頑健な独立成分分析

統計数理研究所／総合研究大学院大学
南 美穂子

独立成分分析とは

独立成分分析とは、いくつかの独立な発生源からの信号が線形に混合されて観測されるとき、その観測された信号からもとの互いに独立な信号を復元する手法である。本発表では、空間的混合のみで、原信号の数と観測信号の数が等しい $(m = l)$ Blind source separation の問題を考える。観測信号は原信号がある未知の正則行列 A によって $\boldsymbol{x}(t) = A\boldsymbol{s}(t), \quad t = 1, 2, ..., n$ と線形変換されたものであるとする。観測信号に行列 W をかけて $\boldsymbol{y}(t) = W\boldsymbol{x}(t), \quad t = 1, 2, ..., n$ として復元信号 $\boldsymbol{y}(t)$ を得る。復元行列 W が A^{-1} と等しければ原信号は完全に復元されたことになるが、独立という条件だけでは順序とスケールは定まらない。また、正規分布に従う原信号は高々一つとする。これは、正規分布に従う独立成分が２つ以上あると復元問題が識別不能となるからである。

Kullback-Leibler ダイバージェンスに基づく方法

まず、同時密度と周辺密度の積の間の Kullback-Leibler 距離（相互情報量）をもとにした方法を示す。復元信号 $\boldsymbol{y}$ の成分が互いに独立であれば、 $\boldsymbol{y}$ の同時密度は周辺密度の積で $\prod_{i=1}^m q_i(y_i)$ と表せ、$\boldsymbol{x}$ の同時密度は $|\det(W)| \prod_{i=1}^m q_i(\boldsymbol{w}_i\boldsymbol{x})$ と表せる。よって推定基準として相互情報量と呼ばれる以下の Kullback-Leibler 情報量：

$$D_{\mathrm{KL}}(W) = \int q(\boldsymbol{y}) \log \frac{q(\boldsymbol{y})}{\prod_{i=1}^m q_i(y_i)} d\boldsymbol{y} = \int f(\boldsymbol{x}) \log \frac{f(\boldsymbol{x})}{|\det(W)| \prod_{i=1}^m q_i(\boldsymbol{w}_i\boldsymbol{x})} d\boldsymbol{x}$$

を最小化することを考えよう。ところが、信号の分布は未知であり、q_i が何であるかわからない。そこで q_i のかわりに適当な密度関数 p_i を用いた

$$\widehat{D}_{\mathrm{KL}}(W) = \frac{1}{n} \sum_{t=1}^n \log \frac{f(\boldsymbol{x}(t))}{|\det(W)| \prod_{i=1}^m p_i(y_i(t))}, \qquad \text{ここで} \qquad y_i(t) = \boldsymbol{w}_i\boldsymbol{x}(t) \tag{1}$$

を最小化することによって推定値を求める。これは、擬似尤度

$$L_{\mathrm{QL}}(W) = \frac{1}{n} \sum_{t=1}^n l(\boldsymbol{x}(t), W) \qquad \text{ここで} \qquad l(\boldsymbol{x}, W) = \log|\det(W)| + \sum_{i=1}^m \log(p_i(y_i))$$

を最大化することと同値である。q_i のかわりに適当な密度関数 p_i を用いたのであるから、最尤推定量が一般的に持つ性質がそのまま保証されるわけではない。しかしながら、対応する推定関数は不偏であり、すそが正規分布より重いかどうかをあわせさえすれば漸近安定性も持つ。密度関数にパラメータを含むことも考えられるが、以降では、簡単のために密度関数にパラメータを含まない場合を述べる。パラメータを含んでいても同様のことが示せる。

推定関数は

$$F(\boldsymbol{x}, W) = \frac{\partial l(\boldsymbol{x}, W)}{\partial W} = \left(I_m - \boldsymbol{h}(\boldsymbol{y})\boldsymbol{y}^T\right) W^{-T} \tag{2}$$

ここで $\boldsymbol{y} = W\boldsymbol{x}$, $\boldsymbol{h}(\boldsymbol{y}) = (h_1(y_1), ..., h_m(y_m))^T$ and $h_i(y_i) = -d\log p_i(y_i)/dy_i = -cp_i'(y_i)/p_i(y_i)$ である。ここでは、相互情報量から (2) を導いたが、Bell & Sejnowski (1995) の相互情報量による方法, クロスキュムラントの最小化法 (Cardoso & Souloumiac, 1993), グラム・チャーリエ展開/自然勾配法 (Amari, Cichocki & Yang, 1996) など他の考え方から導出された多くの方法の推定関数は、これと同じように、$J(W)\left(I_m - \boldsymbol{h}(\boldsymbol{y})\boldsymbol{y}^T\right)G(W)$ という形で表せる（ただし、対角成分は除く）。原信号の平均が 0 であるという仮定をすれば、この推定関数が不偏であることは、容易に示せる。

しかしながら、この形の推定関数を持つ推定方法は、外れ値と思われるデータがある場合にうまく推定できない場合がある。これは、このタイプの推定関数はどのような密度関数を用いたとしても有界にはならない、つまり、B-ロバストではないからであろう。

β ダイバージェンスに基づく方法

そこで、本発表では、これまで提案された推定方法と同様の長所を持ち、なおかつ B-ロバスト性も持つ、β ダイバージェンスに基づく推定方法を提案する。

２つの密度関数間の β-ダイバージェンス (Eguchi & Kano, 2001) は、

$$D_\beta(g,f) = \begin{cases} \dfrac{1}{\beta}\displaystyle\int \left(g^\beta(\boldsymbol{x}) - f^\beta(\boldsymbol{x})\right) g(\boldsymbol{x})\ d\nu(\boldsymbol{x}) - \dfrac{1}{\beta+1}\int \left(g^{\beta+1}(\boldsymbol{x}) - f^{\beta+1}(\boldsymbol{x})\right)\ d\nu(\boldsymbol{x}) & \\ & \text{for} \quad \beta > 0 \\ \int \left(\log g(\boldsymbol{x}) - \log f(\boldsymbol{x})\right) g(\boldsymbol{x})\ d\nu(\boldsymbol{x}) & \text{for} \quad \beta = 0. \end{cases} \tag{3}$$

と定義される。$D_\beta(g,f)$ は 0 以上の値を取り、0 となるのは、$g = f$ a.e. の場合だけである。$\beta = 0$ の場合は、Kullback-Leibler ダイバージェンスとなる。β-ダイバージェンスは、density power ダイバージェンス (Basu, Harris, Hjort & Jones, 1998) の定数倍である。

原信号の平均が 0 であると仮定するかわりに、シフトパラメータ $\boldsymbol{\mu} = (\mu_1, ..., \mu_m)^T$ を導入することとし、適当な密度関数 p_i を用いて $r_0(\boldsymbol{x}, W, \boldsymbol{\mu}) = |\det(W)| \prod_{i=1}^m p_i(\boldsymbol{w}_i\boldsymbol{x} - \mu_i)$ と定義する。$\boldsymbol{x}$ の経験分布 $\tilde{r}$ と $r_0(\boldsymbol{x}, W, \boldsymbol{\mu})$ の間の β-ダイバージェンス $\widehat{D}_\beta(\tilde{r}, r_0(\cdot, W, \boldsymbol{\mu}))$ を基準とした推定方法を考えよう。$\widehat{D}_\beta(\tilde{r}, r_0(\cdot, W, \boldsymbol{\mu}))$ の符号を変えて定数をとると以下の擬似 β-尤度関数を得る。

$$L_\beta(W, \boldsymbol{\mu}) = \begin{cases} \dfrac{1}{n}\dfrac{1}{\beta}\displaystyle\sum_{t=1}^n r_0^\beta(\boldsymbol{x}(t), W, \boldsymbol{\mu}) - b_\beta(W) - \dfrac{1-\beta}{\beta} & \text{for} \quad 0 < \beta < 1 \\ \dfrac{1}{n}\displaystyle\sum_{t=1}^n \log\left(r_0(\boldsymbol{x}(t), W, \boldsymbol{\mu})\right) & \text{for} \quad \beta = 0, \end{cases} \tag{4}$$

ここで、$b_\beta(W) = 1/(\beta+1)\int r_0^{\beta+1}(\boldsymbol{x}, W, \boldsymbol{\mu})d\boldsymbol{x} = |\det(W)|^\beta/(\beta+1)\int q_0^{\beta+1}(\boldsymbol{z})d\boldsymbol{z}$. この関数の最大値を W と $\boldsymbol{\mu}$ の推定量 $\hat{W}$、$\hat{\boldsymbol{\mu}}$ とする。推定関数は、

$$F_1(\boldsymbol{x}, W, \boldsymbol{\mu}) = r_0^\beta(\boldsymbol{x}, W, \boldsymbol{\mu})\left(I_m - \boldsymbol{h}(W\boldsymbol{x} - \boldsymbol{\mu})\ (W\boldsymbol{x})^T\right)W^{-T} - \beta\ b_\beta(W)W^{-T}, \tag{5}$$

$$F_2(\boldsymbol{x}, W, \boldsymbol{\mu}) = r_0^\beta(\boldsymbol{x}, W, \boldsymbol{\mu})\ \boldsymbol{h}(W\boldsymbol{x} - \boldsymbol{\mu}) \tag{6}$$

となる。この推定関数は不偏で、なおかつ、有界であり、この推定量はＢ－ロバストである。発表では、漸近安定性の条件、計算アルゴリズム、スピーチの分離、シミュレーション結果なども示した。