである RAI アルゴリズムに漸近一致性をもつ Bayes factor を用いた CI テストを埋め込むことにより、従来の厳密解探索アプローチに比べ計算時間を大幅に改善しようというものである。

3. CI テスト

制約ベースアルゴリズムによる学習は CI テストに精度が依存する。CI テストには、仮説検定法である χ^2 検定、 G^2 検定または CMI が用いられる。前者では一般に 2 ノード間が独立であることを帰無仮説と定める。その仮説を χ^2 統計量または G^2 統計量を用いて、帰無仮説が正しい確率を表す p 値を求め、ユーザが設定する有意水準により仮説の採択を判定する。後

者では、求まる CMI 値からしきい値により独立性を判定する。

3.1 χ^2 検定を用いた CI テスト

χ^2 検定を用いた CI テストでは、 χ^2 統計量と自由度 df の χ^2 分布から p 値を求め、その値が有意水準より小さいとき、帰無仮説を棄却し 2 ノード間の辺を残す。

今、2 ノード X, Y の共通の親ノード集合 \mathbf{Z} が z 番目のパターンを取るとき (以後、 $\mathbf{Z} = z$ と書く)、2 ノードが $X = x, Y = y$ となる条件付き頻度を n_{xyz} とし、 $\mathbf{Z} = z$ を所与として $X = x$ となる条件付き頻度を n_{xz} 、 $\mathbf{Z} = z$ を所与として $Y = y$ となる条件付き頻度を n_{yz} とすると、 χ^2 統計量 χ_{XYZ}^2 は、

$$\chi_{XYZ}^2 = \sum_{x=1}^{r_X} \sum_{y=1}^{r_Y} \sum_{z=1}^{r_Z} \frac{(n_{xyz} - n_{xz}n_{yz}/n)^2}{n_{xz}n_{yz}/n} \quad (7)$$

と定義される。ここで、 r_X, r_Y はそれぞれ、ノード X, Y の状態値の数を、 r_Z は \mathbf{Z} の取りうるパターン数を表す。 χ_{XYZ}^2 は適当な自由度をもつ χ^2 分布に漸近的に従う。欠測値がないことを仮定すると、自由度 df は、

$$df = (r_X - 1)(r_Y - 1) \prod_{z \in \mathbf{Z}} r_Z \quad (8)$$

となる。ここで、 r_Z は \mathbf{Z} に含まれるノード Z の状態値の数を表す。

3.2 G^2 検定を用いた CI テスト

χ^2 検定は、対数ゆう度の近似に基づく手法であった。一方で、 G^2 検定は対数ゆう度を直接用いる手法である。 G^2 統計量は、

$$G^2 = 2 \sum_{x=1}^{r_X} \sum_{y=1}^{r_Y} \sum_{z=1}^{r_Z} n_{xyz} \log \left(\frac{n_{xyz}}{n_{xz}n_{yz}/n} \right) \quad (9)$$

と定義される。 G^2 統計量は χ^2 統計量と同様、自由度 (8) の χ^2 分布に漸近的に従う。

3.3 CMI を用いた CI テスト

CMI を用いた CI テストは、以下の式 (10) の条件付き相互情報量 $\text{CMI}(X, Y | \mathbf{Z})$ が、しきい値より小さい場合に、2 ノード X, Y が独立と判定する。

$$\begin{aligned} \text{CMI}(X, Y | \mathbf{Z}) & \quad (10) \\ &= \sum_{x=1}^{r_X} \sum_{y=1}^{r_Y} \sum_{z=1}^{r_Z} p(x, y, z) \log \frac{p(x, y | z)}{p(x | z)p(y | z)} \end{aligned}$$

ここで、 $P(x, y, z)$ は $X = x, Y = y, \mathbf{Z} = z$ となる同時確率、 $p(x, y | z)$ は $\mathbf{Z} = z$ を所与として $X = x, Y = y$ となる条件付き確率、 $p(x | z)$ は $\mathbf{Z} = z$ を所与として $X = x$ となる条件付き確率、 $p(y | z)$ は $\mathbf{Z} = z$ を所与として $Y = y$ となる条件付き確率を表す。

3.4 CI テストの問題点

χ^2 検定、 G^2 検定を用いた CI テストは、 p 値と有意水準に精度が依存する。有意水準は、誤って帰無仮説を棄却してしまう第一種の過誤 (Type I error) を犯す確率を表し、一般的に 0.05 が最も用いられている。Cohen [24] では、検定の精度を保証する効果量を基にデータ数を決める必要があり、データ数を増やすことは検定の精度を保証しないと述べられている。実際に Gail ら [16] では、 p 値はデータ数の増加に伴い、小さい値になる傾向があり、Type I error の発生率が高まることが指摘されている。そのため、ベイジアンネットワークの学習でこれらを用いた場合、データ数を増加させても Type I error が発生し、不要な辺を付ける過学習が起こる。一方で、CMI を用いた CI テストは、設定するしきい値に強く影響を受けるため、漸近一致性をもたない。したがって、これらの CI テストを用いた制約ベースアプローチは、漸近的に真の条件付き独立構造を学習できる保証がない。

3.4.1 Type I error の検証

実際に 2 ノード間が真に独立である構造を用いて、データ数に関係なく、Type I error が発生することを検証する。具体的には、2 ノードのみの独立な構造において、各ノードの条件付き確率パラメータを 0.8, 0.7, 0.6 と尖度が高いものから緩やかな条件に設定し、データ数を $n = 10, 50, 100, 500, 1,000, 5,000, 10,000, 20,000$ と増やし、有意水準を 0.05 と定めた χ^2 検定、 G^2 検定を用いて CI テストを行う。

表 1 に各データ数における 100 回あたりの Type I error の発生率と p 値の平均を示す。表中の “ χ^2 ” の行は有意水準を 0.05 と定めた χ^2 検定を、“ G^2 ” の行は有意水準を 0.05 と定めた G^2 検定を表す。表より、 χ^2, G^2 とともにデータ数を大きくしたときに Type I error の発生率が 0.05 を超える場合があり不安定な振る舞いをしていることが分かる。 χ^2 検定では、データ数の増加とともに必ずしも Type I error が減少していないことが分かる。また G^2 検定は、 χ^2 検定と同等か劣る結果を示している。 p 値もデータ数の増加に対して安定していないことが確認できる。

以上の結果から、 χ^2 検定、 G^2 検定を用いた場合、

表 1 各検定の Type I error の発生率
Table 1 The percentage of Type I error each CI test.

条件付き確率パラメータ (0.8)	データ数	10	50	100	500	1,000	5,000	10,000	20,000
χ^2	Type I error の発生率	0.16	0.0	0.0	0.03	0.08	0.07	0.03	0.05
	p 値平均	0.0000	0.5623	0.5712	0.0361	0.0268	0.0232	0.0273	0.0207
G^2	Type I error の発生率	0.17	0.05	0.02	0.03	0.08	0.06	0.03	0.05
	p 値平均	0.0012	0.0383	0.0218	0.0325	0.0260	0.0192	0.0262	0.0214
条件付き確率パラメータ (0.7)	データ数	10	50	100	500	1,000	5,000	10,000	20,000
χ^2	Type I error の発生率	0.08	0.08	0.07	0.07	0.05	0.02	0.04	0.08
	p 値平均	0.0082	0.0198	0.0288	0.0197	0.0252	0.0359	0.0245	0.0264
G^2	Type I error の発生率	0.14	0.11	0.08	0.07	0.05	0.03	0.04	0.08
	p 値平均	0.0210	0.0180	0.0300	0.0194	0.0256	0.0410	0.0244	0.0265
条件付き確率パラメータ (0.6)	データ数	10	50	100	500	1,000	5,000	10,000	20,000
χ^2	Type I error の発生率	0.02	0.06	0.04	0.14	0.03	0.07	0.05	0.04
	p 値平均	0.0157	0.0252	0.0334	0.0276	0.0141	0.0321	0.0188	0.0230
G^2	Type I error の発生率	0.08	0.06	0.04	0.14	0.03	0.06	0.05	0.04
	p 値平均	0.0317	0.0251	0.0334	0.0275	0.0141	0.0293	0.0188	0.0230

有意水準程度の Type I error が発生してしまい、余分な辺を付ける傾向となり、漸近的に真の条件付き独立構造を学習できる保証がないと言える。

4. Bayes factor を組み込んだ RAI アルゴリズム

これまでに漸近一致性を有する CI テストが提案されている。この章では、Steck ら [17] が提案した Bayes factor を用いた CI テストを制約ベースアプローチである RAI (Recursive Autonomy Identification) アルゴリズム [15] に組み込むことで、漸近一致性をもった新しい学習手法を提案する。

Bayes factor は、二つのモデルの周辺ゆう度の比を求めることで厳密なモデル選択が可能な手法である。Steck ら [17] は、2 ノード間が独立・従属モデルの周辺ゆう度の比による Bayes factor を用いた CI テストを提案した。例として、 X と Y 間について各ノードの共通の親ノード集合を \mathbf{Z} としたときの従属モデルを g_1 、独立なモデルを g_2 とし、それぞれ図 1, 2 に示す。このときの Bayes factor を $\text{BF}(X, Y | \mathbf{Z})$ とすると、式 (4) を用いて、

$$\text{BF}(X, Y | \mathbf{Z}) = \frac{p(\mathbf{D} | g_1, \alpha)}{p(\mathbf{D} | g_2, \alpha)} \quad (11)$$

で表される。Steck らでは、これに対数を取った $\log \text{BF}(X, Y | \mathbf{Z})$ を求める。

$$\log \text{BF}(X, Y | \mathbf{Z}) = \log \frac{p(\mathbf{D} | g_1, \alpha)}{p(\mathbf{D} | g_2, \alpha)} \quad (12)$$

ただし Steck らでは、 $p(\mathbf{D} | g_1, \alpha)$, $p(\mathbf{D} | g_2, \alpha)$ に式 (6) を用いる。

Steck ら [17] では、式 (12) の対数 Bayes factor が

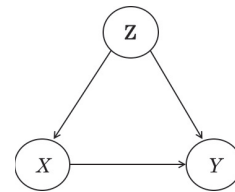


図 1 従属モデル g_1
Fig. 1 dependent model g_1 .

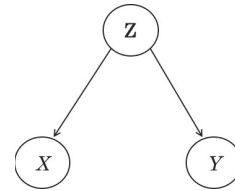


図 2 独立モデル g_2
Fig. 2 independent model g_2 .

0 以上か否かで図 1, 2 のどちらの構造を選択するか判定する CI テストを提案した。しかし、この CI テストは理論的な解析に用いらただけで、ベイジアンネットワークの構造学習に適用されていない。

また Steck ら [17] では言及していないが、BDeu に基づく Bayes factor は以下の漸近一致性をもつ。

[定理 4.1] $n \rightarrow \infty$ のとき、

(1) 真の構造が \mathbf{Z} を所与として X と Y が条件付き独立でないとき、 $\log \text{BF}(X, Y | \mathbf{Z}) > 0$ 。

(2) 真の構造が \mathbf{Z} を所与として X と Y が条件付き独立のとき、 $\log \text{BF}(X, Y | \mathbf{Z}) < 0$ 。

[証明]

(1) 真の構造が図 1 の g_1 のとき、定理 [2.1] より、 $n \rightarrow \infty$ のとき、 $p(\mathbf{D} | g_1, \alpha) > p(\mathbf{D} | g_2, \alpha)$ なので、 $\log \text{BF}(X, Y | \mathbf{Z}) > 0$ 。 g_1 における $X \rightarrow Y$

を $X \leftarrow Y$ に置き換えた g_3 が真の構造のとき、図 1 における g_3 は完全グラフである。ベイジアンネットワークが完全グラフ構造をもつとき、全ての辺の方向に対してマルコフ等価であることが証明されている [19]。更に BDeu はゆう度等価であるので $p(\mathbf{D} | g_1, \alpha) = p(\mathbf{D} | g_3, \alpha)$ が成り立ち、定理 [2.1] より、 $n \rightarrow \infty$ のとき、 $p(\mathbf{D} | g_3, \alpha) > p(\mathbf{D} | g_2, \alpha)$ なので、 $\log \text{BF}(X, Y | \mathbf{Z}) > 0$ 。

(2) 真の構造が \mathbf{Z} を所与として X と Y が条件付き独立のとき、定理 [2.1] より、 $n \rightarrow \infty$ のとき、 $p(\mathbf{D} | g_1, \alpha) < p(\mathbf{D} | g_2, \alpha)$ なので、 $\log \text{BF}(X, Y | \mathbf{Z}) < 0$ 。□

定理 [2.1] では、BDeu は真の構造ではなく真の同時確率分布に対してのみ漸近一致性をもっていたのに対して、定理 [4.1] では、提案された Bayes factor が 2 ノード間の真の条件付き独立構造に対して漸近一致性をもっていることを示している。これより、従来の統計的仮説検定手法が真の条件付き独立構造を判定できる保証がなかったのに対し、提案された Bayes factor がより高精度な検定を提供できる根拠となる。すなわち、 χ^2 検定、 G^2 検定では漸近一致性をもたず Type I error が有意水準に収束することを前提としているが、Bayes factor は Type I error が漸近的に 0.0 に収束する。

本論文では、BDeu に基づく Bayes factor を用いた CI テストを制約ベースアプローチで最も高精度と知られている RAI アルゴリズムに適用することで、大規模ネットワーク学習を実現する。

RAI アルゴリズムは、制約ベースアプローチにおいて最初に提案された PC アルゴリズム [13] を改良したものである。PC アルゴリズムでは、 $N - 2$ 個のノードを所与とした高次の CI テストまで繰り返す。しかし高次の CI テストは、低次のときに比べて信頼性が非常に低くなり、精度が著しく悪化する問題がある。RAI アルゴリズムは、その高次の CI テストを抑えるために開発された学習アルゴリズムである。RAI アルゴリズムは、各次数の CI テスト後にオリエンテーションルールによる辺の方向づけを行い、その結果を用いて全体グラフを部分グラフに分割する処理を繰り返すことで構造を学習する。

今、グラフを $g = (\mathbf{V}, \mathbf{E})$ と表し、 \mathbf{V} 、 \mathbf{E} はそれぞれ g に含まれるノード集合、辺集合を表す。ここで g は有向辺と無向辺を併せ持つとする。また、 $\text{Adj}(X, g)$ はグラフ g におけるノード X の隣接ノード

集合を表し、 $\text{Ch}(X, g)$ はグラフ g におけるノード X の子ノード集合を表す。このとき、 $\text{Pa}_p(X, g)$ は $\text{Adj}(X, g) \setminus \text{Ch}(X, g)$ を表し、 $\text{Pa}(X, g)$ はグラフ g に存在するノード X の親ノード集合を表す。また、 $\text{Pa}(X, \mathbf{g})$ はグラフ集合 \mathbf{g} において $\cup_{g \in \mathbf{g}} \text{Pa}(X, g)$ を表す。ここで、 g の部分構造 $g' = (\mathbf{V}', \mathbf{E}')$ が存在するとき、RAI アルゴリズムのグラフ分割では、以下に定義される外生因果及び自律的部分構造に分割を行う。

[定義 4.1] Y が $g' = (\mathbf{V}', \mathbf{E}')$ の外生因果 $\Leftrightarrow \forall Y \in \mathbf{V} \setminus \mathbf{V}', \forall X \in \mathbf{V}', Y \in \text{Adj}(X, g) \Rightarrow Y \in \text{Pa}(X, g)$

[定義 4.2] g' が自律的部分構造 $\Leftrightarrow \forall X \in \mathbf{V}', \text{Pa}_p(X, g) \subset \{\mathbf{V}' \cup \mathbf{V}_{ex}\}$

RAI アルゴリズムの詳細を Algorithm1 に示す。Algorithm1 では完全無向グラフ g_{uc} とデータ \mathbf{D} を入力として関数 RAI を再帰的に実行することで、学習結果の構造が出力として得られる。また関数 RAI 内の $X \perp Y | \mathbf{Z}$ は \mathbf{Z} を所与として X と Y が条件付き独立であることを表し、 $\mathbf{V}[i]$ はノード集合 \mathbf{V} の i 番目の要素を表す。 $\mathbf{g}[i]$ はグラフ集合 \mathbf{g} の i 番目の要素を表す。関数 RAI の概略は次のとおりである。入力グラフを $g_s = (\mathbf{V}_s, \mathbf{E}_s)$ とし、(1) 各次数の CI テストにおいて $X \perp Y | \mathbf{Z}$ となると、 XY 間の辺を削除する (8 行目から 23 行目)。(2) (1) により得られた無向グラフにオリエンテーションルールを適用して方向付けを行う (24 行目)。(3) 方向づけの結果から自律的部分構造を取り出す。具体的には、 \mathbf{V}_s の要素から子ノードをもつ集合 \mathbf{V}_p と子ノードをもたない集合 \mathbf{V}_c を取り出す。ここでは、 \mathbf{V}_c の要素が \mathbf{E}_s の無向辺集合 \mathbf{E}_U の要素を用いて \mathbf{V}_p のいずれかの要素に到達可能 [25] な場合、その要素を \mathbf{V}_c から削除する。また、 \mathbf{E}_U の要素のうち、 \mathbf{V}_c の要素を頂点にもつ辺集合を \mathbf{E}_c とし、 \mathbf{V}_c と \mathbf{E}_c で構成されるグラフを自律的部分構造として g_s から取り出す (25 行目から 35 行目)。(4) g_s から外生因果を構成するノード集合とそのノードを頂点にもつ辺集合を取り出す。このとき、取り出したノード集合と辺集合で定義されるグラフが非連結グラフとなる場合、非連結グラフ内の個々の連結グラフを列挙する。具体的には、 \mathbf{V}_s の要素がなくなるまで以下の手順を繰り返す。まず、 \mathbf{V}_s の任意の要素 V から到達可能な g_s のノード集合と V の和集合を \mathbf{V}_e とする。次に \mathbf{E}_s において、 \mathbf{V}_e の要素を頂点にもつ辺集合を \mathbf{E}_e とする。 \mathbf{V}_e と \mathbf{E}_e で構成されるグラフをグラフ集合 \mathbf{g}_e に追加し、 g_s から $(\mathbf{V}_e, \mathbf{E}_e)$ を取り除く。(36 行目から 43 行目)。(5) 各部分グラフで再帰的

Algorithm 1 The RAI algorithm

```

1: function MAIN( $g_{uc}, \mathbf{D}$ )
    $g_{uc} = (\mathbf{V}_{uc}, \mathbf{E}_{uc})$ : 完全無向グラフ
    $\mathbf{D}$ : データ
2:   return RAI(0,  $g_{uc}$ ,  $\phi$ ,  $g_{uc}$ ,  $\mathbf{D}$ )
3: end function

4: function RAI( $N_z, g_s, \mathbf{g}_{ex}, g_{all}, \mathbf{D}$ )
    $N_z$ : CI テストの回数
    $g_s = (\mathbf{V}_s, \mathbf{E}_s)$ : 入力グラフ
    $\mathbf{g}_{ex}$ : 分割されたグラフの集合
    $g_{all} = (\mathbf{V}_{all}, \mathbf{E}_{all})$ : CI テストと方向付けによって得られる出力グラフ
5:   if 全ての  $V \in \mathbf{V}_s$  について  $|\mathbf{Pa}_p(V, g_{all})| < N_z + 1$  then
6:     return  $g_{all}$ 
7:   end if
8:   // CI テストによる辺の削除
9:   for  $g_{ex} = (\mathbf{V}_{ex}, \mathbf{E}_{ex}) \in \mathbf{g}_{ex}$  do
10:    for  $X \in \mathbf{V}_s, Y \in \mathbf{V}_{ex}$  do
11:      for  $Z \subset \mathbf{Pa}_p(X, g_s) \cup \mathbf{Pa}(X, \mathbf{g}_{ex}) \setminus \{Y\}$  do
12:        if  $|Z| = N_z$  かつ  $X \perp Y \mid Z$  then
13:           $\mathbf{E}_{all} \leftarrow \mathbf{E}_{all} \setminus \{E_{XY}\}$   $\triangleright E_{XY}: XY$  間の辺
14:        end if
15:      end for
16:    end for
17:  end for
18:  for  $X \in \mathbf{V}_s, Y \in \mathbf{V}_s$  do
19:    for  $Z \subset \mathbf{Pa}_p(X, g_s) \cup \mathbf{Pa}(X, \mathbf{g}_{ex}) \setminus \{Y\}$  do
20:      if  $|Z| = N_z$  かつ  $X \perp Y \mid Z$  then
21:         $\mathbf{E}_{all} \leftarrow \mathbf{E}_{all} \setminus \{E_{XY}\}, \mathbf{E}_s \leftarrow \mathbf{E}_s \setminus \{E_{XY}\}$ 
22:      end if
23:    end for
24:  end for
25:  //  $g_s$  から自立的部分構造を分離
26:   $\mathbf{E}_U \leftarrow \mathbf{E}_s$  の無向辺集合
27:   $\mathbf{V}_c \leftarrow \mathbf{V}_s$  の子ノードをもたないノード集合
28:   $\mathbf{V}_p \leftarrow \mathbf{V}_s \setminus \mathbf{V}_c$ 
29:  for  $k = 1$  to  $|\mathbf{V}_c|$  do
30:    if  $\mathbf{V}_c[k]$  が  $\mathbf{E}_U$  の要素を用いて  $\mathbf{V}_p$  のいずれかの要素に到達可能
31:  then
32:     $\mathbf{V}_c \leftarrow \mathbf{V}_c \setminus \mathbf{V}_c[k]$ 
33:  end if
34:  end for
35:   $\mathbf{E}_c \leftarrow \{e \in \mathbf{E}_U \mid e \cap \mathbf{V}_c \neq \emptyset\}$   $\triangleright e$  の端点が一方でも  $\mathbf{V}_c$  に含まれる
36:   $\mathbf{E}_s \leftarrow \mathbf{E}_s \setminus \mathbf{E}_c$ 
37:   $\mathbf{V}_s \leftarrow \mathbf{V}_s \setminus \mathbf{V}_c$ 
38:  //  $g_s$  から外生因果を分離
39:   $\mathbf{g}_e \leftarrow \phi$ 
40:  for  $V \in \mathbf{V}_s$  do
41:     $\mathbf{V}_e \leftarrow \{V\} \cup \{V \text{ から到達可能な } g_s \text{ のノード集合}\}$ 
42:     $\mathbf{E}_e \leftarrow \{e \in \mathbf{E}_s \mid e \cap \mathbf{V}_e \neq \emptyset\}$   $\triangleright e$  の端点が一方でも  $\mathbf{V}_e$  に含まれる
43:  end for
44:   $\mathbf{g}_e \leftarrow \mathbf{g}_e \cup \{(\mathbf{V}_e, \mathbf{E}_e)\}$ 
45:   $\mathbf{V}_s \leftarrow \mathbf{V}_s \setminus \mathbf{V}_e$ 
46:   $\mathbf{E}_s \leftarrow \mathbf{E}_s \setminus \mathbf{E}_e$ 
47: end for
48: // 再帰的に関数 RAI を呼び出す
49: for  $i = 1$  to  $|\mathbf{g}_e|$  do
50:    $g_{all} \leftarrow \text{RAI}(N_z + 1, \mathbf{g}_e[i], \mathbf{g}_{ex}, g_{all}, \mathbf{D})$ 
51: end for
52:  $\mathbf{g}_{ex} \leftarrow \mathbf{g}_{ex} \cup \mathbf{g}_e$ 
53: return RAI( $N_z + 1, (\mathbf{V}_c, \mathbf{E}_c), \mathbf{g}_{ex}, g_{all}, \mathbf{D}$ )
54: end function

```

にRAIを呼び出す(44行目から48行目)。これにより、これまでの制約ベースアプローチのアルゴリズムにおいて最も高精度な構造学習を実現した。

RAIアルゴリズムの精度は、CIテストの途中経過で部分グラフに分割しながら学習するので、用いるCIテストの性能に大きく依存する。従来のRAIアルゴリズムでは、変数間のCMIの値がしきい値より小さいときに条件付き独立と判定する。しかし、この手法はしきい値に依存し漸近一致性をもたない問題がある。本提案では、RAIアルゴリズムにBayes factorを用いたCIテストを埋め込むので、条件付き独立構造の推定に関して漸近一致性が保証される。更に、精度の高いCIテストを組み入れることにより、RAIアルゴ

リズムにおける各ステップでの部分グラフへの分割精度も高くなり、学習精度のみならず高速化にも寄与できると期待される。

5. 評価実験

本章では、Bayes factorを用いたCIテストの有意性を示すために数種類の条件で実験を行う。具体的には、BDeuに基づくBayes factor (ESS = 1.0, 5.0, 10.0, 20.0), G^2 検定 (有意水準 0.05) (以後 G^2 と書く), χ^2 検定 (有意水準 0.05) (以後 χ^2 と書く), CMI (しきい値 0.003 (Yahezkel ら [15])) をRAIアルゴリズムに適用して、様々な規模のベンチマークネットワークの構造学習の精度を比較する。また、厳密解探索アプローチの最先端手法である整数計画法を用いた構造学習 (以後IPと書く) [10]を行い、学習精度を比較する。RAIアルゴリズムとIPの計算環境を表2に示す。

5.1 小規模ネットワークを用いた評価

本節では、図3に示す5変数で構成されるDAGに対して、図4, 5, 6に示す3パターンの条件付き確率を設定した場合の学習精度を比較する。これは、条件付き確率パラメータが一様分布から尖度の高い分布について、提案手法がIPと同様に真の構造を推定できることを示すためである。

本比較では、条件付き確率の三つのパターンにおいて、データ数を $n = 100, 200, 500, 1,000, 2,000, 5,000, 10,000$ でそれぞれ実験を行い、学習精度の変移を検証した。実験手順は以下のとおりである。

(1) 各パターンの真のネットワーク構造からデータセットをランダムに $n = 100 \sim 10,000$ 個発生させる。

(2) 手順(1)で発生させたデータを用いて、RAIアルゴリズムに各CIテストを適用して構造学習する。

(3) 手順(2)を30回繰り返す。

本実験のStructural Hamming Distance (SHD) [14], Missing Edge, Extra Edgeの結果を表3に表す。Missing Edgeは真の構造では辺が存在するが学習の過程で削除された辺、Extra Edgeは真の構造では辺が存在しないが学習の結果残った辺、SHDは真の構造と推定された構造の距離を表す。SHDが0.0に収束することで、真の構造と推定された構造が一致したことを表す。表中では、SHD, Missing Edge, Extra Edgeの標準偏差を括弧書きで示している。また、各データ数における各手法の計算時間を表4に表す。

表 2 計算環境
Table 2 Computational environment.

RAI アルゴリズム	
CPU	12-Cores Intel XEON
System Memory	128GB
OS	MacOS X 10.11.6
ソフトウェア	MATLAB
IP	
CPU	12-Cores Intel XEON
System Memory	128GB
OS	MacOS X 10.11.6
ソフトウェア	Gobnilp(C)

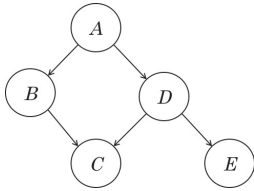


図 3 5 変数の DAG
Fig. 3 5 variables DAG.

B	A	$p(B A)$	A	$p(A)$	
1	0	0.1	1	0.5	
1	1	0.9			
$p(C B,D)$			D	$p(D A)$	
C	B	D	1	0	0.1
1	0	0	1	1	0.9
1	0	1	E	D	$p(E D)$
1	1	0	1	0	0.1
1	1	1	1	1	0.9

図 4 偏りが非常に強い分布
Fig. 4 Strongly skewed Distribution.

B	A	$p(B A)$	A	$p(A)$	
1	0	0.2	1	0.5	
1	1	0.8			
$p(C B,D)$			D	$p(D A)$	
C	B	D	1	0	0.2
1	0	0	1	1	0.8
1	0	1	E	D	$p(E D)$
1	1	0	1	0	0.2
1	1	1	1	1	0.8

図 5 偏りが強い分布
Fig. 5 Skewed Distribution.

B	A	$p(B A)$	A	$p(A)$	
1	0	0.3	1	0.5	
1	1	0.7			
$p(C B,D)$			D	$p(D A)$	
C	B	D	1	0	0.3
1	0	0	1	1	0.7
1	0	1	E	D	$p(E D)$
1	1	0	1	0	0.3
1	1	1	1	1	0.7

図 6 偏りが弱い分布
Fig. 6 Uniform Distribution.

SHD の結果より、データ数の増加に伴い、Bayes factor, IP は共に SHD が 0.0 に収束し、最も高い学習精度を示した。また Skewed Distribution, Uniform Distribution と条件付き確率パラメータが一樣分布に近づく、IP がまず 0.0 に収束し、次に ESS を 5.0, 10.0, 20.0 に設定した Bayes factor が 0.0 に収束した。CMI は、Strongly Skewed Distribution, Skewed Distribution の二つの分布で、SHD が 0.0 に収束した。しかし、Uniform Distribution では、SHD が極端に大きい値となった。この理由として、CMI はしきい値に強く影響を受け、学習するネットワークによって精度が安定しないためである。 G^2 , χ^2 は、少ないデータ数では三つの条件ともに Bayes factor に比べて高い精度を示した。特に Skewed Distribution ではいったん SHD が 0.0 となる場合があった。しかし、データ数の増加により、再び SHD が上昇していることが分かる。3.4.1 で示したとおり、 G^2 検定、 χ^2 検定はデータ数が増加しても Type I error の発生率が減少する保証がないため、最終的に過学習してしまい Extra Edge を多くつけすぎってしまったことに原因する。

Missing Edge の結果では、三つの条件付きパラメータにおいて、全ての手法で 0.0 に収束した。少ないデータ数では ESS を 20.0 と定めた Bayes factor が最も小さい値を示した。また Uniform Distribution では、Bayes factor を用いた手法の中で ESS を 20.0 と定めた Bayes factor が最も小さい値を示した。これは、Ueno [21] において、条件付き確率パラメータが一樣分布に近づくにつれ、最適な ESS が大きくなることを理論的に示していることから理解できる。

SHD, Missing Edge, Extra Edge の標準偏差の結果より、Bayes factor, IP ではデータ数の増加に伴い、この値が単調減少したことが分かる。一方で、 G^2 , χ^2 , CMI では単調減少しなかった。これは、Bayes factor, IP は漸近一致性をもつため、十分なデータ数による学習において、安定した学習精度が保証されていることを示している。一方、 G^2 , χ^2 , CMI は漸近一致性をもたず学習精度が不安定であることが分かる。

計算時間の評価では、全般に IP が最も速いことが分かる。一方で、Bayes factor は G^2 , χ^2 , CMI に比べて学習時間が遅い。この理由として、Bayes factor がそれらに比べて CI テストの実施回数が多かったためと考えられる。

表 3 5 変数の実験結果
Table 3 The experiment results for 5 variables networks.

SHD								
Strongly Skewed	Bayes factor (1.0)	Bayes factor(5.0)	Bayes factor(10.0)	Bayes factor(20.0)	G^2	χ^2	CMI	IP
100	2.8(1.32)	2.3(1.42)	3.6(1.43)	4.7(1.06)	2.4(1.08)	2.1(1.10)	5.3(1.77)	2.7(1.89)
200	1.9(0.74)	1.3(1.42)	2.1(1.52)	3.4(1.78)	1.1(0.88)	0.9(1.29)	4.4(1.58)	1.1(1.52)
500	0.8(1.32)	0.3(0.95)	0.4(0.97)	0.8(1.48)	0.4(0.97)	0.4(0.97)	2.3(1.64)	0.0(0.00)
1,000	0.0(0.00)	0.0(0.00)	0.0(0.00)	0.0(0.00)	0.3(0.68)	0.0(0.00)	0.1(0.32)	0.2(0.42)
2,000	0.0(0.00)	0.0(0.00)	0.0(0.00)	0.4(0.97)	0.4(0.97)	0.4(0.97)	0.1(0.32)	0.0(0.00)
5,000	0.0(0.00)	0.4(1.26)	0.2(0.63)	0.3(0.68)	0.1(0.32)	0.1(0.32)	0.0(0.00)	0.0(0.00)
10,000	0.0(0.00)	0.0(0.00)	0.0(0.00)	0.0(0.00)	1.2(1.55)	0.0(0.00)	0.0(0.00)	0.0(0.00)
Skewed	Bayes factor (1.0)	Bayes factor(5.0)	Bayes factor(10.0)	Bayes factor(20.0)	G^2	χ^2	CMI	IP
100	4.6(0.70)	2.9(1.52)	3.5(1.65)	3.6(1.43)	3.1(1.52)	3.2(1.52)	6.7(0.95)	2.4(0.84)
200	3.5(1.65)	2.5(1.18)	2.6(1.17)	1.9(1.60)	2.5(1.18)	2.3(1.18)	4.4(2.32)	2.1(1.66)
500	2.0(1.33)	1.6(1.27)	1.5(1.65)	1.0(1.63)	1.2(1.62)	1.1(1.62)	1.8(1.40)	0.0(0.00)
1,000	1.8(1.40)	0.6(1.27)	0.3(0.95)	0.3(0.95)	0.0(0.00)	0.0(0.00)	0.4(0.97)	0.0(0.00)
2,000	0.6(1.27)	0.5(1.08)	0.5(1.08)	0.5(1.08)	0.8(1.32)	0.7(1.32)	0.5(1.08)	0.0(0.00)
5,000	0.0(0.00)	0.0(0.00)	0.0(0.00)	0.0(0.00)	0.0(0.00)	0.9(1.45)	0.0(0.00)	0.0(0.00)
10,000	0.0(0.00)	0.0(0.00)	0.0(0.00)	0.0(0.00)	0.1(0.32)	0.4(0.32)	0.0(0.00)	0.0(0.00)
Uniform	Bayes factor (1.0)	Bayes factor(5.0)	Bayes factor(10.0)	Bayes factor(20.0)	G^2	χ^2	CMI	IP
100	4.8(0.63)	4.7(0.82)	4.8(0.63)	4.7(0.68)	4.9(0.57)	4.9(0.50)	6.2(0.79)	2.6(0.84)
200	4.7(0.68)	4.6(0.97)	4.4(1.08)	4.7(0.95)	4.7(1.16)	4.7(1.16)	5.9(0.88)	2.1(0.32)
500	4.8(0.42)	4.5(0.71)	4.6(0.70)	4.4(0.70)	4.5(0.71)	4.5(0.71)	5.1(1.20)	2.0(1.63)
1,000	4.4(0.52)	3.6(1.58)	3.4(1.51)	3.1(1.52)	3.1(1.60)	3.1(1.60)	4.1(0.88)	0.3(0.95)
2,000	2.8(1.14)	2.5(1.18)	2.0(1.33)	1.9(1.16)	1.3(1.16)	1.3(1.16)	3.2(0.79)	0.3(0.95)
5,000	1.0(1.33)	0.0(0.00)	0.0(0.00)	0.0(0.00)	0.3(0.95)	0.3(0.95)	3.6(0.52)	0.0(0.00)
10,000	0.0(0.00)	0.0(0.00)	0.0(0.00)	0.0(0.00)	0.4(0.97)	0.4(0.97)	3.7(0.68)	0.0(0.00)
Missing Edge								
Strongly Skewed	Bayes factor (1.0)	Bayes factor(5.0)	Bayes factor(10.0)	Bayes factor(20.0)	G^2	χ^2	CMI	IP
100	0.8(0.63)	0.4(0.52)	0.4(0.52)	0.2(0.42)	0.5(0.53)	0.5(0.53)	0.0(0.00)	0.5(0.53)
200	0.4(0.70)	0.1(0.32)	0.1(0.32)	0.1(0.32)	0.3(0.68)	0.1(0.32)	0.1(0.32)	0.1(0.32)
500	0.0(0.00)	0.0(0.00)	0.0(0.00)	0.0(0.00)	0.0(0.00)	0.0(0.00)	0.0(0.00)	0.0(0.00)
1,000	0.0(0.00)	0.0(0.00)	0.0(0.00)	0.0(0.00)	0.0(0.00)	0.0(0.00)	0.0(0.00)	0.0(0.00)
2,000	0.0(0.00)	0.0(0.00)	0.0(0.00)	0.0(0.00)	0.0(0.00)	0.0(0.00)	0.0(0.00)	0.0(0.00)
5,000	0.0(0.00)	0.0(0.00)	0.0(0.00)	0.0(0.00)	0.0(0.00)	0.0(0.00)	0.0(0.00)	0.0(0.00)
10,000	0.0(0.00)	0.0(0.00)	0.0(0.00)	0.0(0.00)	0.0(0.00)	0.0(0.00)	0.0(0.00)	0.0(0.00)
Skewed	Bayes factor (1.0)	Bayes factor(5.0)	Bayes factor(10.0)	Bayes factor(20.0)	G^2	χ^2	CMI	IP
100	1.8(0.63)	0.7(0.68)	0.8(1.03)	0.3(0.48)	0.6(0.70)	0.7(0.68)	0.0(0.00)	0.8(0.42)
200	1.0(0.67)	0.4(0.52)	0.3(0.48)	0.3(0.48)	0.3(0.48)	0.3(0.48)	0.0(0.00)	0.3(0.48)
500	0.2(0.63)	0.0(0.00)	0.0(0.00)	0.0(0.00)	0.0(0.00)	0.0(0.00)	0.0(0.00)	0.0(0.00)
1,000	0.0(0.00)	0.0(0.00)	0.0(0.00)	0.0(0.00)	0.0(0.00)	0.0(0.00)	0.0(0.00)	0.0(0.00)
2,000	0.0(0.00)	0.0(0.00)	0.0(0.00)	0.0(0.00)	0.0(0.00)	0.0(0.00)	0.0(0.00)	0.0(0.00)
5,000	0.0(0.00)	0.0(0.00)	0.0(0.00)	0.0(0.00)	0.0(0.00)	0.0(0.00)	0.0(0.00)	0.0(0.00)
10,000	0.0(0.00)	0.0(0.00)	0.0(0.00)	0.0(0.00)	0.0(0.00)	0.0(0.00)	0.0(0.00)	0.0(0.00)
Uniform	Bayes factor (1.0)	Bayes factor(5.0)	Bayes factor(10.0)	Bayes factor(20.0)	G^2	χ^2	CMI	IP
100	2.4(0.70)	1.6(0.52)	1.5(0.53)	1.3(0.68)	1.7(0.48)	1.7(0.48)	0.1(0.32)	1.7(0.48)
200	1.7(0.68)	1.5(0.53)	1.3(0.68)	1.0(0.67)	1.2(0.63)	1.2(0.63)	0.3(0.48)	1.2(0.42)
500	1.1(0.32)	0.7(0.48)	0.3(0.48)	0.2(0.42)	0.1(0.32)	0.1(0.32)	0.1(0.32)	0.4(0.52)
1,000	0.8(0.42)	0.1(0.32)	0.1(0.32)	0.0(0.00)	0.0(0.00)	0.0(0.00)	0.0(0.00)	0.0(0.00)
2,000	0.1(0.32)	0.0(0.00)	0.0(0.00)	0.0(0.00)	0.0(0.00)	0.0(0.00)	0.0(0.00)	0.0(0.00)
5,000	0.0(0.00)	0.0(0.00)	0.0(0.00)	0.0(0.00)	0.0(0.00)	0.0(0.00)	0.0(0.00)	0.0(0.00)
10,000	0.0(0.00)	0.0(0.00)	0.0(0.00)	0.0(0.00)	0.0(0.00)	0.0(0.00)	0.0(0.00)	0.0(0.00)
Extra Edge								
Strongly Skewed	Bayes factor (1.0)	Bayes factor(5.0)	Bayes factor(10.0)	Bayes factor(20.0)	G^2	χ^2	CMI	IP
100	0.2(0.42)	0.2(0.42)	1.0(0.94)	1.8(0.63)	0.1(0.32)	0.0(0.00)	3.5(0.70)	0.7(0.68)
200	0.0(0.00)	0.3(0.48)	0.5(0.53)	0.8(0.42)	0.3(0.48)	0.3(0.48)	2.6(0.97)	0.2(0.42)
500	0.0(0.00)	0.0(0.00)	0.1(0.32)	0.2(0.42)	0.1(0.32)	0.1(0.32)	1.5(0.85)	0.0(0.00)
1,000	0.0(0.00)	0.0(0.00)	0.0(0.00)	0.0(0.00)	0.1(0.32)	0.1(0.32)	0.2(0.42)	0.0(0.00)
2,000	0.0(0.00)	0.0(0.00)	0.0(0.00)	0.2(0.42)	0.2(0.42)	0.2(0.42)	0.1(0.32)	0.0(0.00)
5,000	0.0(0.00)	0.0(0.00)	0.0(0.00)	0.0(0.00)	0.1(0.32)	0.1(0.32)	0.0(0.00)	0.0(0.00)
10,000	0.0(0.00)	0.0(0.00)	0.0(0.00)	0.0(0.00)	0.4(0.52)	0.4(0.52)	0.0(0.00)	0.0(0.00)
Skewed	Bayes factor (1.0)	Bayes factor(5.0)	Bayes factor(10.0)	Bayes factor(20.0)	G^2	χ^2	CMI	IP
100	0.2(0.42)	0.2(0.42)	0.2(0.42)	0.3(0.48)	0.2(0.48)	0.1(0.32)	3.4(1.08)	0.0(0.00)
200	0.1(0.32)	0.1(0.32)	0.2(0.42)	0.1(0.32)	0.1(0.32)	0.0(0.00)	2.1(0.99)	0.0(0.00)
500	0.0(0.00)	0.0(0.00)	0.1(0.32)	0.1(0.32)	0.1(0.32)	0.2(0.42)	1.2(0.92)	0.0(0.00)
1,000	0.0(0.00)	0.0(0.00)	0.0(0.00)	0.0(0.00)	0.0(0.00)	0.0(0.00)	0.2(0.42)	0.0(0.00)
2,000	0.0(0.00)	0.0(0.00)	0.0(0.00)	0.0(0.00)	0.3(0.48)	0.2(0.42)	0.0(0.00)	0.0(0.00)
5,000	0.0(0.00)	0.0(0.00)	0.0(0.00)	0.0(0.00)	0.2(0.42)	0.2(0.42)	0.0(0.00)	0.0(0.00)
10,000	0.0(0.00)	0.0(0.00)	0.0(0.00)	0.0(0.00)	0.1(0.32)	0.1(0.32)	0.0(0.00)	0.0(0.00)
Uniform	Bayes factor (1.0)	Bayes factor(5.0)	Bayes factor(10.0)	Bayes factor(20.0)	G^2	χ^2	CMI	IP
100	0.1(0.32)	0.2(0.42)	0.4(0.52)	0.3(0.68)	0.3(0.48)	0.3(0.48)	2.3(0.82)	0.2(0.42)
200	0.2(0.42)	0.3(0.48)	0.3(0.48)	0.4(0.52)	0.4(0.52)	0.4(0.52)	1.3(0.68)	0.1(0.32)
500	0.0(0.00)	0.0(0.00)	0.1(0.32)	0.1(0.32)	0.1(0.32)	0.1(0.32)	0.8(0.79)	0.0(0.00)
1,000	0.0(0.00)	0.0(0.00)	0.0(0.00)	0.0(0.00)	0.1(0.32)	0.1(0.32)	0.1(0.32)	0.0(0.00)
2,000	0.0(0.00)	0.0(0.00)	0.0(0.00)	0.0(0.00)	0.2(0.42)	0.2(0.42)	0.0(0.00)	0.0(0.00)
5,000	0.0(0.00)	0.0(0.00)	0.0(0.00)	0.0(0.00)	0.0(0.00)	0.0(0.00)	0.0(0.00)	0.0(0.00)
10,000	0.0(0.00)	0.0(0.00)	0.0(0.00)	0.0(0.00)	0.2(0.42)	0.2(0.42)	0.0(0.00)	0.0(0.00)

5.2 ベンチマークネットワークを用いた評価
本節では、ベイジアンネットワークのリポジトリ *bnlearn* [26] に登録されている 8 種類のベンチマークネットワークを用いて実験を行う。ベンチマークネッ

トワークの情報と実験を行ったデータ数について表 5 に示す。表中の変数数、辺数、最大親変数数、パラメータ数はネットワークの規模を表している。

実験手順は前節と同様である。また IP による学

表 4 5 変数の計算時間 (s)
Table 4 The results of computational time (s) for 5 variables networks.

Strongly Skewed	Bayes factor (1.0)	Bayes factor(5.0)	Bayes factor(10.0)	Bayes factor(20.0)	G^2	χ^2	CMI	IP
100	0.0474	0.0528	0.0659	0.0878	0.0478	0.0471	0.0784	0.0092
200	0.0558	0.0515	0.0585	0.0763	0.0435	0.0468	0.0670	0.0095
500	0.0582	0.0575	0.0569	0.0594	0.0515	0.0469	0.0555	0.0103
1,000	0.0629	0.0582	0.0569	0.0616	0.0535	0.0572	0.0482	0.0109
2,000	0.0635	0.0609	0.0604	0.0626	0.0516	0.0501	0.0536	0.0151
5,000	0.0777	0.0679	0.0690	0.0710	0.0606	0.0608	0.0530	0.0229
10,000	0.0882	0.0792	0.0803	0.0814	0.0654	0.0640	0.0641	0.0378
Skewed	Bayes factor (1.0)	Bayes factor(5.0)	Bayes factor(10.0)	Bayes factor(20.0)	G^2	χ^2	CMI	IP
100	0.0362	0.0395	0.0452	0.0523	0.0335	0.0335	0.0792	0.0105
200	0.0404	0.0429	0.0465	0.0521	0.0386	0.0403	0.0551	0.0108
500	0.0609	0.0507	0.0548	0.0574	0.0455	0.0472	0.0577	0.0111
1,000	0.0614	0.0574	0.0580	0.0580	0.0564	0.0475	0.0520	0.0127
2,000	0.0642	0.0592	0.0626	0.0622	0.0544	0.0550	0.0497	0.0164
5,000	0.0755	0.0636	0.0686	0.0681	0.0637	0.0540	0.0562	0.0238
10,000	0.0846	0.0783	0.0782	0.0797	0.0644	0.0654	0.0618	0.0375
Uniform	Bayes factor (1.0)	Bayes factor(5.0)	Bayes factor(10.0)	Bayes factor(20.0)	G^2	χ^2	CMI	IP
100	0.0244	0.0262	0.0287	0.0348	0.0265	0.0248	0.0554	0.0110
200	0.0273	0.0329	0.0352	0.0387	0.0313	0.0293	0.0409	0.0110
500	0.0437	0.0400	0.0427	0.0441	0.0423	0.0360	0.0392	0.0124
1,000	0.0464	0.0455	0.0499	0.0510	0.0502	0.0442	0.0383	0.0137
2,000	0.0557	0.0498	0.0526	0.0535	0.0467	0.0505	0.0451	0.0175
5,000	0.0710	0.0679	0.0670	0.0683	0.0612	0.0594	0.0474	0.0240
10,000	0.0879	0.0776	0.0790	0.0794	0.0624	0.0621	0.0545	0.0396

表 5 ベンチマークネットワーク
Table 5 Benchmark networks.

network	変数数	辺数	最大 親変数数	パラ メータ数	データ数 n
cancer	5	4	2	10	10,000 ~ 2,000,000
earthquake	5	4	2	10	10,000 ~ 200,000
survey	6	6	2	21	10,000 ~ 200,000
sachs	11	17	3	178	10,000 ~ 200,000
child	20	25	2	230	10,000 ~ 10,000,000
alarm	37	46	4	509	10,000 ~ 20,000,000
win95pts	76	112	7	574	10,000 ~ 2,000,000
andes	223	338	6	1157	10,000 ~ 2,000,000

習については 12 時間で打ち切った。この理由は、Cussens [10] では 6 時間の制限時間を定めているが、本実験では、より長い制限時間を設けても厳密探索アプローチでは大規模ネットワーク学習を実現できないことを示すためである。

表 6~13 に各ベンチマークネットワークの SHD, Missing Edge, Extra Edge と計算時間の結果を示す。表中の“-”は制限時間内に学習できなかったことを表す。SHD の結果より、データ数が少ないとき、Bayes factor は G^2 , χ^2 に比べ精度は劣るが、データ数の増加に対して単調減少し最も高い精度を示した。cancer, earthquake, survey, sachs では、Bayes factor はデータ数の増加により、SHD が 0.0 に収束し IP と同等の精度をもち、真の構造を推定できることが分かる。IP は、cancer, earthquake, survey, sachs, child では制限時間内に学習でき、少ないデータ数から SHD が 0.0 に収束した。また alarm では、データ数が 2,000,000 以上では制限時間内に学習できず、大規模ネットワークの win95pts, andes では、制限時間内で学習できなかった。Bayes factor は、child, alarm,

win95pts, andes で SHD が 0.0 に収束しなかったが、alarm, win95pts, andes では IP を除いた手法の中で最も高い精度を示した。また Bayes factor では、ESS が大きい値になるほど Missing Edge が減少した。大規模構造ではパラメータ数が多く、各パラメータに対してデータがスパースになる。前節で述べたように、スパースなデータでは最適な ESS が理論的に大きくなることから知られている (Ueno [21])。このことから、win95pts, andes での Missing Edge は、ESS を 10.0 または 20.0 に設定した Bayes factor が最も減少した。しかし、ESS を大きな値にするほど過学習する傾向があるため、ESS を小さい値 (ESS = 1.0, 5.0) に設定した場合に比べ Extra Edge が極端に多い結果が示された。これより、win95pts, andes の SHD は ESS を小さい値にした場合の方が高い学習精度を示した。

一方で、 G^2 , χ^2 はデータ数の増加に対して、SHD は単調減少せず精度が不安定となった。cancer, earthquake, sachs において途中まで SHD が減少傾向であったが、データ数の増加に伴い再び上昇した。特に earthquake の結果では、データ数が 100,000 で SHD が 0.0 に収束したが、データ数が 200,000 で再び上昇した。survey の結果では序盤から SHD が上昇した。これら四つのネットワークの実験結果では、 G^2 , χ^2 の Missing Edge は少ないデータ数で 0.0 に収束した。しかし Extra Edge は途中まで減少または 0.0 に収束したが、データ数の増加に伴い、再び上昇している。child, alarm, win95pts, andes における G^2 , χ^2 の SHD は、データ数の増加に対して増加するかほぼ一定の値を示した。Missing Edge は 0.0 に収束するか

表 6 cancer の学習結果と計算時間
Table 6 The experiment results for cancer.

SHD								
データ数	Bayes factor (1.0)	Bayes factor(5.0)	Bayes factor(10.0)	Bayes factor(20.0)	G^2	χ^2	CMI	IP
10,000	2.7(0.68)	2.8(0.79)	2.8(0.79)	2.9(0.88)	3.4(1.08)	3.4(1.08)	3.0(0.00)	0.0(0.00)
20,000	2.8(0.79)	2.7(0.48)	2.8(0.63)	3.2(0.63)	3.2(1.32)	3.2(1.32)	3.0(0.00)	0.0(0.00)
200,000	2.8(1.99)	2.2(1.62)	1.9(1.29)	1.7(0.68)	0.7(0.95)	0.7(0.95)	3.0(0.00)	0.0(0.00)
500,000	1.2(1.03)	1.2(1.03)	1.2(1.03)	1.2(1.03)	1.1(1.60)	1.1(1.60)	3.0(0.00)	0.0(0.00)
1,000,000	0.6(0.97)	0.2(0.63)	0.2(0.63)	0.2(0.63)	0.3(0.48)	0.3(0.48)	3.0(0.00)	0.0(0.00)
2,000,000	0.0(0.00)	0.0(0.00)	0.0(0.00)	0.0(0.00)	0.0(0.00)	0.1(0.32)	3.0(0.00)	0.0(0.00)
Missing Edge								
10,000	0.7(0.48)	0.2(0.42)	0.1(0.32)	0.1(0.32)	0.1(0.32)	0.1(0.32)	1.8(0.42)	0.0(0.00)
20,000	0.4(0.52)	0.0(0.00)	0.0(0.00)	0.0(0.00)	0.0(0.00)	0.0(0.00)	1.4(0.52)	0.0(0.00)
200,000	0.0(0.00)	0.0(0.00)	0.0(0.00)	0.0(0.00)	0.0(0.00)	0.0(0.00)	1.9(0.32)	0.0(0.00)
500,000	0.0(0.00)	0.0(0.00)	0.0(0.00)	0.0(0.00)	0.0(0.00)	0.0(0.00)	2.0(0.00)	0.0(0.00)
1,000,000	0.0(0.00)	0.0(0.00)	0.0(0.00)	0.0(0.00)	0.0(0.00)	0.0(0.00)	2.0(0.00)	0.0(0.00)
2,000,000	0.0(0.00)	0.0(0.00)	0.0(0.00)	0.0(0.00)	0.0(0.00)	0.0(0.00)	2.0(0.00)	0.0(0.00)
Extra Edge								
10,000	0.2(0.42)	0.6(0.70)	0.7(0.68)	0.8(0.79)	1.2(1.14)	1.2(1.14)	0.0(0.00)	0.0(0.00)
20,000	0.4(0.52)	0.7(0.48)	0.8(0.63)	1.2(0.63)	1.5(0.85)	1.5(0.85)	0.0(0.00)	0.0(0.00)
200,000	1.5(1.35)	1.1(1.10)	0.9(0.88)	0.8(0.42)	0.4(0.52)	0.4(0.52)	0.0(0.00)	0.0(0.00)
500,000	0.6(0.52)	0.6(0.52)	0.6(0.52)	0.6(0.52)	0.5(0.53)	0.5(0.53)	0.0(0.00)	0.0(0.00)
1,000,000	0.3(0.48)	0.1(0.32)	0.1(0.32)	0.1(0.32)	0.3(0.48)	0.3(0.48)	0.0(0.00)	0.0(0.00)
2,000,000	0.0(0.00)	0.0(0.00)	0.0(0.00)	0.0(0.00)	0.1(0.32)	0.1(0.32)	0.0(0.00)	0.0(0.00)
計算時間 (s)								
10,000	0.0882	0.0761	0.0806	0.0807	0.0652	0.0627	0.0250	0.0427
20,000	0.1467	0.1214	0.1182	0.1259	0.0991	0.0973	0.0410	0.0679
200,000	0.9085	0.7051	0.6840	0.7692	0.4954	0.5022	0.1550	0.4646
500,000	2.4587	1.7595	1.6995	2.0760	1.4797	1.5308	0.4735	1.0926
1,000,000	6.1644	4.6628	4.6821	4.9112	3.5117	3.5669	0.7826	2.1212
2,000,000	13.2400	9.8458	9.9045	11.5970	6.9856	7.1896	1.8872	4.2360

表 7 earthquake の学習精度と計算時間
Table 7 The experiment results for earthquake.

SHD								
データ数	Bayes factor (1.0)	Bayes factor(5.0)	Bayes factor(10.0)	Bayes factor(20.0)	G^2	χ^2	CMI	IP
10,000	1.5(2.42)	0.2(0.42)	2.0(1.63)	6.2(1.75)	1.6(2.37)	0.3(0.48)	0.1(0.32)	1.5(2.42)
20,000	0.0(0.00)	0.7(1.57)	1.8(2.25)	5.2(1.99)	0.9(1.52)	0.3(0.48)	0.1(0.32)	0.0(0.00)
50,000	0.0(0.00)	0.5(1.58)	0.6(1.58)	5.0(2.11)	0.5(1.58)	0.5(1.58)	0.0(0.00)	0.0(0.00)
100,000	0.0(0.00)	0.0(0.00)	0.0(0.00)	0.9(1.91)	0.0(0.00)	0.0(0.00)	0.0(0.00)	0.0(0.00)
200,000	0.0(0.00)	0.0(0.00)	0.0(0.00)	0.0(0.00)	0.6(1.58)	0.7(1.57)	0.0(0.00)	0.0(0.00)
Missing Edge								
10,000	0.0(0.00)	0.0(0.00)	0.0(0.00)	0.0(0.00)	0.0(0.00)	0.0(0.00)	0.1(0.32)	0.0(0.00)
20,000	0.0(0.00)	0.0(0.00)	0.0(0.00)	0.0(0.00)	0.0(0.00)	0.0(0.00)	0.1(0.32)	0.0(0.00)
50,000	0.0(0.00)	0.0(0.00)	0.0(0.00)	0.0(0.00)	0.0(0.00)	0.0(0.00)	0.0(0.00)	0.0(0.00)
100,000	0.0(0.00)	0.0(0.00)	0.0(0.00)	0.0(0.00)	0.0(0.00)	0.0(0.00)	0.0(0.00)	0.0(0.00)
200,000	0.0(0.00)	0.0(0.00)	0.0(0.00)	0.0(0.00)	0.0(0.00)	0.0(0.00)	0.0(0.00)	0.0(0.00)
Extra Edge								
10,000	0.3(0.48)	0.2(0.42)	1.2(0.42)	2.7(0.82)	0.4(0.52)	0.3(0.48)	0.0(0.00)	0.3(0.48)
20,000	0.0(0.00)	0.3(0.48)	0.6(0.52)	2.1(0.57)	0.5(0.53)	0.3(0.48)	0.0(0.00)	0.0(0.00)
50,000	0.0(0.00)	0.1(0.32)	0.2(0.42)	1.8(0.42)	0.1(0.32)	0.1(0.32)	0.0(0.00)	0.0(0.00)
100,000	0.0(0.00)	0.0(0.00)	0.0(0.00)	0.5(0.85)	0.0(0.00)	0.0(0.00)	0.0(0.00)	0.0(0.00)
200,000	0.0(0.00)	0.0(0.00)	0.0(0.00)	0.0(0.00)	0.2(0.42)	0.3(0.48)	0.0(0.00)	0.0(0.00)
計算時間 (s)								
10,000	0.1579	0.0995	0.1109	0.1486	0.0730	0.0674	0.0792	0.0386
20,000	0.1408	0.1407	0.1473	0.1789	0.0882	0.0857	0.1000	0.0655
50,000	0.2579	0.2550	0.2542	0.3089	0.1491	0.1492	0.1826	0.1420
100,000	0.4689	0.4481	0.4456	0.4927	0.2598	0.2566	0.3182	0.2564
200,000	0.8217	0.7765	0.7710	0.8395	0.4441	0.4590	0.5345	0.4486

減少傾向であったが、Extra Edge は増加傾向となった。前述のとおり、 G^2 検定、 χ^2 検定は、データ数の増加に対して Type I error の発生率が単調減少しないので過学習し Extra Edge が多く発生している。CMI は、earthquake のみ少ないデータ数から SHD が IP と同等の精度となったが、その他のネットワークでは、データ数を増加してもほぼ一定の値を示した。特に child では SHD が他手法に比べ極端に大きな値を取っている。また、それらのネットワークにおいて、CMI の Missing Edge または Extra Edge はデータ数の増加に関係なくほぼ一定の値を示している。これは前述

のとおり、CMI はしきい値に強い影響を受け漸近一致性をもたないためである。

SHD, Missing Edge, Extra Edge の標準偏差の結果より、Bayes factor は、データ数が十分に多いときは単調減少するのに対し、データ数が少ないときは単調減少するとは限らないことが示された。これは、データ数が少ないときは CI テストが誤る確率が高くなり、学習精度が安定しないためである。IP は、cancer, earthquake, survey, sachs, child では、SHD, Missing Edge, Extra Edge の平均値が常に 0.0 に収束したため標準偏差も 0.0 に収束した。ただし alarm

表 8 survey の実験結果と計算時間
Table 8 The experiment results for survey.

SHD								
データ数	Bayes factor (1.0)	Bayes factor(5.0)	Bayes factor(10.0)	Bayes factor(20.0)	G^2	χ^2	CMI	IP
10,000	0.6(0.97)	0.2(0.63)	0.2(0.63)	0.2(0.63)	0.5(1.08)	0.4(0.84)	5.7(0.48)	2.8(2.82)
20,000	0.0(0.00)	0.0(0.00)	0.0(0.00)	0.0(0.00)	0.7(1.06)	1.0(1.23)	5.5(0.53)	0.8(1.87)
50,000	0.0(0.00)	0.0(0.00)	0.0(0.00)	0.0(0.00)	1.5(1.43)	1.5(1.43)	5.7(0.48)	0.0(0.00)
100,000	0.0(0.00)	0.0(0.00)	0.0(0.00)	0.0(0.00)	4.4(0.97)	4.4(0.97)	5.7(0.48)	0.0(0.00)
200,000	0.0(0.00)	0.0(0.00)	0.0(0.00)	0.0(0.00)	4.8(0.79)	4.8(0.79)	6.0(0.00)	0.0(0.00)
Missing Edge								
10,000	0.4(0.52)	0.0(0.00)	0.0(0.00)	0.0(0.00)	0.0(0.00)	0.0(0.00)	3.7(0.48)	0.3(0.48)
20,000	0.0(0.00)	0.0(0.00)	0.0(0.00)	0.0(0.00)	0.0(0.00)	0.0(0.00)	3.8(0.42)	0.0(0.00)
50,000	0.0(0.00)	0.0(0.00)	0.0(0.00)	0.0(0.00)	0.0(0.00)	0.0(0.00)	3.7(0.48)	0.0(0.00)
100,000	0.0(0.00)	0.0(0.00)	0.0(0.00)	0.0(0.00)	0.0(0.00)	0.0(0.00)	3.7(0.48)	0.0(0.00)
200,000	0.0(0.00)	0.0(0.00)	0.0(0.00)	0.0(0.00)	0.0(0.00)	0.0(0.00)	4.0(0.00)	0.0(0.00)
Extra Edge								
10,000	0.0(0.00)	0.0(0.00)	0.0(0.00)	0.0(0.00)	0.2(0.42)	0.1(0.32)	0.0(0.00)	0.0(0.00)
20,000	0.0(0.00)	0.0(0.00)	0.0(0.00)	0.0(0.00)	0.3(0.48)	0.4(0.52)	0.0(0.00)	0.0(0.00)
50,000	0.0(0.00)	0.0(0.00)	0.0(0.00)	0.0(0.00)	0.7(0.68)	0.7(0.68)	0.0(0.00)	0.0(0.00)
100,000	0.0(0.00)	0.0(0.00)	0.0(0.00)	0.0(0.00)	1.9(0.57)	1.9(0.57)	0.0(0.00)	0.0(0.00)
200,000	0.0(0.00)	0.0(0.00)	0.0(0.00)	0.0(0.00)	2.0(0.67)	2.0(0.67)	0.0(0.00)	0.0(0.00)
計算時間 (s)								
10,000	0.0441	0.0521	0.0414	0.0461	0.0361	0.0356	0.0273	0.0431
20,000	0.0617	0.0615	0.0614	0.0619	0.0533	0.0580	0.0373	0.0718
50,000	0.0960	0.0959	0.0957	0.1004	0.1106	0.1084	0.0549	0.1412
100,000	0.1760	0.1820	0.1836	0.1823	0.3356	0.3349	0.1058	0.2797
200,000	0.3155	0.3178	0.3227	0.3227	0.6317	0.6332	0.1538	0.5230

表 9 sachs の実験結果と計算時間
Table 9 The experiment results for sachs.

SHD								
データ数	Bayes factor (1.0)	Bayes factor(5.0)	Bayes factor(10.0)	Bayes factor(20.0)	G^2	χ^2	CMI	IP
10,000	16.2(0.79)	16.6(0.70)	17.3(0.82)	16.3(1.16)	12.5(3.44)	11.4(4.97)	18.3(0.95)	0.0(0.00)
20,000	17.3(0.82)	15.4(1.51)	14.2(1.48)	13.6(1.78)	1.6(1.90)	7.5(1.269)	18.3(0.82)	0.0(0.00)
50,000	14.0(0.00)	14.0(0.00)	12.6(2.95)	9.1(3.38)	7.7(6.58)	8.1(6.44)	17.9(0.88)	0.0(0.00)
100,000	12.6(2.95)	6.3(6.13)	6.3(6.13)	1.4(4.43)	7.7(9.27)	7.9(9.47)	18.4(0.84)	0.0(0.00)
200,000	0.0(0.00)	0.0(0.00)	0.0(0.00)	0.0(0.00)	12.5(11.01)	12.5(11.01)	17.9(0.74)	0.0(0.00)
Missing Edge								
10,000	2.0(0.00)	2.0(0.00)	2.0(0.00)	2.0(0.00)	1.3(0.68)	1.6(0.52)	0.0(0.00)	0.0(0.00)
20,000	2.0(0.00)	2.0(0.00)	2.0(0.00)	1.2(0.42)	1.0(0.00)	1.0(0.00)	0.0(0.00)	0.0(0.00)
50,000	1.0(0.00)	1.0(0.00)	1.0(0.00)	1.0(0.00)	0.3(0.48)	0.0(0.00)	0.0(0.00)	0.0(0.00)
100,000	1.0(0.00)	0.6(0.52)	0.6(0.52)	0.0(0.00)	0.0(0.00)	0.0(0.00)	0.0(0.00)	0.0(0.00)
200,000	0.0(0.00)	0.0(0.00)	0.0(0.00)	0.0(0.00)	0.0(0.00)	0.0(0.00)	0.0(0.00)	0.0(0.00)
Extra Edge								
10,000	2.2(0.79)	2.6(0.70)	3.3(0.82)	2.3(1.16)	0.6(1.27)	0.7(1.27)	4.3(0.95)	0.0(0.00)
20,000	3.3(0.82)	1.4(1.51)	0.5(1.27)	0.2(0.63)	0.0(0.00)	0.0(0.00)	4.3(0.82)	0.0(0.00)
50,000	0.0(0.00)	0.0(0.00)	0.0(0.00)	0.0(0.00)	1.9(2.73)	1.9(2.73)	3.9(0.88)	0.0(0.00)
100,000	0.0(0.00)	0.0(0.00)	0.0(0.00)	0.0(0.00)	2.4(4.30)	2.6(4.50)	4.4(0.84)	0.0(0.00)
200,000	0.0(0.00)	0.0(0.00)	0.0(0.00)	0.0(0.00)	5.9(6.12)	5.9(6.12)	3.9(0.74)	0.0(0.00)
計算時間 (s)								
10,000	1.7460	1.9227	2.0077	2.5766	1.0297	1.1388	1.6797	2.5984
20,000	3.3350	3.3664	3.5497	3.7597	2.0025	2.0696	2.2844	4.0001
50,000	6.6541	7.0688	7.2192	7.7939	2.4137	2.4396	3.9784	8.7195
100,000	13.3620	12.2720	12.3060	15.1930	4.8982	4.8863	7.0577	14.3050
200,000	27.2280	22.8960	23.1190	25.2860	12.5510	12.4760	14.0710	24.5650

では、単調減少する傾向が見られなかった。これは、Bayes factor と同様に、データ数が少ない場合では IP の学習精度が安定しないためである。一方で、 G^2 , χ^2 は八つのネットワークでデータ数に関係なく SHD, Missing Edge, Extra Edge の標準偏差が単調減少しなかった。CMI では標準偏差の単調減少傾向は確認できるものの、平均値が単調減少せず一定の値に収束する傾向が見られた。この理由は、 G^2 , χ^2 , CMI が漸近一致性をもたず精度が不安定であるためである。

計算時間については、child では IP が最も速く学習できたが、その他のネットワークでは CMI が最も速

く学習できた。しかしこれは Missing Edge 数が極端に多く CI テストの実施回数が減少したためと解釈できる。 G^2 , χ^2 でも alarm を除いた七つのネットワークで Bayes factor より計算時間が短かった。alarm では、Bayes factor が G^2 , χ^2 に比べて速く学習できた。これは、Bayes factor が G^2 , χ^2 に比べ Extra Edge 数が少なく、CI テストの実施回数も少なくなるためである。

以上から、従来の独立検定手法を用いた制約ベースアルゴリズムでは、最適な構造を得る保証がなく、従来の厳密解探索アプローチでは大規模ネットワークを

表 10 child の実験結果と計算時間
Table 10 The experiment results for child.

SHD								
データ数	Bayes factor (1.0)	Bayes factor(5.0)	Bayes factor(10.0)	Bayes factor(20.0)	G^2	χ^2	CMI	IP
10,000	35.9(3.52)	36.3(3.13)	36.9(2.80)	37.5(3.22)	45.1(6.31)	43.5(5.85)	63.2(4.87)	0.0(0.00)
50,000	32.2(3.44)	32.0(4.31)	30.8(4.33)	30.4(4.14)	30.0(4.13)	29.5(4.06)	51.8(3.95)	0.0(0.00)
1,000,000	18.5(1.57)	17.7(1.06)	17.5(0.97)	17.5(1.04)	19.2(5.97)	19.0(5.90)	46.0(0.85)	0.0(0.00)
2,000,000	16.8(0.42)	16.7(0.45)	16.4(1.14)	16.2(1.49)	17.7(5.84)	18.3(4.73)	45.9(0.78)	0.0(0.00)
5,000,000	15.3(2.02)	14.6(2.40)	14.3(2.40)	14.0(2.49)	17.4(5.86)	17.4(5.95)	45.6(0.68)	0.0(0.00)
10,000,000	12.0(2.03)	11.3(1.27)	11.2(0.91)	11.2(0.91)	19.2(6.21)	19.1(6.20)	45.3(0.52)	0.0(0.00)
Missing Edge								
10,000	5.4(0.85)	5.4(0.62)	5.0(0.26)	4.9(0.43)	3.1(0.31)	3.1(0.35)	0.0(0.00)	0.0(0.00)
50,000	3.8(0.53)	3.2(0.43)	3.1(0.35)	3.0(0.32)	1.9(0.25)	1.9(0.25)	0.0(0.00)	0.0(0.00)
1,000,000	1.8(0.38)	1.0(0.00)	1.0(0.00)	1.0(0.00)	0.0(0.00)	0.0(0.00)	0.0(0.00)	0.0(0.00)
2,000,000	1.0(0.00)	1.0(0.00)	1.0(0.00)	1.0(0.00)	0.0(0.00)	0.0(0.00)	0.0(0.00)	0.0(0.00)
5,000,000	0.3(0.48)	0.0(0.00)	0.0(0.00)	0.0(0.00)	0.0(0.00)	0.0(0.00)	0.0(0.00)	0.0(0.00)
10,000,000	0.0(0.00)	0.0(0.00)	0.0(0.00)	0.0(0.00)	0.0(0.00)	0.0(0.00)	0.0(0.00)	0.0(0.00)
Extra Edge								
10,000	12.9(3.10)	13.5(2.62)	14.3(2.31)	15.1(2.50)	24.3(5.06)	22.8(4.64)	41.6(4.10)	0.0(0.00)
50,000	11.9(2.87)	12.7(3.32)	12.1(3.88)	12.0(3.64)	12.3(3.15)	11.8(2.98)	30.8(3.18)	0.0(0.00)
1,000,000	5.1(1.01)	5.0(0.85)	4.8(0.86)	4.7(0.94)	5.0(2.87)	4.9(2.86)	25.9(0.74)	0.0(0.00)
2,000,000	3.9(0.57)	3.7(0.45)	3.5(0.68)	3.4(0.81)	4.5(2.85)	4.6(2.65)	25.9(0.94)	0.0(0.00)
5,000,000	2.7(0.83)	2.3(1.21)	2.2(1.23)	1.9(1.40)	4.5(3.17)	4.3(3.09)	25.5(0.57)	0.0(0.00)
10,000,000	0.6(1.22)	0.2(0.76)	0.1(0.55)	0.1(0.55)	5.4(3.61)	5.3(3.61)	25.3(0.60)	0.0(0.00)
計算時間 (s)								
10,000	2.2	3.4	3.5	3.8	4.1	4.6	388.7	1.2
50,000	12.1	13.1	13.5	15.0	11.3	13.0	20.0	13.0
1,000,000	385.8	452.5	444.1	397.2	350.5	323.6	416.8	565.3
2,000,000	919.8	948.8	907.0	874.5	488.9	542.8	825.9	1106.1
5,000,000	3033.0	2887.3	2917.0	2824.1	1461.3	1565.6	2800.1	1803.1
10,000,000	6904.6	6618.0	6599.3	6466.5	3292.8	3647.3	5927.5	2126.4

表 11 alarm の実験結果と計算時間
Table 11 The experiment results for alarm.

SHD								
データ数	Bayes factor (1.0)	Bayes factor(5.0)	Bayes factor(10.0)	Bayes factor(20.0)	G^2	χ^2	CMI	IP
10,000	24.4(2.86)	30.7(3.73)	35.7(4.90)	50.4(6.50)	52.2(6.52)	51.9(7.32)	36.8(5.32)	2.30(1.16)
20,000	26.0(3.66)	31.2(4.24)	35.4(4.06)	46.8(5.47)	53.5(9.44)	54.6(9.08)	35.2(3.29)	2.00(1.25)
2,000,000	17.1(4.18)	17.9(4.52)	19.7(4.50)	21.1(4.31)	53.3(16.35)	52.3(15.58)	31.9(0.31)	-
10,000,000	18.4(5.61)	13.9(5.47)	13.9(5.47)	13.7(5.45)	52.8(11.78)	52.3(11.27)	32.0(0.18)	-
20,000,000	10.1(3.10)	10.7(3.20)	11.4(3.43)	11.9(3.03)	56.2(14.45)	55.9(13.84)	32.0(0.00)	-
Missing Edge								
10,000	3.8(0.61)	3.6(0.62)	3.7(0.70)	3.9(0.31)	4.3(0.79)	3.8(0.81)	3.4(0.56)	1.8(0.42)
20,000	3.4(0.77)	3.9(0.40)	3.9(0.25)	3.8(0.46)	3.0(0.87)	3.2(0.70)	3.2(0.43)	1.0(0.47)
2,000,000	1.0(0.00)	1.0(0.00)	1.0(0.00)	1.0(0.00)	0.4(0.50)	0.1(0.31)	3.0(0.00)	-
10,000,000	1.0(0.18)	0.2(0.38)	0.2(0.38)	0.1(0.31)	0.0(0.00)	0.0(0.00)	3.0(0.00)	-
20,000,000	0.0(0.00)	0.0(0.00)	0.0(0.00)	0.0(0.00)	0.0(0.00)	0.0(0.00)	3.0(0.00)	-
Extra Edge								
10,000	10.1(1.55)	14.9(2.47)	18.8(3.06)	30.4(4.59)	33.2(4.71)	33.5(5.42)	22.6(4.25)	0.3(0.48)
20,000	12.3(2.13)	15.6(2.34)	18.9(2.42)	27.9(3.83)	35.4(7.11)	36.1(6.86)	21.4(2.36)	0.7(0.67)
2,000,000	11.9(2.98)	12.6(3.20)	14.0(3.22)	15.1(3.15)	41.4(13.16)	40.6(12.46)	19.0(0.00)	-
10,000,000	12.7(3.97)	9.8(4.31)	9.8(4.31)	9.8(4.28)	40.7(9.53)	40.2(8.97)	19.0(0.00)	-
20,000,000	7.3(2.39)	7.8(2.48)	8.4(2.48)	8.7(2.27)	43.7(11.88)	43.4(11.30)	19.0(0.00)	-
計算時間 (s)								
10,000	5.2	5.4	5.5	7.2	5.3	5.5	4.4	156.3
20,000	8.5	9.1	9.6	11.6	8.6	9.3	5.9	777.9
2,000,000	1194.8	1158.2	1191.0	1233.0	1561.8	1521.9	412.1	-
10,000,000	12234.0	12777.0	12777.0	14182.0	13934.0	13561.0	3412.3	-
20,000,000	28416.0	28938.0	28340.0	28498.0	34836.0	33122.0	6957.9	-

学習できないという欠点があることが確認できた。一方で、提案した手法では、従来手法が 60 ノードを限界としていたのに対して、200 ノードを超える大規模ネットワーク学習を実行できることが分かった。更に、本手法は従来の独立検定手法と比べ、データ数が増加したときに、高い学習精度をもつことが分かった。

6. む す び

本論文では、ベイジアンネットワークの構造学習に

おいて、Bayes factor を用いた CI テストを RAI アルゴリズムに組み込み、漸近一致性をもち大規模なネットワーク学習を可能にする手法を提案した。シミュレーション実験により、Bayes factor は漸近一致性をもつため、これを適用した RAI アルゴリズムは、厳密解探索アプローチと同様にデータ数の増加により真の条件付き独立構造を推定できることを示した。更に、制約ベースアルゴリズムを採用しているため、厳密解探索アプローチでは学習できないノード数の構造

表 12 win95pts の実験結果と計算時間
Table 12 The experiment results for win95pts.

SHD								
データ数	Bayes factor (1.0)	Bayes factor(5.0)	Bayes factor(10.0)	Bayes factor(20.0)	G^2	χ^2	CMI	IP
10,000	59.9(4.80)	92.2(7.28)	137.4(5.15)	203.9(7.25)	139.2(7.28)	135.8(5.79)	57.3(3.37)	-
20,000	53.8(4.76)	84.5(6.43)	119.2(6.30)	184.0(6.42)	137.6(5.13)	136.0(4.55)	57.7(2.68)	-
500,000	35.6(2.69)	45.3(3.75)	67.3(6.13)	96.8(6.24)	59.5(7.43)	60.4(7.36)	56.1(1.49)	-
1,000,000	33.6(1.79)	40.5(4.11)	58.3(4.68)	83.5(4.42)	60.0(5.57)	60.2(5.79)	56.1(1.11)	-
2,000,000	33.1(1.96)	36.0(2.95)	45.8(5.13)	69.4(4.05)	58.8(5.49)	58.5(6.16)	56.1(1.06)	-
Missing Edge								
10,000	27.0(1.67)	16.6(1.97)	15.8(1.78)	14.4(1.57)	95.1(3.41)	94.7(2.67)	45.1(1.51)	-
20,000	24.3(1.11)	14.5(1.17)	13.4(1.25)	11.7(1.11)	94.1(2.73)	94.3(2.50)	45.6(0.93)	-
500,000	17.9(0.73)	8.6(0.82)	9.3(0.61)	7.8(0.83)	17.1(0.76)	17.0(0.72)	45.2(0.66)	-
1,000,000	17.3(0.45)	8.4(0.56)	9.1(0.37)	7.5(0.82)	17.0(0.56)	17.0(0.63)	45.4(0.68)	-
2,000,000	17.2(0.55)	8.2(0.38)	8.5(0.57)	7.0(0.53)	16.5(0.78)	16.4(0.73)	45.5(0.63)	-
Extra Edge								
10,000	20.8(3.86)	61.7(7.35)	107.4(4.80)	173.6(7.00)	40.8(4.37)	38.2(3.19)	6.9(1.93)	-
20,000	18.0(3.57)	56.0(6.52)	91.2(6.12)	155.8(5.67)	40.8(2.90)	39.5(2.76)	6.9(1.69)	-
500,000	7.7(2.00)	24.5(3.23)	45.2(5.39)	76.0(5.84)	29.3(6.86)	30.1(6.72)	5.8(0.87)	-
1,000,000	6.9(1.48)	20.7(3.56)	37.3(3.98)	63.1(4.12)	29.8(5.14)	29.8(5.20)	5.6(0.49)	-
2,000,000	6.6(1.59)	17.1(2.37)	26.2(4.37)	50.3(3.68)	29.0(5.37)	28.7(5.94)	5.6(0.44)	-
計算時間 (s)								
10,000	18.5	27.9	119.1	1631.6	6.3	6.2	10.1	-
20,000	29.7	38.9	144.2	1853.0	8.0	9.8	13.8	-
500,000	606.3	660.4	826.4	5146.2	388.5	400.9	187.8	-
1,000,000	1301.7	1435.0	1707.6	5426.9	950.7	929.9	371.4	-
2,000,000	3187.2	3486.3	4660.9	9732.7	2312.3	1867.6	786.9	-

表 13 andes の実験結果と計算時間
Table 13 The experiment results for andes.

SHD								
データ数	Bayes factor (1.0)	Bayes factor(5.0)	Bayes factor(10.0)	Bayes factor(20.0)	G^2	χ^2	CMI	IP
10,000	70.4(4.75)	113.8(8.90)	148.9(8.14)	205.9(6.45)	172.7(7.87)	170.0(7.52)	84.3(2.93)	-
20,000	48.7(6.04)	81.8(6.42)	109.9(7.49)	158.8(6.17)	158.9(7.02)	156.2(6.87)	84.9(2.41)	-
500,000	16.1(3.17)	34.9(4.37)	44.9(3.70)	61.1(4.19)	150.4(5.77)	147.9(5.94)	86.2(1.34)	-
1,000,000	13.9(3.53)	28.4(3.50)	37.1(3.79)	51.1(4.30)	148.5(6.61)	147.2(6.75)	86.3(1.34)	-
2,000,000	12.1(3.02)	26.2(3.89)	32.8(3.60)	46.4(3.80)	149.2(8.16)	148.7(7.94)	85.7(0.74)	-
Missing Edge								
10,000	40.5(2.96)	34.4(3.46)	31.8(3.13)	31.2(3.16)	26.1(3.55)	26.0(3.46)	80.3(2.31)	-
20,000	26.3(3.09)	20.8(2.98)	19.3(3.09)	18.0(2.67)	14.9(2.83)	15.0(2.81)	81.5(1.48)	-
500,000	1.6(0.68)	0.7(0.55)	0.4(0.50)	0.4(0.56)	2.1(0.25)	2.1(0.31)	82.8(1.05)	-
1,000,000	1.3(0.48)	0.1(0.31)	0.0(0.00)	0.0(0.00)	2.0(0.00)	2.0(0.00)	83.0(0.96)	-
2,000,000	1.0(0.00)	0.0(0.00)	0.0(0.00)	0.0(0.00)	2.0(0.00)	2.0(0.00)	82.7(0.74)	-
Extra Edge								
10,000	24.0(4.56)	71.5(6.51)	108.4(6.10)	166.1(5.38)	140.0(5.77)	137.3(5.72)	2.1(1.11)	-
20,000	16.1(3.87)	53.4(5.62)	82.2(6.36)	132.1(5.82)	137.2(5.73)	134.4(5.80)	1.9(0.83)	-
500,000	8.3(2.61)	27.4(3.72)	37.7(3.24)	53.6(3.98)	141.6(5.73)	139.0(5.71)	2.3(0.55)	-
1,000,000	7.2(2.90)	22.6(2.77)	31.2(3.15)	44.9(3.90)	139.6(6.62)	138.2(6.80)	2.3(0.55)	-
2,000,000	5.5(2.33)	21.1(3.27)	27.7(2.97)	40.8(3.31)	140.4(8.01)	139.8(7.80)	2.0(0.00)	-
計算時間 (s)								
10,000	58.9	56.8	59.8	81.0	66.4	66.1	36.8	-
20,000	96.0	93.6	97.2	122.7	95.7	85.3	46.6	-
500,000	1668.7	1571.1	1637.0	2292.2	1733.7	1733.6	568.4	-
1,000,000	3934.2	4015.2	4181.2	5895.8	4853.2	4760.9	1055.3	-
2,000,000	11641.0	11060.0	11188.0	14475.0	14005.0	13121.0	2286.6	-

を学習できることを示した。Bayes factor では、一様分布に近いパラメータ設定のネットワーク及び大規模ネットワークの学習において、ESS の値を大きくするほど、Missing Edge を減少でき、Ueno [21] で証明された結果と同様になることを示した。

一方で G^2 検定、 χ^2 検定は、データ数の増加に対して、Type I error の発生率が単調減少しないので、過学習を起こし、漸近的に真の構造を推定できないことが示された。また、CMI を用いた検定は、しきい値に大きく影響を受け漸近一致性がないことから、ほとんどのネットワークで最も精度が悪く、漸近的に真の構造を推定できないことが示された。

今後の課題として、数千以上のノード数をもつベン

チマークネットワーク及び実データを用いて実験を行い、本手法の有意性を検証する。

文 献

- [1] 植野真臣, ベイジアンネットワーク, コロナ社, 2013.
- [2] 本村陽一, “チュートリアル講演 ベイジアンネットワーク,” 信学技報, vol.103, no.228, pp.25–30, 2003.
- [3] D.M. Chickering, “Learning Bayesian networks is NP-Complete,” in Learning from Data: Artificial Intelligence and Statistics, vol.V, pp.121–130, Springer, 1996.
- [4] R.G. Cowell, “Efficient maximum likelihood pedigree reconstruction,” Theoretical Population Biology, vol.76, no.4, pp.285–291, Dec. 2009.
- [5] M. Koivisto and K. Sood, “Exact bayesian structure discovery in bayesian networks,” J. Machine Learning

- Research, vol.5, pp.549–573, Dec. 2004.
- [6] A. Singh and A. Moore, “Finding optimal Bayesian networks by dynamic programming,” Technical Report, Carnegie Mellon University, pp.1–16, June 2005.
- [7] T. Silander and P. Myllymaki, “A simple approach for finding the globally optimal Bayesian network structure,” in *Uncertainty in Artificial Intelligence (UAI)*, pp.445–452, AUAI Press, 2006.
- [8] B. Malone, C. Yuan, and E.A. Hansen, “Memory-efficient dynamic programming for learning optimal bayesian networks,” *Proc. 25th AAAI Conference*, pp.1057–1062, 2011.
- [9] C. Yuan, B. Malone, and W. Xiaojian, “Learning optimal Bayesian networks using A* search,” *International Joint Conference on Artificial Intelligence (IJCAI)*, pp.2186–2191, 2011.
- [10] J. Cussens, “Bayesian network learning with cutting planes,” in *Uncertainty in Artificial Intelligence (UAI)*, pp.153–160, AUAI Press, 2011.
- [11] C. Li and M. Ueno, “An extended depth-first search algorithm for optimal triangulation of Bayesian networks,” *Int. J. Approximate Reasoning*, vol.80, pp.294–312, 2017.
- [12] J. Pearl, *Causality: Models, Reasoning, and Inference*, Cambridge University Press, 2000.
- [13] P. Spirtes, C. Glymour, and R. Scheines, *Causation, Prediction, and Search*, MIT Press, 2000.
- [14] I. Tsamardinos, L.E. Brown, and C.F. Aliferis, “The max-min hill-climbing Bayesian network structure learning algorithm,” *Mach. Learn.*, vol.65, no.1, pp.31–78, 2006.
- [15] R. Yehezkel and B. Lerner, “Bayesian network structure learning by recursive autonomy identification,” *J. Machine Learning Research*, vol.10, pp.1527–1570, 2009.
- [16] S. GM and R. Feinn, “Using effect size—or why the p value is not enough,” *J. Grad. Med. Educ.*, vol.4, no.3, pp.279–282, 2012.
- [17] H. Steck and T.S. Jaakkola, “On the dirichlet prior and Bayesian regularization,” *Neural Information Processing Systems (NIPS 2002)*, pp.697–704, 2002.
- [18] W. Buntine, “Theory Refinement on Bayesian Networks,” in *Uncertainty in Artificial Intelligence (UAI)*, pp.52–60, 1991.
- [19] D. Heckerman, D. Geiger, and D.M. Chickering, “Learning Bayesian networks: The combination of knowledge and statistical data,” *Mach. Learn.*, vol.20, pp.197–243, 1995.
- [20] M. Ueno, “Learning likelihood-equivalence Bayesian networks using an empirical Bayesian approach,” *Behaviormetrika*, vol.35, no.2, pp.115–135, 2008.
- [21] M. Ueno, “Learning networks determined by the ratio of prior and data,” in *Uncertainty in Artificial Intelligence (UAI)*, pp.598–605, 2010.
- [22] M. Ueno, “Robust learning Bayesian networks for prior belief,” in *Uncertainty in Artificial Intelligence (UAI)*, pp.698–707, 2011.
- [23] D. Koller and N. Friedman, *Probabilistic Graphical Models: Principles and Techniques*, MIT Press, 2009.
- [24] J. Cohen, “A power primer,” *Psychological Bulletin*, pp.155–159, 1992.
- [25] R. Sedgewick and K. Wayne, *Algorithms*, 4th ed., Pearson, 2011.
- [26] M. Scutari, “Learning Bayesian networks with the bnlearn R package,” *J. Statistical Software*, vol.35, no.3, pp.1–22, 2011.
- (平成 29 年 11 月 7 日受付, 30 年 1 月 19 日早期公開)



名取 和樹 (学生会員)

2014 年電気通信大学情報理工学部卒。2016 年同大学院情報システム学研究科社会知能情報学専攻博士前期課程修了。同年, 同大学院情報理工学研究科情報・ネットワーク工学専攻博士後期課程入学, 現在に至る。



宇都 雅輝 (正員)

2013 年電気通信大学大学院情報システム学研究科博士後期課程修了。博士(工学)。長岡技術科学大学を経て, 2015 年より電気通信大学助教に就任, 現在に至る。e テスティング, e ラーニング, 人工知能, ベイズ統計, 自然言語処理などの研究に従事。



植野 真臣 (正員)

1992 年神戸大学大学院教育学研究科修了, 1994 年東京工業大学大学院総合理工学研究科修了。博士(工学)。東京工業大学, 千葉大学, 長岡技術科学大学を経て 2006 年より電気通信大学助教授, 2013 年より教授, 現在に至る。