

동일 오차 분산 가우시안 구조방정식모델의 식별가능성 (Identifiability of Gaussian structural equation models with equal error variances)

요약(SUMMARY)

우리는 변수를 부모들의 함수로 사용될 수 있는 구조방정식모델(structural equation model)과 공동적으로 독립이라 가정된 잡음항(noise term)을 고려한다. 각 구조방정식 모델(structural equation model)에 해당하는 것은 변수들 간의 관계를 설명하는 방향성 비순환 그래프(directed acyclic graph, DAG)이다. 선형함수가 있는 가우시안 구조방정식 모델(Gaussian structural equation model)에서 그래프는 충실성(faithfulness)이 가정된 마코프 동등클래스(Markov equivalence classe)까지만 결합분포에서 식별할 수 있다. 본 연구에서는 모든 잡음 변수들의 분산이 동일한 경우 완전한 식별가능성을 입증한다: 방향성 비순환 그래프(directed acyclic graph)는 결합 가우시안 분포(joint Gaussian distribution)로부터 복구할 수 있다. 우리의 결과는 인과적 추론(causal inference)에 직접적인 의미를 지닌다: 만약 데이터가 동일 오차 분산(equal error variance)인 가우시안 구조 방정식 모델(Gaussian structural equation model)을 따른다면, 모든 변수가 관측된다고 가정 할 때, 인과 구조(causal structure)는 오직 관측 데이터에서만 추론 할 수 있다. 우리는 우리의 이론적 발견에 근거한 통계 방법과 알고리즘을 제안한다.

1 소개(INTRODUCTION)

1.1 그래피컬과 구조방정식 모델(Graphical and structural equation models)

확률변수들이 X_1, \dots, X_p 인 경우, 우리는 그래피컬 모델(graphical model)을 쌍 $\{\mathcal{G}, \mathcal{L}(\mathcal{X})\}$ 으로 정의한다. 여기서 $\mathcal{L}(\mathcal{X}) = \mathcal{L}(X_1, \dots, X_p)$ 는 방향성 비순환 그래프(directed acyclic graph) \mathcal{G} 와 관련한 마코프(Markov)의 결합 확률분포(joint probability distribution)이다 (Lauritzen, 1996, Ch. 3.2). 구조방정식 모델(Structural equation model)은 기능 모델(functional model)이라고도 하며, 그래피컬 모델(graphical model)과 관련이 있다. 그들은 p 방정식들의 집합 $S = \{S_1, \dots, S_p\}$ 과 잡음 변수들(noise variable)의 결합분포 $\mathcal{L}(\mathcal{N}) = \mathcal{L}(\mathcal{N}_1, \dots, \mathcal{N}_p)$ 에 의해 지정된다.

$$S_j : X_j = f_j(X_{PA}, N_j) \quad (j = 1, \dots, p) \cdots \cdots (1)$$

여기서, $PA_j \subset \{1, \dots, p\} \setminus \{j\}$ 는 j 의 부모를 나타낸다. 우리는 잡음항들(noise terms)이 공동적으로 독립(jointly independent)일 것을 필요로하므로, $\mathcal{L}(\mathcal{N})$ 은 생산분포(product distribution)이다. 구조방정식 모델(structural equation model)의 그래프 \mathcal{G} 는 (1)의 우측에서 X_j 까지 발생하는 각 변수 $X_k (k \in PA_j)$ 로부터 방향성 엣지들(directed edges)을 그려서 얻는다. 그래프 \mathcal{G} 는 비순환(acyclic)이어야 한다. 또한, 구조방정식 모델(structural equation model)이 주어졌을 때, 결합분포(joint distribution) $\mathcal{L}(\mathcal{X})$ 는 완전히 결정되고, $\mathcal{L}(\mathcal{X})$ 는 그래프 \mathcal{G} 에 대한 마코프(Markov)이다 (Pearl, 2009, Theorem 1.4.1).

1.2 분포의 식별 가능성(Identifiability from the distribution)

우리는 다음의 표현을 해결한다. 그래피컬 모델(graphical model) 또는 방향성 비순환 그래프(directed acyclic graph) \mathcal{G}_0 의 구조방정식 모델(structural equation model)에서 주어진 결합분포 $\mathcal{L}(\mathcal{X}) = \mathcal{L}(X_1, \dots, X_p)$ 가 주어졌을때, 우리는 그래프 \mathcal{G}_0 를 복구 할 수 있을까? 먼저 그래피컬모델들(graphical models)을 고려해보면, 답이 부정적이라는 것을 쉽게 알 수 있다: 결합분포(joint distribution) $\mathcal{L}(\mathcal{X})$ 는 모든 완전히 연결된 방향성 비순환 그래프들(directed acyclic graphs)과 같이 다른 방향성 비순환

그래프들(directed acyclic graphs)에 대한 마코프(Markov)이다. 따라서 동일한 분포 $\mathcal{L}(\mathcal{X})$ 에 대한 가능한 그래피컬 모델들(graphical models) $\{\mathcal{G}, \mathcal{L}(\mathcal{X})\}$ 이 많다. 마찬가지로, 분포 $\mathcal{L}(\mathcal{X})$ 를 생성할 수 있는 구조가 다른 구조방정식 모델들(structural equation models)이 있다. 추가적인 가정을 함으로써, 제한된 그래피컬 모델들(restricted graphical models)과 제한된 구조방정식 모델들(restricted structural equation models)을 얻을 수 있는데, 이 모델은 결합분포(joint distribution)로부터 그래프를 식별할 수 있다. 그래피컬 모델들(graphical models)과 기능 모델들(functional models)간의 차이가 분명해지는 것은 정확하게 바로 여기에 있다.

그래피컬 모델(graphical model)이 주어지면, $\mathcal{L}(\mathcal{X})$ 에서 발견된 각 조건부 독립(conditional independence)이 마코프 조건(Markov condition)에 의해 암시되면 분포 $\mathcal{L}(\mathcal{X})$ 는 방향성 비순환 그래프(directed acyclic graph) \mathcal{G}_0 에 대해 충실(faithful)하다. 만약 충실성(faithfulness)이 유지되면, 참인 방향성 비순환 그래프(true directed acyclic graph) \mathcal{G}_0 의 마코프 동등 그래프(Markov equivalence graph)를 얻을 수 있다 (Spirtes et al., 2000). 그러나 마코프 동등 클래스는(Markov equivalence class)는 여전히 클 수 있으며(cf. Andersson et al., 1997), 방향성 비순환 그래프(directed acyclic graph) \mathcal{G}_0 는 식별되지 않는다. 또한, 전체적으로 일반적인 충실성(faithfulness)은 데이터에서 시험 할 수 없다 (Zhang & Spirtes, 2008). 마코프 조건(Markov condition)과 충실성(faithfulness) 모두 결합 분포(joint distribution)의 조건부 독립(conditional independence)만을 제한하기 때문에, 동일한 조건부 독립(conditional independence)을 수반하는 두개의 그래프를 서로 구별할 수 없다는 것은 놀라운 일이 아니다.

구조방정식 모델(Structural equation model)을 통해 다른 유형의 제한을 활용할 수 있다. 먼저, 일반적인 가우시안 구조방정식 모델(general Gaussian structural equation model)은 가우시안 그래피컬 모델(Gaussian graphical model) $\{\mathcal{G}_0, \mathcal{L}(\mathcal{X})\}$ 과 같으므로, 구조 \mathcal{G}_0 는 $\mathcal{L}(\mathcal{X})$ 로부터 식별되지 않는다. 하지만, 최근에 다음과 같은 경우는 예외적이라는 것이 밝혀졌다: (i) 만약 선형 함수들(linear functions)과 non-가우시안(non-Gaussian) 잡음을 고려할 경우, 근본적인 방향성 비순환 그래프(underlying directed acyclic graph) \mathcal{G}_0 를 식별할 수 있습니다. (Shimizu et al., 2006); (ii) 함수를 잡음요소에 대해 추가적으로 제한하면, 선형 가우시안 경우와 기타 몇가지의 병리학적 함수-노이즈 조합(pathological function-noise combinations)을 배제하는 경우, \mathcal{G}_0 은 $\mathcal{L}(\mathcal{X})$ 로부터 식별 가능함을 나타낼 수 있다 (Hoyer et al., 2009; Peters et al., 2011). 이 글에서, 우리는 일반적인 선형 가우시안을 벗어날 세 번째 방법이 있음을 증명한다: (iii) 모든 함수들이 선형이지만 동일 분산 σ^2 을 갖는 정규분포 잡음 변수인 가우시안 구조방정식 모델(Gaussian structural equation models)은 다시 식별이 가능하다. 식별가능성 결과 (i)와 (ii)는 인과 최소성(causal minimality)이라고 불리는 조건을 요구한다. 원래 형태에서, Zhang & Spirtes (2008)는 인과 최소성(causal minimality)을 다음과 같이 정의한다: 참인 원인 그래프(true causal graph) \mathcal{G}_0 의 경우, $\mathcal{L}(\mathcal{X})$ 는 \mathcal{G}_0 의 로스 적절한 부분그래프와 관련하여 마코프(Markov)가 아니다. 따라서 인과 최소성(causal minimality)은 충실성(faithfulness)의 약한 형태이다. Remark2는 (iii)의 증명하기 위해 인과 최소성(causal minimality)을 가정함을 보여준다.

가우시안 구조방정식 모델(Gaussian structural equation models)의 클래스에서 근본적인 방향성 비순환 그래프(underlying directed acyclic graph)가 식별된다는 것은 놀라운 일이 될 수 있다. 동일 오차 분산들의 가정은 유사한 도메인의 변수를 사용하는 실제적용에 대해 자연스러운 것으로 보이며, 일반적으로 시계열 모델에서 사용된다.

1.3 원인 해석(Causal interpretation)

우리의 결과는 인과적 추론에 영향을 미친다. 만약 \mathcal{G}_0 가 X_1, \dots, X_p 에 대한 데이터 생성 프로세스(data-generating process)의 인과 그래프(causal graph)로 해석될 경우, 여기서 고려되는 문제는 결합 분포(joint distribution)로부터 인과 구조(causal structure)를 추론하는 것이다. 이것은 인과 그래프(causal graph)가 관심이 있을 때 특히 흥미롭지만 개입 실험(interventional experiments)은 너무 비싸고, 비윤리적이고, 심지어 수행하기도 불가능하다. 인과 설정에서, 우리의 결과는 다음과 같다. 만약

관측 데이터(observational data)가 인과 관계(causal relationships)를 나타내고 동일 오차 분산을 갖는 가우시안 구조 방정식 모델(Gaussian structural equation model)에 의해 생성된다면, 원인 그래프(causal graph)는 결합 분포(joint distribution)로부터 식별가능하다. 인과 추론(causal inference)에 잠재적으로 중요한 적용에도 불구하고, 우리는 인과관계 용어를 사용하지 않고 주요 진술과 그 증거를 제시한다; 특히, 방정식 (1)과 (2)는 분포에서 holding으로 해석될 수 있다.

2 동일 오차 분산 가우시안 모델의 식별성 (IDENTIFIABILITY FOR GAUSSIAN MODELS WITH EQUAL ERROR VARIANCES)

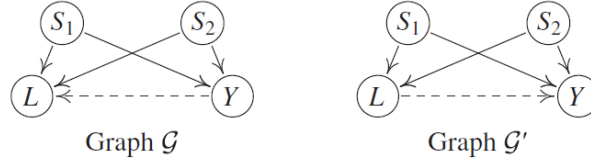


Figure 1: The situation dealt with in the second part of case (ii) in the proof of Theorem 1, with $S = S_1, S_2$ and $\mathcal{D} = \emptyset$. It contains the proof's main argument.

먼저 몇가지 표기법을 소개한다. 인덱스 집합(index set) $\mathcal{J} = \{1, \dots, p\}$ 는 그래프의 점들(vertices) 집합에 해당한다. $j \in \mathcal{J}$ 와 관련된 것은 $\mathcal{X} = (X_1, \dots, X_p)$ 의 임의의 확률변수 X_j 이다. 방향성 비순환 그래프(directed acyclic graph) \mathcal{G} 가 주어졌을 때, 우리는 노드 j 의 부모(parents)를 $PA_j^{\mathcal{G}}$, 자식(children)을 $CH_j^{\mathcal{G}}$, 자손(descendants)을 $DE_j^{\mathcal{G}}$, 그리고 비자손(nondescendants)을 $ND_j^{\mathcal{G}}$ 로 나타낸다.

우리는 아래 형식의 방향성 비순환 그래프(directed acyclic graph) \mathcal{G}_0 를 가진 구조방정식 모델(structural equation model)을 고려한다

$$X_j = \sum_{k \in PA_j^{\mathcal{G}_0}} \beta_{jk} X_k + N_j \quad (j = 1, \dots, p), \quad (1)$$

모든 N_j 가 독립이고 $N(0, \sigma^2)$, $\sigma^2 > 0$ 인 동일한 분포를 따른다. 또한, 각 $j \in \{1, \dots, p\}$ 에 대해, 우리는 모든 $k \in PA_j^{\mathcal{G}_0}$ 에 대해 $\beta_{jk} \neq 0$ 이라는 관계를 필요로 한다.

THEOREM 1. 모델(2)에서 $\mathcal{L}(\mathcal{X})$ 를 생성되었다고 하자. 그러면 \mathcal{G}_0 는 $\mathcal{L}(\mathcal{X})$ 로부터 식별되고 계수 β_{jk} 는 모든 j 와 $k \in PA_j^{\mathcal{G}_0}$ 에 대해 재구성 될 수 있다.

Remark 1. 증명에대한 아이디어는 동일한 결합 분포(joint distribution)에서 유도한 별개의 그래프들 \mathcal{G} 와 \mathcal{G}' 를 가진 두개의 구조방정식 모델(structural equation model)이 있다고 가정하는 것이다. 우리는 Figure 1에서 보면 \mathcal{G} 와 \mathcal{G}' 에서 방향이 반대인 엣지(edges)를 가진 두 그래프에서 같은 부모 집합 $\mathcal{S} = \{S_1, S_2\}$ 을 가진 변수 L 와 Y 를 찾으려면 마코프 조건(Markov condition)과 인과 최소성(causal minimality)을 이용한다.(see Remark 2) 뒤의 내용을 정의함으로써 $L^* = L|_{\mathcal{S}=s}$ 의 몇 개의 값 $s \in \mathbb{R}^2$, 우리는 동일 오차 분산을 사용하여 L^* 가 두 그래프에서 다른 분산을 갖는 것을 보일 수 있다. 이것은 모순으로 이어진다.

Remark 2. Theorem 1은 계수 $\beta_{jk} \neq 0$ 가 어떤 $k \in PA_j^{\mathcal{G}_0}$ 에 대해 사라지지 않는다고 가정한다. Lemma A4 below and Proposition 2 in Peters et al. (2011)는 이 조건이 인과 최소성(causal minimality)을 의미함을 보여준다. 우리의 관점에서, 인과 최소성(causal minimality)은 변수들 사이의 인과

관계(causal influence)에 대한 직관적인 이해에 따르는 자연적인 조건이다.

Remark 3. 오차 공분산 행렬이 미리 정해진 $\alpha_1, \dots, \alpha_p$ 와 알려지지 않은 σ^2 인 $\text{cov}(N_1, \dots, N_p) = \sigma^2 \text{diag}(\alpha_1, \dots, \alpha_p)$ 형태를 가진 경우로 Theorem 1은 다음과 같은 케이스로 일반화 될 수 있다.

3 별점 최대 우도 추정량 (PENALIZED MAXIMUM LIKELIHOOD ESTIMATOR)

참인 계수 β_{jk}^0 의 모델 (2)에서 독립이고 동일한 분포를 갖는 $X^{(1)}, \dots, X^{(n)}$ 인 실현(realizations) 데이터를 고려한다. 벡터 형태의 표현은 $X = BX + N$ 이다, 여기서 B 는 $B_{jk} = \beta_{jk}$ 항목을 가진 $p \times p$ 행렬이다. 이 글을 읽기 쉽게 하기 위해, 각각 모수의 행렬 또는 벡터를 B 또는 β 로 쓴다. 계수 $B^0 = (\beta_{jk}^0)_{j,k}$ 와 오차 분산(error variance) σ^2 의 추정량으로 다음을 고려한다

$$\{\hat{\beta}(\lambda), \hat{\sigma}^2(\lambda)\} = \arg \min_{\beta \in \mathcal{B}, \sigma^2 \in \mathbb{R}^+} -\ell(\beta, \sigma^2; X^{(1)}, \dots, X^{(n)}) + \lambda \|\beta\|_0, \quad (2)$$

where

$$-\ell(\beta, \sigma^2; X^{(1)}, \dots, X^{(n)}) = \frac{np}{2} \log(2\pi\sigma^2) + \frac{n}{2\sigma^2} \text{tr}\{(I - B)^T(I - B)\hat{\Sigma}\},$$

표본 공분산 행렬(sample covariance matrix) $\hat{\Sigma}$ 은 동일 오차 분산(equal error variances) σ^2 을 가정한 음의 로그우도(negative loglikelihood) 그리고 $\|\beta\|_0 = |\{(j, k) : \beta_{jk} \neq 0\}|$. 또한, $\mathcal{B} = \{B \in \mathbb{R}^{p \times p} : \text{Adj}(B) \text{는 오직 } 0 \text{인 고유값(eigenvalues)을 가진다.}\}$ 는 주기(cycles)가 없는 그래프의 계수행렬(coefficient matrices)만 포함한다 (Cvetkovic et al., 1995, p. 81). 여기서, $\text{Adj}(B)_{kj} = 1_{\beta_{kj} \neq 0}$ 는 인접행렬(adjacency matrix)이다. 모든 $\beta \in \mathcal{B}$ 에 대한 최소화는 모든 방향성 비순환 그래프에 대한 최적화를 포함한다. (§4를 보.) 유도된 방향성 비순환 그래프에서 유도된 $\hat{\beta}(\lambda)$ 은 $\hat{\mathcal{G}}$ 으로 나타낸다. $\lambda = \log(n)/2$ 인 경우, 방정식 (3)의 목표 함수(objective function)는 BIC의 점수이다.

참인 계수 β_{jk}^0 와 참인 구조 \mathcal{G}_0 에 대한 별점 최대우도 추정량(penalized maximum likelihood estimator)의 수렴속도(convergence rate)와 일치성(consistency)은 정규성조건(regularity conditions)하에서 van de Geer & Bühlmann (2013, Theorem 5.1)의 분석을 따른다. 더 정확하게, $\lambda_n = \log(n)/2$ 인 경우 우리는 갖는다, $n \rightarrow \infty$ 일 때,

$$\sum_{j,k=1}^p \{\hat{\beta}_{jk}(\lambda_n) - \beta_{jk}^0\}^2 = O_p\{\log(n)n^{-1}\}, \quad \text{pr}(\hat{\mathcal{G}}_n = \mathcal{G}_0) \rightarrow 1.$$

van de Geer & Bühlmann (2013, § 5)의 결과는 $p = p_n = O\{n/\log(n)\}$ 인 고차원 스파스 설정(high-dimensional sparse setting)에도 적용된다.

방정식(3)에서 PC-algorithm과 별점 최대우도의 최소화(penalized likelihood)를 결합하여 사용할 수 있다: 계산적으로 매우 효율적인 전자는, 마코프 동등 클래스(Markov equivalence class)를 추정하는데 사용할 수 있고, 후자는 남은 방향성이 없는 엣지(undirected edges)들을 지정하는데 사용할 수 있다. Tillman et al이 관련된 접근을 제안했다.(2010). 첫번째 단계의 일치성(consistency)을 위해, 반드시 강력한 충실도(strong faithfulness) 가정이 필요하며, 이는 매우 제한적일 수 있다 (Uhler et al., 2013). 별점 최대우도 추정(Penalized maximum likelihood estimation)은 그러한 가정이 필요하지 않지만(van de Geer & Bühlmann, 2013), 계산상 복잡도(computational complexity) 측면에서 비용을 지불한다.

4 욕심쟁이 조사 알고리즘(GREEDY SEARCH ALGORITHM)

(3)의 최적화(optimization)는 모든 방향성 비순환 그래프(directed acyclic graphs)의 공간을 넘어서기 때문에 추정량은 계산하기 어렵다. $p = 20$ 인 경우 이미 $2 \cdot 3 \times 10^{72}$ 개의 방향성 비순환 그래프들(directed acyclic graphs)이 있어(OEIS Foundation Inc., 2011), 철저한 조사를 실행할 수 없다. 대신 우리는 욕심쟁이(greedy) 절차를 제안하는데, 우리는 이것을 동일 오차 분산을 갖는 욕심쟁이 방향성 비순환 그래프 조사(greedy directed acyclic graph search)라고 부른다. 각 t 반복에서, 우리는 방향성 비순환 그래프(directed acyclic graph) \mathcal{G}_t 를 받고, BIC점수에서 가장 크게 하락하는 인접한 방향성 비순환 그래프(directed acyclic graph)로 이동한다. 만약 모든 이웃들이 \mathcal{G}_t 보다 (3)에서 더 높은 BIC 점수를 가지고 있다면, 알고리즘은 종료된다. 여기서, 우리는 두 방향성 비순환 그래프(directed acyclic graph)가 하나의 엣지를 추가, 제거, 반전하여 서로 변형할 수 있다면 이웃이라고 말한다. Chickering (2002)은 비슷한 조사 전략을 제안 했지만, 방향성 비순환 그래프(directed acyclic graphs)보다는 마코프 동등 클래스(Markov equivalence classes) 공간에서 조사를 수행했다.

실행시간을 줄이기 위해, 우리는 \mathcal{G}_t 보다 나은 점수를 가진 방향성 비순환 그래프(directed acyclic graph)를 찾을 때까지 이웃된 방향성 비순환 그래프(neighbouring directed acyclic graphs)를 통해 무작위로 조사하고, 그리고 나면 우리는 \mathcal{G}_{t+1} 에 대해서 이를 방향성 비순환 그래프(directed acyclic graph)로 사용한다. 우리는 적어도 k 이웃들을 고려한다; 만약 \mathcal{G}_t 보다 더 좋은 점수를 갖는 첫번째 k 중에 몇개의 방향성 비순환 그래프(directed acyclic graph)가 있다면, 우리는 가장 좋은 그래프를 선택한다. 전체 절차는 높은 분산을 갖는 잔차를 갖는 노드들을 가르키는 엣지를 바꿀 확률을 증가시키면 더욱 개선된다. 이 수정(modification)과 점수 함수(score function)는 알고리즘에서 동일 오차 분산을 사용하는 유일한 부분이다. 추가적으로, 우리는 다시 $k = p$, $k = 2p$, $k = 3p$, $k = 5p$ 및 $k = 300$ 일때 랜덤 드문 그래프(random sparse graph)에서 다섯 번 다시 이 방법을 시행한다. 이 선택은 특별하지만 실제적으로 잘 작동하는데, 이것이 로컬 최적(local optimum)에 빠질 위험을 감소시키기 때문이다. 이방법의 R코드는 보조 자료에서 확인할 수 있다.

5 실험(EXPERIMENTS)

5.1 기존 방법(Existing methods)

우리는 우리의 방법을 PC-algorithm(Spirites et al., 2000)과 욕심쟁이 동등 조사(greedy equivalence search) (Chickering, 2002)와 비교한다. 후자는 제한되지 않은 가우시안 구조방정식모델(nonrestricted Gaussian structural equation models)에 BIC-정규화된(regularized) 최대우도추정량을 근사시킨다. 두 방법 모두 마코프 동등 클래스(Markov equivalence class)만 복구 할 수 있고, (see §1.2), 그래서 화살표는 방향성 없이 남겨둔다. 마코프 동등 클래스(Markov equivalence class)는 완전히 부분적으로 방향성 비순환 그래프(directed acyclic graph)로 나타낼 수 있다. 실험적으로, 우리는 참과 추정된 부분적인 방향성 비순환 그래프(estimated partially directed acyclic graphs) 사이의 구조적 해밍 거리(Hamming distance)를 보고한다; 예를들어 참이 \rightarrow 이고, 추정된 그래프가 \leftarrow 와 같이 상반된 엣지 쌍에 대해 2의 거리를 할당한다; 모든 다른 엣지 오류는 1로 계산된다.

5.2 랜덤그래프(Random graphs)

n 과 p 가 달라지는 경우, 우리는 세가지의 방법을 비교한다. p 가 주어졌을 때, 우리는 균일분포에서 변수의 순서를 랜덤하게 선택하고, p_{edge} 의 확률로 $p(p-1)/2$ 의 가능한 엣지들을 포함한다. 모든 잡음 분산들은 1로 설정된다, 모든 잡음 변수들은 공통요인으로 스케일링한 결과가 정확히 똑같은 추정치 $\hat{\beta}$ 과 $\hat{\mathcal{G}}$ 를 만들기 때문이다. 계수 β_{jk}^0 들은 균일하게 $[-1, -0.1] \cup [0.1, 1]$ 에서 선택한다. 우리는 기대되는 엣지의 갯수가 $3p/4$ 인 $p_{edge} = 3/(2p-2)$ 을 가진 드문 설정(sparse setting)과 $p_{edge} = 0.3$

인 조밀한 경우(dense setting)을 고려한다. Table 1은 드문 설정(sparse setting)인 경우 100회이상의 시뮬레이션에서 참인 방향성 비순환 그래프(true directed acyclic graph)와 참인 부분 방향성 비순환 그래프(true completed partially directed acyclic graph)의 구조적 해밍 거리(Hamming distance)의 평균을 보여준다. $p = 40$ 이고 $n = 100$ 인 경우를 제외하고, 제안한 방법으로 추정된 그래프들은 참인 마코프 동등 클래스(true Markov equivalence class)만 복구할 수 있는 그래프들보다 참인 방향성 비순환 그래프(true directed acyclic graph)에 가깝다; 또한 욕심쟁이 방향성 비순환 그래프 조사(greedy directed acyclic graph search)는 참인 참인 부분 방향성 그래프(true completed partially directed acyclic graph)의 거리와 비교하였을때 수행이 더 잘된다. Table 2는 동일 오차 분산인 욕심쟁이 방향성 비순환 그래프 조사(greedy directed acyclic graph search)를 통해 얻은 향상이 훨씬 큰 조밀한 설정(dense setting)인 경우 유사한 결과를 보여준다.

개념의 증명으로, 우리는 또한 non-faithful distribution: $X_1 = N_1$, $X_2 = -X_1 + N_2$, $X_3 = X_1 + X_2 + N_3$ 로 부터 얻은 $n = 500$ 인 데이터로 시뮬레이션한다. 이론에서 기대한 대로, PC-algorithm과 욕심쟁이 동등 조사(greedy equivalence search)는 실패한다; 모든 100번의 실험에서 두 알고리즘은 올바른 마코프 동등 클래스(Markov equivalence class)가 아닌 $X_1 \rightarrow X_2 \leftarrow X_3$ 을 출력한다. 반대로, 욕심쟁이 방향성 비순환 그래프 조사(greedy directed acyclic graph search)는 항상 올바른 방향성 비순환 그래프(directed acyclic graph)을 식별한다.

5.3 동일 오차 분산의 편차(Deviation from equal error variances)

Table 1. Average structural Hamming distance between the estimated and true directed acyclic graphs and between the estimated and true Markov equivalence classes, for sparse graphs with p nodes and sample size n

p		$n = 100$			$n = 500$			$n = 1000$		
		GDS _{EEV}	PC	GES	GDS _{EEV}	PC	GES	GDS _{EEV}	PC	GES
5	DAG	1.5	3.9	3.6	0.5	2.9	2.8	0.4	3.0	2.5
	CPDAG	1.5	2.9	2.3	0.5	1.4	1.2	0.3	1.0	0.7
20	DAG	12.2	14.1	18.0	4.5	11.1	10.3	2.7	10.1	8.7
	CPDAG	13.9	10.9	17.0	5.2	7.7	7.6	3.0	6.9	5.6
40	DAG	44.7	29.6	53.0	15.7	22.6	26.1	10.7	20.1	21.9
	CPDAG	50.0	24.4	53.1	18.9	15.9	23.4	13.4	13.3	17.5

DAG, directed acyclic graph; CPDAG, completed partially directed acyclic graph; GDS_{EEV}, greedy directed acyclic graph search with equal error variances; PC, PC-algorithm; GES, greedy equivalence search.

Table 2. Average structural Hamming distance between the estimated and true directed acyclic graphs and between the estimated and true Markov equivalence classes, for dense graphs with p nodes and sample size n

p		$n = 100$			$n = 500$			$n = 1000$		
		GDS _{EEV}	PC	GES	GDS _{EEV}	PC	GES	GDS _{EEV}	PC	GES
5	DAG	1.2	2.9	3.0	0.6	2.4	2.2	0.3	2.1	2.1
	CPDAG	1.3	2.1	1.9	0.5	1.2	0.7	0.2	0.8	0.5
20	DAG	30.0	56.6	63.9	12.5	55.7	66.3	8.2	57.6	69.1
	CPDAG	31.0	56.1	63.2	13.1	55.5	66.2	8.8	57.5	68.5
40	DAG	216.1	242.8	323.1	185.2	247.2	430.4	172.0	248.9	470.6
	CPDAG	217.1	242.4	323.0	185.7	247.0	430.1	172.2	248.5	470.4

DAG, directed acyclic graph; CPDAG, completed partially directed acyclic graph; GDS_{EEV}, greedy directed acyclic graph search with equal error variances; PC, PC-algorithm; GES, greedy equivalence search.

데이터가 다른 오차 분산들을 가진 가우시안 구조방정식 모델(Gaussian structural equation model)로 생성되었을 때, 이 방법은 올바른 방향성 비순환 그래프(directed acyclic graph) 또는 마코프 동등 클래스(Markov equivalence class)를 찾는다고 보장할 수 없다. 실제 생성된 데이터 프로세스가 분산이 다른 가우시안 구조방정식 모델(Gaussian structural equation model)을 따를 때, 우리는 항상 변수를 fine-tuned rescaling을 적용하여 동일 오차 분산을 갖는 모델로 표현할 수 있다, $X_i \mapsto a_i X_i$, 여기서 a_i 는 i th 구조방정식(structural equation)의 오차의 표준화 편차 역과 동일한 값이다. 물론, rescaling은 오직 우리가 오차 분산을 알고있을 때만 가능하므로 the term fine-tuned이다. 데이터가 deceptive fine-tuned standardization와 함께 scale된 가상의 경우, 우리의 방법으로 식별된 그래프는 올바른 마코프 동등 클래스(Markov equivalence class)에 속할 것이다. 그러나, 우리는 다른 오차 분산을 갖는 가우시안 구조방정식 모델(Gaussian structural equation model)의 원데이터를 가진 상황과 다른 인공의 시나리오를 위한 것임을 강조한다.

중요한 질문은 우리 방법이 동일 오차 분산의 가정으로 부터 벗어나면 얼마나 민감한지이다. 우리는 이것을 경험적으로 조사한다. $p = 10$ 이고 $n = 500$ 에서, 우리는 잡음 변수를 $[1 - a, 1 + a]$ 부터 균일하게 추출하고, a 는 0과 0.9사이에서 달라진다. Theorem 1은 오직 $a=0$ 에서 그래프의 식별가능성을 보장한다. 전과 같이 계수 β_{jk}^0 은 $[-1, -0.1] \cup [0.1, 1]$ 로 부터 균일하게 선택한다. 모수 p_{edge} 는 평균이 p 앳지가 되는 $2/(p-1)$ 을 선택한다; 이는 드문과 조밀한 설정 사이 이다. Figure 2는 모수 a 가 변할 때 욕심쟁이 방향성 비순환 그래프 조사(greedy directed acyclic graph search)의 성능은 상대적으로 robust한 것을 보여준다. a 의 값이 크면 이 방법은 PC-algorithm보다 나쁘지 않다. The bestscore method는 욕심쟁이 방향성 비순환 그래프 조사(greedy directed acyclic graph search) 또는 욕심쟁이 동등 조사(greedy equivalence search) 중에 더 좋은 점수를 얻는 방법에 의존하여 결과를 얻는다. 욕심쟁이 방향성 비순환 그래프 조사(Greedy directed acyclic graph search)는 $a = 0, 0.1, \dots, 0.9$ 에서 각각 100%, 100%, 88%, 36%, 7%, 1%, 2%, 0%, 0% and 0%이 선택되었다.

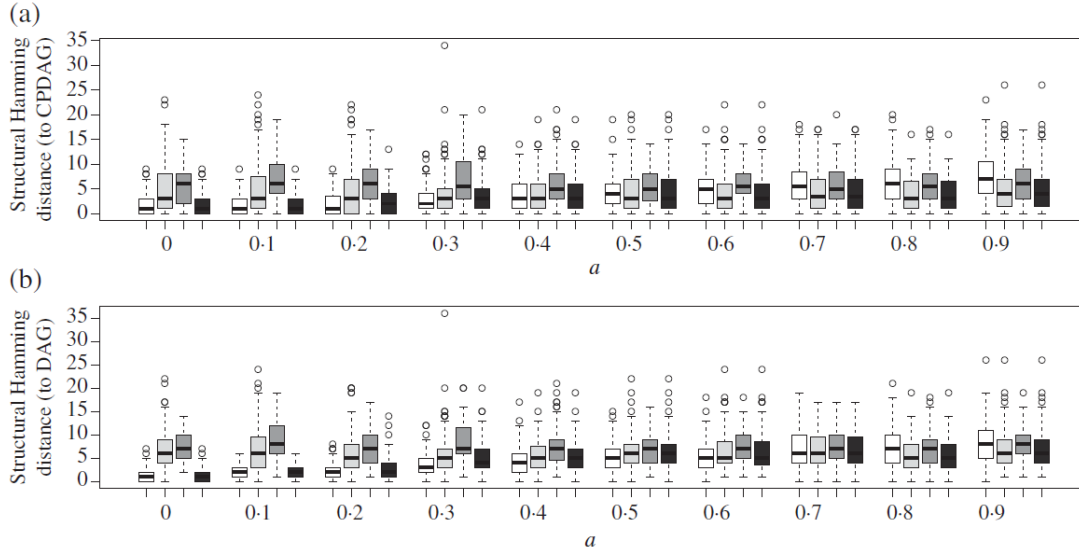


Fig. 2. Box plots for the structural Hamming distance of greedy directed acyclic graph search (white), greedy equivalence search (light grey), the PC-algorithm (grey) and a best-score method (dark grey) to: (a) the true directed acyclic graph, DAG; and (b) the true partially directed acyclic graph, CPDAG. The horizontal axis shows values of a parameter a which measures the perturbation from the case of equal error variances; $a = 0$ corresponds to equal error variances.

Table 3. BIC scores of greedy equivalence search and greedy directed acyclic graph search with equal error variances obtained for different types of microarray data; smaller is better

	Prostate	Lymphoma	Riboflavin	Leukaemia	Brain	Cancer	Colon
GES	4095	4560	2711	5456	1411	5891	3224
GDS _{EEV}	6057	5404	3236	5481	1343	6288	3201

GES, greedy equivalence search; GDS_{EEV}, greedy directed acyclic graph search with equal error variances.

5.4 실제 데이터(Real data)

우리는 이제 욕심쟁이 동등 조사(greedy equivalence search)와 욕심쟁이 방향성 비순환 그래프 조사(greedy directed acyclic graph search)을 microarray data가 포함 된 7개의 데이터 셋들에 적용하고 두 방법을 BIC 점수로 비교한다, Dettling & Buhlmann (2003) and Buhlmann et al. (2014). 욕심쟁이 동등 조사(greedy equivalence search)가 더 나은 점수를 줄 때는 동일 오차 분산 가정이 만족되지 않았을 때를 나타낸다. Fig. 2에서 우리는 가끔 욕심쟁이 방향성 비순환 그래프 조사 (greedy directed acyclic graph search)을 보는 것이 유용할 수 있다는 것을 알 수 있다. 반면에, 욕심쟁이 방향성 비순환 그래프 조사(greedy directed acyclic graph search)가 욕심쟁이 동등 조사(greedy equivalence search)보다 더 좋은 점수를 준다면 우리는 욕심쟁이 방향성 비순환 그래프 조사(greedy directed acyclic graph search)로 얻은 답을 선호한다, 이것은 마코프 동등 클래스(Markov equivalence class)이기 보다는 그래프이기 때문이다. $p > n$ 인 고차원(high-dimensional) 설정을 피하기 위해, 우리는 항상 분산이 높은 $0.8n$ 유전자들을 선택했다. Table3을 보면 7개의 데이터들 중 2개에서 욕심쟁이 동등 조사(greedy equivalence search)보다 욕심쟁이 방향성 비순환 그래프 조사(greedy directed acyclic graph search)가 더 좋은 점수를 제공한다는 것을 보여준다.

Colon의 예제에서, 욕심쟁이 방향성 비순환 그래프 조사 (greedy directed acyclic graph search)는 192 엣지들을 가진 방향성 비순환 그래프(directed acyclic graph)를 제안했고, 욕심쟁이 동등 조사(greedy equivalence search)는 217 엣지들을 가진 그래프를 제안했다. 두 solutions은 91 엣지들과 61개의 같은 방향을 가지고 있다. 그러므로 그래프들은 대략 절반의 엣지들이 다르다.