平成２９年度 共同利用研究報告書

平成　30 年　10　 月 23　 日

九州大学 マス・フォア・インダストリ研究所長 殿

所属・職名　 大阪大学大学院基礎工学研究科 教授

提案者 氏(ふりがな)名　 鈴木 譲（すずきじょう）

下記の通り共同研究の報告をいたします.

記

		※整理番号	20170005	
1.研究計画題目	スパースな統計モデリングと情報量基準：因子分析への応用			
2.種別（○で囲む）	a. 研究集会Ⅰ　b. 研究集会Ⅱ　c.短期共同研究　✓d.短期研究員			
3.研究代表者	氏(ふりがな)名	鈴木譲		
	所属部局名	大阪大学 大学院基礎工学研究科	職名	教授
	連絡先			
	e-mail		TEL	
4.研究実施期間	平成 29 年 12 月 6 日（水 曜日）～平成 29 年 12 月 8 日（金 曜日）			

5.参加者数・参加者リスト（＊別紙「共同利用研究報告書作成上の注意」参照）

(a,b は参加者数のみ記入し，集会参加者リストを添付．c.の非公開プログラム参加者と d.は参加者リストに記入.

c.は公開プログラムを含めた全参加者数を記入し，公開プログラム参加者リストを添付.）

参加者数：　2　人

参加者リスト（a,b は記入不要，c.は非公開プログラム参加者，d.は共同研究参加者を記入）

氏(ふりがな)名	所属	職名	氏(ふりがな)名	所属	職名
廣瀬　慧	九州大学IMI	准教授			

6.本研究で得られた成果の概要

滞在期間が3日間という短く、証明が完成したとか新しいアルゴリズムが提案できたというような成果はなかった。しかし、スパース推定に興味がもてたこと、その後の共同研究に発展する素地が作れたことは収穫であると考えている。まず、滞在(2017 年 12 月)から 3 ヶ月たった 2018 年 3 月に阪大で開催されたスパース推定に関するシンポジウム(阪大 MMDS 主催 機械学習・統計学スプリングキャンプ)を開催した。知名度の高い九大 MMI の廣瀬慧准教授を招待することで、200 名の参加者を集めることができた。また、日本行動計量学会第 46 回大会(2018 年 9 月)で、「多変量解析におけるスパース推定」(4 件の講演、各 1 ページを添付)というラウンドテーブルを開催できた(廣瀬氏にもそのイベントにご参加いただけた)。また、2018 年 11 月には、日本行動計量学会秋の行動計量セミナー「スパース推定：手を動かしてみる」を開催することになった。代表者の鈴木は、2017 年の秋までスパース推定に関してはほとんど知識がなかった。2017 年 12 月の九大 IMI 短期共同研究で、廣瀬氏の研究室に滞在することがなければ、こうしたことは実現しなかったと考えている。このような機会をくださった九大 IMI に感謝申し上げたい。

スパース推定の概要をわかりやすく

鈴木 讓

大阪大学大学院基礎工学研究科

統計パラメータの各値の絶対値の和に対して罰則を加える (L1 正則化) ことによってパラメータ数を適度なレベルに調節する、いわゆるスパース推定が、データサイエンスや機械学習の諸問題で検討されるようになっている。しかしながら、どのような動作をするのか、どのような問題に適用できるのかなどについて、一部の研究者をのぞいては、十分な理解が得られていないように思われる。

本講演は、ラウンドテーブル「多変量解析におけるスパース推定」の冒頭にさきだち、諸概念をわかりやすく説明し、参加される方々が予備知識などなくても、違和感なく議論の中に入っていけるようにすることを目的する。

1. 情報量基準と比較して、Lasso を適用するとどのような点が有利なのか。尤度とパラメータ数のバランスをとるという意味では同じではないか。
2. Lasso は、尤度とパラメータの絶対値の和を重み付け (後者を λ 倍) した和を最小にしているが、それによってパラメータ数が削減されるのはなぜか。絶対値の和ではなく、2 乗の和を最小にする (Ridge 回帰) のでは、なぜパラメータが削減されないのか。
3. 最適な λ はどのように求めるのか。
4. Lasso は、新しい定式化を提案するたびに、その最適解を求める方法を自分で提案する必要があるのか。プログラミングなどせずに、自分の問題やデータを既存の Lasso の方法に適用するには、どのようにすればよいのか。
5. 線形回帰の場合の Lasso の話はよく聞くが、ロジスティック回帰や Cox 平均ハザードのような一般化線形回帰には適用できるのか。一般に、スパース推定の応用にはどのようなものがあるのか。

引き続き、本ラウンドテーブル「多変量解析におけるスパース推定」は、以下の 3 テーマについて議論する。

- 主成分分析と正準相関分析 (文献 [1] の 8 章 8.2, 8.3)
- 判別分析とクラスタ分析 (文献 [1] の 8 章 8.4, 8.5)
- グラフィカルモデル (文献 [1] の 9 章)

それぞれについて、「どのような研究課題が残されているか」(研究者向け)、「その方法がどのような応用をもっているか」(実務者向け) を議論する。入門的な書籍として、文献 [1] のようなものがあるが、カバーされていない部分もあるので、当日は事前知識をそろえるためのビデオを提供する予定である。

参考文献

[1] T. Hastie, R. Tibshirani, and M. Wainwright. *Statistical Learning with Sparsity: The Lasso and Generalizations*. Chapman & Hall/CRC, 2015.

[2] 川野秀一, 松井秀俊, 廣瀬慧. スパース推定法による統計モデリング. 共立出版, 2018.

多変量解析におけるスパース推定 (1)
スパース主成分分析、スパース正準相関分析

五條喜仁、鈴木 譲

大阪大学大学院基礎工学研究科

スパース主成分分析

サンプル数 N、変数の個数が p のデータ行列 $X \in \mathbb{R}^{N\times p}$(各行を $x_i \in \mathbb{R}^p$ とかくものとする) から第 1 主成分を求める際に、定数 $t > 0$ に対して、$||v||_2 = 1$, $||v||_1 \le t$ のもとで、射影した分散 v^TX^TXv を最大化にする v を求める SCoTLASS [1] と、$||\theta||_2 = 1$, $\lambda_1, \lambda_2 > 0$ のもとで、

$$\frac{1}{N}\sum_{i=1}^{N} |x_i - \theta v^T x_i||_2^2 + \lambda_1 ||v||_1 + \lambda_2 ||v||_2^2 \tag{1}$$

を最小にする $\theta, v \in \mathbb{R}^p$ を求める SPCA [2] がある ([3] の 5 章に日本語の解説)。いずれも、L_1 正則化によって、$\lambda > 0$ の値に応じて、第 1 主成分の負荷ベクトル中の 0 の個数を多くしているが、前者は第 1 主成分 v に射影した分散 v^TX^TXv を最大に、後者は第 1 主成分 $(v^Tx_1, \cdots, v^Tx_N)$ を $\theta \in \mathbb{R}^p$ 倍して、x_i に復元しようとしたときの 2 乗誤差を最小にすることを目的としている。(1) の最後の 2 項 3 項のような正則化の設定は、エラスティックネットとよばれ、p 個の変数で特性の近い変数に対して、近い係数の値にする役割がある。

また、第 2 主成分以降を求める際に、スパース推定の場合、各成分で独立に最適化をはかることはできない。SPCA [2] は、(1) を一般化することによって、任意個の成分まで同時に解を求めることができるとしている。当日は、SCoTLASS を凸問題に緩和した定式化と計算方法、SCoTLASS および SPCA の計算方法、両者の定式化の関連性、ディープラーニングとの関連について議論する。

スパース正準相関分析

正準相関分析は、$p, q \ge 1$ として、$X \in \mathbb{R}^{N\times p}$ および $Y \in \mathbb{R}^{N\times q}$ をデータ行列として、$\beta \in \mathbb{R}^p$, $\theta \in \mathbb{R}^q$ を $\frac{1}{N}\beta^TX^TX\beta = 1$ かつ $\frac{1}{N}\theta^TY^TY\theta = 1$ の範囲で動かして、$\frac{1}{N}\beta^TX^TY\theta$ の最大化として定式化される。これに、定数 $c_1, c_2 > 0$ を用意して、$||\beta||_1 \le c_1$, $||\theta||_1 \le c_2$ の制限をおいたのが、スパース正準相関分析である。p, q が大きい場合、計算ができないので、$\frac{1}{N}\beta^TX^TX\beta = 1$, $\frac{1}{N}\theta^TY^TY\theta = 1$ の制約を $||\beta||_2^2 \le 1$, $||\theta||_2^2 \le 1$ におきかえて計算する方法も提案されている。

参考文献

[1] I. Jolliffe, N. Trendafilov, and M. Uddin. A modified principal component technique based on the lasso. *Journal of Computational and Graphical Statistics*, 12(3):531–547, 2003.

[2] H. Zou, T. Hastie, and R. Tibshirani. Sparse principal component analysis. *Journal of Computational and Graphical Statistics*, 15(2):265–286, 2006.

[3] 川野秀一, 松井秀俊, and 廣瀬慧. スパース推定法による統計モデリング. 共立出版, 2018.

多変量解析におけるスパース推定 (2)
判別分析とクラスタリング

吉岡凜太郎、鈴木讓

大阪大学大学院基礎工学研究科

参考文献 [1] で紹介されている事例を中心に紹介する。当日は、デモなどを交える。

判別分析

以下の判別分析にスパースのオプションを設定する

- 最近傍法
- Fisher の線形判別
- 最適スコアリング

$i = 1, \cdots, N$ 番目のサンプルのクラスが $k = 1, \cdots, K$ であるときに $y_{i,k} = 1$、それ以外で $y_{i,k} = 0$ とし、$i = 1, \cdots, N$ 番目のサンプルの $j = 1, \cdots, p$ 番目の共変量が $x_{i,j}$ であるときに、$X = (x_{i,j})$ および $Y = (y_{i,k})$ とさだめる。そして、各 $k = 1, \cdots, K$ で、$\theta_k^T Y^T Y \theta_k = 1$ および $\theta_k^T Y^T Y \theta_j = 0$, $j = 1, 2, \cdots, k-1$ のもとで、$||Y\theta_k - X\beta_k||_2^2$ を最小にする $\beta_k \in \mathbb{R}^p$ および $\theta_k \in \mathbb{R}^K$ を求める処理が最適スコアリングである。

クラスタリング

以下の 2 クラスタリングにスパースのオプションを設定する

- 階層的クラスタリング
- K-means クラスタリング

また、凸クラスタリングといって、観測された $x_1, \cdots, x_N$ から、$\lambda > 0$ に対して、

$$\frac{1}{2}\sum_{i=1}^{N} ||x_i - u_i||_2^2 + \lambda \sum_{i<i'} w_{i,i'} ||u_i - u_{i'}||_q$$

を最小にするように $S = \{u_1, \cdots, u_N\}$ を定める方法もある。ここで、$q = 1$ または $q = 2$ とし、重み $w_{i,i'}$ は 1 としてもよいし、$x_i, x_{i'}$ の距離に依存してもよい。必然的に、S の要素が N より小さくなる。

参考文献

[1] T. Hastie, R. Tibshirani, and M. Wainwright. *Statistical Learning with Sparsity: The Lasso and Generalizations.* Chapman & Hall/CRC, 2015.

多変量解析におけるスパース推定 (3)
グラフィカルモデル

張秉元、鈴木讓

大阪大学大学院基礎工学研究科

確率変数 $X = (X^{(1)}, \cdots, X^{(p)})$ が、平均が 0 の p 変量正規分布にしたがうものとする。ここで、その共分散行列の逆行列が $\Theta \in \mathbb{R}^{p\times p}$ であるとき、$\theta_{st} = 0$ であることと、$X^{(s)}, X^{(t)}$ が何らかの変数のもとで条件付き独立になることが同値であることが知られている。すなわち、$\theta_{st} \neq 0$ であるということと $X^{(s)}$ と $X^{(t)}$ が辺でむすばれていることが同値になる。以下では、この分布を $P_\Theta(x)$, $x \in \mathbb{R}^p$ とかく。

そして、共分散行列の推定量 $S := \frac{1}{N}X^TX$ を用いて、N 組のサンプル $x_1, \cdots, x_N \in \mathbb{R}^p$ のもとでの尤度 $\dfrac{1}{N}\log \mathbb{P}_\Theta(x)$ が計算できる。この値を最大にするするように Θ を決めることができるが、ここでは L1 正則化を適用して、一部のパラメータ θ_{st} を強制的に 0 にして、p 変数の間の依存関係を無向グラフで表現することを考える。$\lambda > 0$ として、以下の値を最小化する Θ を求める。

$$-\frac{1}{N}\log \mathbb{P}_\Theta(x) + \lambda \sum_{s\neq t} |\theta_{st}| \tag{1}$$

最後の項は、微分できないが、Θ で劣勾配 (微分の一般化) を施すと (詳細は、文献 [1] を参照されたい)、

$$\Theta^{-1} - S - \lambda\Psi = 0 \tag{2}$$

の解を求める問題に帰着できる。ここで、$\Psi = (\psi_{st})$ は、$s = t$ のとき $\psi_{st} = 0$。それ以外で、$\theta_{st} > 0$ のとき $\psi_{st} = 1$、$\theta_{st} < 0$ のとき $\psi_{st} = -1$、$\theta_{st} = 0$ のとき $\psi_{st} \in [-1, 1]$ となる。この問題は、(2) の Θ の解を計算する問題に帰着する。$\lambda = 0$ であれば、すべての辺が結ばれるが、λ が大きくなるにつれ、辺の個数が減少する。p の値が大きくても、λ の値が大きければ、非常に短い時間で解が求まる。

本講演では、上記の他に、以下の問題を扱う。

- 正規分布を仮定した劣尤度を用いた推定
- 離散分布を仮定した劣尤度を用いた推定
- 混合分布を仮定した劣尤度を用いた推定
- 隠れ変数をもったグラフィカルモデル

最初の 3 方法は、各変数が依存する変数を回帰 (正規分布の場合は線形回帰、離散分布の場合はロジスティック回帰) によって求めるもので、一方の変数が他方の変数に依存するがその逆がない場合は、相互に依存し合う場合、少なくとも一方が依存する場合というようにルールを決めて、無効グラフを構成する。

参考文献

[1] 川野秀一, 松井秀俊, and 廣瀬慧. スパース推定法による統計モデリング. 共立出版, 2018.