

潜在クラスが存在する場合のベイズ的アプローチによる非ガウス因果構造推定法

A Bayesian estimation approach to analyze non-Gaussian data-generating process with latent classes

田中 直樹^{1*} 清水 昌平¹ 鷲尾 隆¹
Naoki Tanaka¹ Shohei Shimizu¹ Takashi Washio¹

¹ 大阪大学 産業科学研究所

¹ The Institute of Scientific and Industrial Research, Osaka University

Abstract: A large amount of observational data has been accumulated in various fields in recent times, and there is a growing need to estimate the generating processes of these data. A linear non-Gaussian acyclic model (LiNGAM) based on the non-Gaussianity of external influences has been proposed to estimate the data-generating processes of variables. However, the results of the estimation can be biased if there are latent classes. In this paper, we first review LiNGAM, its extended model, as well as the estimation procedure for LiNGAM in a Bayesian framework. We then propose a new Bayesian estimation procedure that solves the problem.

1 はじめに

データマイニングの分野では、データの生成モデルに非ガウス分布を仮定することで変数間の因果関係を向きも含めて導出する因果推論に関する手法が盛んに研究されている。しかし、**潜在クラス**と呼ばれる、データ生成過程の異なるデータが混在する場合、推定結果が歪められる可能性がある。近年、潜在クラスが存在する場合の観測変数間の因果関係を推定する方法が研究されている [12] が、局所解に収束してしまい、推定精度が低下する等の問題がある。そこで本稿では非ガウス性に基づく既存手法を、潜在クラスが存在する場合でも因果構造を正しく推定できるように改良し、局所解に陥ることを避けるよう試みる。そして、人工的に生成したデータと実データについて、それぞれ評価実験を行い、その結果について考察する。

2 先行研究

2.1 LiNGAM (Linear Non-Gaussian Acyclic Model) モデル

まずは本研究の基礎となる LiNGAM モデルについて説明する。LiNGAM モデルは以下の 4 つの仮定から

成る因果モデルである。

1. 観測変数 $x_i (i = 1, \dots, n)$ の間の因果関係は、図 1 のような **DAG (a directed acyclic graph)** によって表現可能である。
2. x_i の値は、既に値が決まっている観測変数 (DAG における x_i の親変数)、外的影響 e_i 、任意の定数 μ_i を用いて以下の**線形関数**によって生成される。

$$x_i = \sum_{k(j) < k(i)} b_{ij}(x_j - \mu_j) + e_i + \mu_i \quad (1)$$

ここで、 $k(i)$ は DAG 内の x_i の因果的順序を表す。なので、もし DAG 内で x_j から x_i へと有向辺が存在する場合、 $k(j) < k(i)$ である。また、 b_{ij} は x_j から x_i への影響の強さを表す。

3. 外的影響 e_i は平均 0、分散が非零の**非ガウス分布**に従う連続な確率変数とする。また、 e_i は互いに独立である。すなわち、 $p(e_1, \dots, e_n) = \prod_i p_i(e_i)$ を満たす。
4. データセット $\mathbf{D} = \{\mathbf{x}^1, \dots, \mathbf{x}^N\}$ (それぞれの \mathbf{x} は要素 x_i を含む) が観測され、各データベクトル \mathbf{x} は上記のデータ生成過程に従って同じ DAG、係数 b_{ij} 、定数 μ_i 、そして同じ分布から独立にサンプルされた外的影響 e_i を用いて生成される。この仮定は**潜在交絡変数が存在しない**ことを意味する [10]。潜在交絡変数とは、図 2 のような未観測かつ複数の観測変数に影響を与える変数である。

*連絡先：大阪大学 産業科学研究所
〒 567-0047 大阪府茨木市美穂ヶ丘 8-1
E-mail: tanaka@ar.sanken.osaka-u.ac.jp

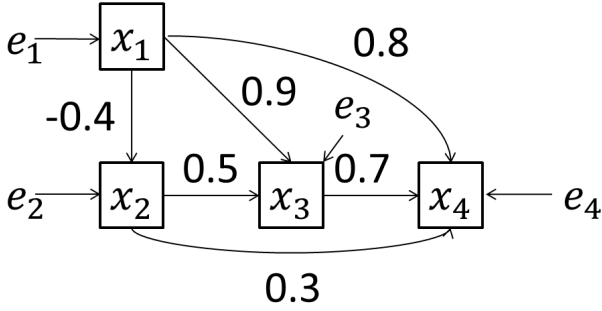


図 1: LiNGAM モデルを DAG で表した図. ノードは変数を表し, 辺は変数間の因果関係の向きとその結合の強さを表す.

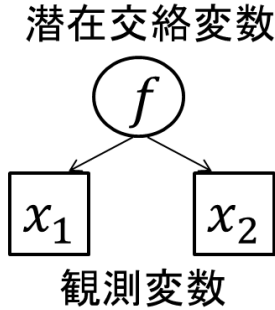


図 2: DAG における潜在交絡変数を表した図.

2.2 LiNGAM 混合モデル

LiNGAM モデルを用いて観測変数間の因果関係を推定する際に, 潜在クラスが存在する場合, 2.1 節で説明した最後の仮定が崩れてしまい, 推定結果が歪められてしまう場合がある. そこで, 潜在クラスが存在する場合の因果モデルとして LiNGAM モデルを拡張した LiNGAM 混合モデル [12] がある. 潜在クラスが存在する場合, データは各クラスで異なる「構造」を持つ. ここで構造とは, 因果的順序 $k(i)$, 係数 b_{ij} , 定数 μ_i , 外的影響が従う分布の確率密度 p_i から成る, というものを指す. そこで, クラス c に属するデータの構造を $k^{(c)}(i)$, $b_{ij}^{(c)}$, $\mu_i^{(c)}$, $p_i^{(c)}$ で表す. ($p_i^{(c)}$ を用いて外的影響 $e_i^{(c)}$ が生成されるということに注意.) それぞれのクラス ($c = 1, \dots, l$) に属するデータはそれぞれ異なる構造を持つ LiNGAM モデルから生成されると仮定すると, クラス c におけるデータ生成モデルは次のように表される.

$$x_i = \sum_{k^{(c)}(j) < k^{(c)}(i)} b_{ij}^{(c)}(x_j - \mu_j^{(c)}) + \mu_i^{(c)} + e_i^{(c)} \quad (2)$$

そして, データベクトル \mathbf{x} は次の混合分布に従って生成されるとする.

$$p(\mathbf{x}|\Theta) = \sum_{c=1}^l \left\{ p(\mathbf{x}|\theta^{(c)}) \right\} P(c) \quad (3)$$

ここで l は潜在クラスの数であり, $\Theta = [\theta^{(1)}, \dots, \theta^{(l)}]$ である. また, $\theta^{(c)}$ は式 (2) における全てのパラメータ, すなわち, $k^{(c)}(j)$, $k^{(c)}(i)$, $b_{ij}^{(c)}$, $\mu_j^{(c)}$, $\mu_i^{(c)}$, $p_i^{(c)}(e_i^{(c)})$ を含む. もし潜在クラスの数 l が 1 ならば, LiNGAM 混合モデルと LiNGAM モデルは等しい. $p_i^{(c)}(e_i^{(c)})$ をモデリングすることで $p(\mathbf{x}|\theta^{(c)})$ を定式化することができる. 本研究における $p_i^{(c)}(e_i^{(c)})$ のモデリングについては 3.1 節で述べる.

2.3 LiNGAM 混合モデルを用いた従来の因果構造推定法

LiNGAM 混合モデルを用いた従来の因果構造推定法として, LiNGAM 混合モデルを ICA (Independent Component Analysis[6], 独立成分分析) 混合モデルに変形し, ICA 混合モデルを推定するアルゴリズム (例として, [7] や [9] など) を適用するというアプローチが提案されている [12].

LiNGAM 混合モデル (式 (2)) は次のように行列を用いて表すことができる.

$$\mathbf{x} = \mathbf{B}^{(c)}\mathbf{x} + (\mathbf{I} - \mathbf{B}^{(c)})\boldsymbol{\mu}^{(c)} + \mathbf{e}^{(c)} \quad (4)$$

ここで, $\mathbf{B}^{(c)}$, $\boldsymbol{\mu}^{(c)}$, $\mathbf{e}^{(c)}$ はそれぞれ $b_{ij}^{(c)}$, $\mu_i^{(c)}$, $e_i^{(c)}$ を要素にもつ行列またはベクトルである. また, \mathbf{I} は単位行列である. この式 (4) を \mathbf{x} について解くことで ICA 混合モデルに変形できる.

$$\mathbf{x} = \boldsymbol{\mu}^{(c)} + \mathbf{A}^{(c)}\mathbf{e}^{(c)} \quad (5)$$

ここで, $\mathbf{A}^{(c)} = (\mathbf{I} - \mathbf{B}^{(c)})^{-1}$ である. ICA 混合モデルを推定するアルゴリズムを用いて $\mathbf{A}^{(c)}$ を求めることで, $\mathbf{B}^{(c)}$ を計算から得られる. しかしながら, ICA では独立成分のスケールと生成順序を決定することができないため, $\mathbf{B}^{(c)}$ を計算する前に $\mathbf{A}^{(c)}$ を正規化し, 置換を行う必要がある (詳細については [11] を参照).

2.4 BayesLiNGAM 法

LiNGAM 混合モデルの既存の推定手法 ([7][9]) は局所解に陥る場合があるため, 本稿ではそれを避けるためにベイズ的アプローチをとる. この節では 2.1 節で述べた LiNGAM モデルのベイズ的アプローチによる推定手法である, BayesLiNGAM 法 [5] について述べ

る。ベイズ推定は、ある複数の仮説の中からそれらの事後確率が最大である仮説を選択する推定法である。ここで BayesLiNGAM における仮説とは考えられ得る観測変数間の因果関係、すなわち DAG である。よって BayesLiNGAM はデータが与えられた際に事後確率が最大となる DAG を選択し、出力する。その事後確率はベイズの定理から計算することができ、次式で表される。

$$P(G_m|\mathbf{D}) = \frac{p(\mathbf{D}|G_m)P(G_m)}{p(\mathbf{D})} \quad (6)$$

ここで、 $G_m (m = \{1, \dots, N_g\})$ は考えられ得る DAG を表し、 N_g は n 個の観測変数について考えられ得る DAG の数である。また、 $\mathbf{D} (\mathbf{D} = \{\mathbf{x}^1, \dots, \mathbf{x}^N\})$, N はサンプルサイズ) は観測データセットであり、データが独立かつ同一の分布に従って生成されると仮定した時、 $p(\mathbf{D}) = \prod_{s=1}^N p(\mathbf{x}^s)$ となる。式 (6) から、事後確率を計算するには尤度 $p(\mathbf{D}|G_m)$ 、DAG に関する事前分布 $P(G_m)$ 、そして正規化定数 $p(\mathbf{D})$ を求める必要がある。まず、事前分布 $P(G_m)$ は、推定を行う前にわかっている専門知識や事前情報を加味して決定する。もしそれらが無ければ全ての G_m が等確率、すなわち $P(G_m) = \frac{1}{N_g}$ であるとする。次に、 $p(\mathbf{D})$ は事後確率を正規化する定数であり、以下のようにして計算できる。

$$P(\mathbf{D}) = \sum_{m=1}^{N_g} p(\mathbf{D}|G_m)P(G_m) \quad (7)$$

最後に尤度 $p(\mathbf{D}|G_m)$ は、次式のようにモデル $p(\mathbf{D}|\Theta, G_m)$ をパラメータ Θ について周辺化することで計算できる。

$$P(\mathbf{D}|G_m) = \int p(\mathbf{D}|\Theta, G_m)p(\Theta|G_m)d\Theta \quad (8)$$

Θ は式 (1) の全てのパラメータ ($k(j), k(i), b_{ij}, \mu_j, \mu_i, p_i$) をまとめた変数であり、 $p(\mathbf{D}|\Theta, G_m)$ はモデル (ここでは LiNGAM モデル) の尤度を表す。さらに、 $p(\Theta|G_m)$ はパラメータ Θ に関する事前分布である。ここで、もし仮説となる DAG (G_m) が与えられた時、その DAG における各観測変数 x_i の親がわかるので因果的順序 $k(i)$ を決定することができる。すなわち、式 (1) の右辺第 1 項は G_m が与えられた場合特定することができる。本研究で用いるモデルの尤度とパラメータ Θ に関する事前分布については 3 節で述べる。

3 提案手法

3.1 データ生成モデル

本研究では各クラス ($c = 1, \dots, l$) のデータが LiNGAM モデルに従って生成されると仮定する。よってクラス c のデータ生成モデルは式 (2) のように書くことができ、

そのときのデータベクトル \mathbf{x} の確率密度は式 (3) のように表される。そして LiNGAM 混合モデルの推定の際には、2.4 節で述べた BayesLiNGAM 法と同様にしてベイズ的アプローチをとる。

ここで、式 (8) における $p(\mathbf{D}|\Theta, G_m)$ を計算するためには、確率密度 p_i を式で表す必要がある。本研究では LiNGAM モデルの仮定の一つである、 e_i が非ガウス分布に従う連続変数であるという事から一般化ガウス分布を用いる [8]。一般化ガウス分布は左右対称であり、ガウス分布、ラプラス分布、連続一様分布を含む様々な非ガウス分布を表現可能である。そしてその確率密度関数は以下の式で表される。

$$p_i(e_i) = \frac{\lambda_i \sqrt{\frac{\Gamma(3/\lambda_i)}{\Gamma(1/\lambda_i)}} \exp(-(\sqrt{\frac{\Gamma(3/\lambda_i)}{\Gamma(1/\lambda_i)}} \frac{|e_i|}{\sigma_i})^{\lambda_i})}{2\sigma_i \Gamma(1/\lambda_i)} \quad (9)$$

σ_i は標準偏差、 λ_i は形状パラメータである。また、 $\Gamma()$ はガンマ関数である。

さらに、 $p(\mathbf{D}|\Theta, G_m)$ を計算するためには、確率密度の変数 e_i を x_i に変換する必要がある。式 (4) より、LiNGAM 混合モデルを行列の形で表した式を \mathbf{x} について整理すると、

$$\begin{aligned} \mathbf{x} &= \mathbf{f}(\mathbf{e}^{(c)}) \\ &= \boldsymbol{\mu}^{(c)} + (\mathbf{I} - \mathbf{B}^{(c)})^{-1} \mathbf{e}^{(c)} \end{aligned} \quad (10)$$

となる。 $\mathbf{f}()$ は写像ベクトルである。

クラス c における \mathbf{x} の確率密度 ($p(\mathbf{x}|\theta^{(c)}, G_m)$) は $\mathbf{e}^{(c)}$ の確率密度 ($p_{\mathbf{e}^{(c)}}^{(c)}(\mathbf{e}^{(c)})$) を用いて次式のように求められる [6]。

$$p(\mathbf{x}|\theta^{(c)}, G_m) = \frac{1}{|\det \mathbf{Jf}(\mathbf{f}^{-1}(\mathbf{x}))|} p_{\mathbf{e}^{(c)}}^{(c)}(\mathbf{f}^{-1}(\mathbf{x})) \quad (11)$$

\mathbf{Jf} はヤコビ行列である。式 (10) と式 (11)、そして LiNGAM モデルにおける仮定の一つ、非巡回性より $\det \mathbf{Jf}(\mathbf{f}^{-1}(\mathbf{x})) = 1$ であるので、次式が得られる。

$$p(\mathbf{x}|\theta^{(c)}, G_m) = p_{\mathbf{e}^{(c)}}^{(c)}(\mathbf{e}^{(c)}) \quad (12)$$

LiNGAM モデルにおいて e_i は互いに独立であると仮定しているので、

$$p_{\mathbf{e}^{(c)}}^{(c)}(\mathbf{e}) = \prod_{i=1}^n p_i^{(c)}(e_i^{(c)}) \quad (13)$$

であり、この時クラス c における \mathbf{x} の確率密度は式 (2)、(12)、(13) を用いて、

$$\begin{aligned} &p(\mathbf{x}|\theta^{(c)}, G_m) \\ &= \prod_{i=1}^n p_i^{(c)}(e_i^{(c)}) \\ &= \prod_{i=1}^n p_i^{(c)}(x_i - \mu_i^{(c)} - \sum_{k^{(c)}(j) < k^{(c)}(i)} b_{ij}^{(c)}(x_j - \mu_j^{(c)})) \end{aligned} \quad (14)$$

と表される。よって式 (3), (14), そしてデータが独立かつ同一の分布に従うという仮定から $p(\mathbf{D}|\Theta, G_m)$ を以下のように表すことができる。

$$\begin{aligned}
p(\mathbf{D}|\Theta, G_m) &= \prod_{s=1}^N p(\mathbf{x}^s|\Theta, G_m) \quad (15) \\
p(\mathbf{x}|\Theta, G_m) &= \sum_{c=1}^l \left\{ p(\mathbf{x}|\theta^{(c)}, G_m) \right\} P(c) \\
&= \sum_{c=1}^l \left\{ \prod_{i=1}^n p_i^{(c)}(x_i - \mu_i^{(c)} \right. \\
&\quad \left. - \sum_{k^{(c)}(j) < k^{(c)}(i)} b_{ij}^{(c)}(x_j - \mu_j^{(c)}) \right\} \times P(c) \quad (16)
\end{aligned}$$

3.2 パラメータの事前分布

本研究では、式 (3) における $P(c)$ に多項分布を用いる。この分布は、 N 回の独立な試行によってそれぞれ起こりうる事象 $c = 1, \dots, l$ のうちの 1 つが起こるといった状況において、それぞれの事象が起きた回数 ($\mathbf{z} = [z^{(1)}, \dots, z^{(l)}]$) を表現する分布である。 $w^{(c)}$ をそれぞれの事象が起こる確率とすると、多項分布の確率質量関数は次式で表される。

$$P(\mathbf{z}) = \frac{N!}{\prod_{c=1}^l z^{(c)}!} \prod_{c=1}^l (w^{(c)})^{z^{(c)}} \quad (17)$$

ここで $w^{(c)}$ は $w^{(c)} > 0$, $\sum_{c=1}^l w^{(c)} = 1$ を満たす。

多項分布のパラメータ ($w^{(1)}, \dots, w^{(l)}$) はディリクレ分布に従うとする。これはディリクレ分布が多項分布に対する共役事前分布であり、多項分布に対してディリクレ分布が一般的によく用いられるからである。ディリクレ分布の確率密度関数は次式で表される。

$$p(\mathbf{w}) = \frac{\Gamma(\sum_{c=1}^l a^{(c)})}{\prod_{c=1}^l \Gamma(a^{(c)})} \prod_{c=1}^l (w^{(c)})^{a^{(c)}-1} \quad (18)$$

$\mathbf{w} = w^{(1)}, \dots, w^{(l)}$ であり、 $a^{(c)} (> 0)$ は集中パラメータである。ディリクレ分布に従う乱数 \mathbf{w} を発生させるには、形状パラメータが $a^{(c)}$ 、スケールパラメータが 1 のガンマ分布から乱数 $\gamma^{(1)}, \dots, \gamma^{(l)}$ を独立に生成し、それらを正規化すればよい [4]:

$$w^{(c)} = \frac{\gamma^{(c)}}{\sum_{c=1}^l \gamma^{(c)}}, \quad (19)$$

$$p(\gamma^{(c)}) = \frac{1}{\Gamma(a^{(c)})} (\gamma^{(c)})^{a^{(c)}-1} \exp(-\gamma^{(c)}). \quad (20)$$

次に、 $b_{ij}^{(c)}$ と $\mu_i^{(c)}$ は平均がそれぞれ 0, $\varphi^{(c)}$ 、分散がそれぞれ v^2 , τ^2 のガウス分布に従うとする。

$$p(b_{ij}^{(c)}) = \frac{1}{\sqrt{2\pi v^2}} \exp\left\{-\frac{(b_{ij}^{(c)})^2}{2v^2}\right\} \quad (21)$$

$$p(\mu_i^{(c)}) = \frac{1}{\sqrt{2\pi\tau^2}} \exp\left\{-\frac{(\mu_i^{(c)} - \varphi^{(c)})^2}{2\tau^2}\right\} \quad (22)$$

ここで、 $\varphi^{(c)}$ の値には、混合ガウス分布モデルを、EM (expectation-maximization) アルゴリズム [2] を用いて推定した結果を用いるとする。これは、クラス数が増えるに従って $\mu_i^{(c)}$ の推定結果が事前分布により影響を受けるので、よりよい事前分布の設定を行うためである。

最後に、 $(\sigma_i^{(c)})^2$, $\lambda_i^{(c)}$, v^2 は形状パラメータがそれぞれ α , η , χ 、スケールパラメータがそれぞれ β , ζ , ϵ である逆ガンマ分布 [3] に従うとする。

$$p((\sigma_i^{(c)})^2) = \frac{\beta^\alpha}{\Gamma(\alpha)} (\sigma_i^{(c)})^{2(-\alpha-1)} \exp\left\{\frac{-\beta}{(\sigma_i^{(c)})^2}\right\} \quad (23)$$

$$p(\lambda_i^{(c)}) = \frac{\zeta^\eta}{\Gamma(\eta)} (\lambda_i^{(c)})^{-\eta-1} \exp\left(\frac{-\zeta}{\lambda_i^{(c)}}\right) \quad (24)$$

$$p(v^2) = \frac{\epsilon^\chi}{\Gamma(\chi)} v^{2(-\chi-1)} \exp\left(\frac{-\epsilon}{v^2}\right) \quad (25)$$

逆ガンマ分布に従う乱数は、ガンマ分布に従う乱数を用いて生成することができる。もし確率変数 X が形状パラメータ α 、スケールパラメータ β のガンマ分布に従うとすると、 $Y = 1/X$ は形状パラメータが α 、スケールパラメータが $1/\beta$ の逆ガンマ分布に従う [3]。

$a^{(c)}$, α , β , η , ζ , χ , ϵ , τ については、値が与えられているものとする。以上により、式 (8) における $P(\mathbf{D}|G_m)$ を、モンテカルロ積分 [2] を用いて計算することができる。

4 評価実験

今回の実験では、簡単のため二変数間に因果関係が存在するとし、各クラスで因果関係の向きは同じであるとする。そして真のモデルがどちらの向きであるか ($x_1^{(c)} \rightarrow x_2^{(c)}$ または $x_1^{(c)} \leftarrow x_2^{(c)}$) を推定する。

サンプルサイズ $N = 50, 100, 500$ 、クラス数 $l = 2, 4, 6$ についてそれぞれデータセットを 1,000 個生成させた。各クラスのデータは LiNGAM モデル (式 (1)) に従って生成させ、それを全てのクラスについて混合させた。全てのデータセットにおいて、真のモデルは $x_1^{(c)} \rightarrow x_2^{(c)}$ であるとし、観測変数間の影響の強さ $b_{ij}^{(c)} = [-1.5, -0.5] \cap [0.5, 1.5]$ とした。外生変数 $e_i^{(c)}$ が従う分布は平均 0、分散 1 のラプラス分布、連続一様分布、t 分布からランダムに選択した。

表 1: シミュレーション結果
サンプルサイズ

		50	100	500
提案手法	$l = 2$	913	947	981
	$l = 4$	908	937	973
	$l = 6$	922	957	967
既存手法	$l = 2$	649	657	684
	$l = 4$	663	655	729
	$l = 6$	646	700	762

また、 G_m に関する事前情報は無い、すなわち $P(G_m) = \frac{1}{2}$ であるとした。ハイパーパラメータの値の設定については、 $a^{(c)}$ は3, 5, 7からランダムに選択し、 $\alpha = 3$, $\beta = 3$, $\eta = 3$, $\zeta = 3$, $\chi = 3$, $\epsilon = 3$, $\tau = 0.5$ であるとした。

さらに、クラス数 l の決め方については、クラス数を1から $2\log N$ でそれぞれ固定して対数周辺尤度を計算し、その値が最も大きい時のクラス数を選択した。最大クラス数を $2\log N$ としたのは、ディリクレ混合過程において、 N が無限大に近づくと、 l は $\log N$ に近づくとという性質 [1] を参考にした。

そして、提案手法と既存手法 [12] を用いて因果関係の向きの推定を行い、真のモデルと向きが一致する回数を比較した。その結果を表 1 に示す。表内の数字は真のモデルと推定したモデルの因果関係の向きが一致した回数である。実験結果より提案手法は、従来の手法よりも精度良く推定を行えることがわかった。

5 むすび

本稿では、因果推論分野における既存の非ガウス性に基づく手法に、潜在クラスが存在しても正しい推定を可能にする改良を加え、その性能を評価した。従来の非ガウス性に基づく手法では、局所解に収束する等の問題点がある。本稿の成果により、潜在クラスが存在する場合でも精度良く推定を行えることを確認できた。今後の課題としては、人工的に生成したデータではなく、現実に蓄積されたデータに対して提案手法を適用し性能評価を行うことが挙げられる。

参考文献

- [1] C. E. Antoniak. Mixtures of dirichlet processes with applications to bayesian nonparametric problems. *The annals of statistics*, pp. 1152–1174, 1974.
- [2] C. M. Bishop. *Pattern Recognition and Machine Learning*. Springer, 2006.

- [3] John D Cook. Inverse gamma distribution. online: http://www.johndcook.com/inverse_gamma.pdf, Tech. Rep, 2008.
- [4] L Devroye. *Non-uniform Random Variate Generation*. Springer-Verlag, 1986.
- [5] P. O. Hoyer and A. Hyttinen. Bayesian discovery of linear acyclic causal models. In *Proc. 25th Conf. on Uncertainty in Artificial Intelligence (UAI2009)*, pp. 240–248, 2009.
- [6] A. Hyvärinen, J. Karhunen, and E. Oja. *Independent component analysis*. Wiley, New York, 2001.
- [7] M. N. H. Mollah, M. Minami, and S. Eguchi. Exploring latent structure of mixture ICA models by the minimum β -divergence method. *Neural Computation*, Vol. 18, pp. 166–190, 2006.
- [8] S. Nadarajah. A generalized normal distribution. *Journal of Applied Statistics*, Vol. 32, No. 7, pp. 685–694, 2005.
- [9] J. A. Palmer, S. Makeig, K. Kreutz-Delgado, and B. D. Rao. Newton method for the ICA mixture model. In *Proc. IEEE Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP2008)*, pp. 1805–1808, 2008.
- [10] J. Pearl. *Causality: Models, Reasoning, and Inference*. Cambridge University Press, 2000.
- [11] S. Shimizu, P. O. Hoyer, A. Hyvärinen, and A. Kerminen. A linear non-gaussian acyclic model for causal discovery. *J. Machine Learning Research*, Vol. 7, pp. 2003–2030, 2006.
- [12] S. Shimizu and A. Hyvärinen. Discovery of linear non-gaussian acyclic models in the presence of latent classes. In *Proc. 14th Int. Conf. on Neural Information Processing (ICONIP2007)*, pp. 752–761, 2008.

