

2011 年度科学研究費シンポジウム

## 生命科学と統計学

会 場：大阪大学 基礎工学国際棟(シグマホール)セミナー室  
豊中市待兼山町 1-3

日 時：平成 23 年 11 月 3 日(木)～11 月 5 日(土)

世 話 人：狩野 裕(大阪大学 基礎工学研究科)

基盤研究(A)「非対称・非線形統計理論と経済・生体科学への応用」

(研究代表：谷口正信, 研究課題番号:23244011)

基盤研究(B)「予測・因果・不完全データ解析とサイエンスの基礎」

(研究代表：狩野 裕, 研究課題番号:22300096)

挑戦的萌芽研究「統計的正則化理論と神経生理学:脳科学との接点」

(研究代表：狩野 裕, 研究課題番号:23650145)

# Scientific Program

November 3, Thursday

Theme 1: Factor Rotation and Sparseness (English session)

Chair: Etsuo Kumagai (Osaka Univ)

10:00–10:10 Yutaka Kano (Osaka Univ. Grad Sch Engineering Science)

「Brief introduction to this session」

10:10–11:10 Kohei Adachi (Osaka Univ. Grad Sch Human Science)

Nickolay T. Trendafilov (The Open University, UK)

「Sparse component and factor analysis procedures via directly  
constraining the sparseness of loadings」

11:10–11:55 Michio Yamamoto (Osaka Univ. Grad Sch Engineering Science, D2)

「A cluster-based factor rotation and its properties」

11:55–13:30 break for lunch

Chair: Yutaka Kano (Osaka Univ)+Kohei Adachi (Osaka Univ)

13:30–14:30 Robert I. Jennrich (UCLA, USA)

「Rotation algorithms from the beginning and exploratory bi-factor analysis」

14:30–15:15 Sho Hashimoto (Osaka Univ. Grad Sch Human Science, D2)

「A Bayesian exploratory factor analysis solution with the simplimax target」

15:15–15:45 break for coffee

15:45–16:45 Nickolay T. Trendafilov (The Open University, UK)

「Exploratory factor analysis: modern aspects」

16:45–17:30 Hironori Satomura (Osaka Univ. Grad Sch Human Science, M2)

「Exploratory data-fitting factor analysis with clustering of individuals」

Informal reception

(to be continued)

11月4日（金）

テーマ2： シンボリックデータ解析

座長：藤本翔太（大阪大学）

- 10:00-11:00 水田正弘（北海道大学 情報基盤センター）  
「シンボリックデータ解析の考え方と手法」
- 11:00-11:45 寺田吉壱（大阪大学 基礎工学研究科 D1）  
「シンボリックデータ解析における MDS について」
- 11:45-13:30 昼休憩

テーマ3： 統計的正則化理論と神経生理学

座長：狩野裕（大阪大学）＋廣瀬慧（大阪大学）

- 13:30-13:45 狩野 裕（大阪大学 基礎工学研究科）  
「本テーマの序」
- 13:45-14:45 赤穂昭太郎（産業技術総合研究所, ヒューマンライフテクノロジー研究部門）  
「カーネル法と正則化」
- 14:45-15:30 廣瀬 慧（大阪大学 基礎工学研究科）  
「正則化法に基づくスパース推定と調整パラメータの選択」
- 15:30-16:00 休憩
- 16:00-16:45 小林 康（大阪大学 生命機能研究科）  
「学習神経回路における活動同期現象について」
- 16:45-17:45 乾 敏郎（京都大学 情報科学研究科）  
「予測的処理の神経生理学」
- 18:30-20:30 総合討論（ミュージアムカフェ坂, キャンパス内）

[次ページへ続く](#)

11月5日(土)

テーマ4： 生物統計・生物情報

座長：内藤貫太（島根大学）＋大西俊郎(九州大学)

- 10:00-10:45 大西俊郎(九州大学 経済学研究院)・柳本武美(中央大学 理工学部)  
「Dual saddlepoint equalities in model averaging」
- 10:45-11:30 柳本武美(中央大学 理工学部)・大西俊郎(九州大学 経済学研究院)  
「Bayesian モデル e-統合：複数の証拠と意見」
- 11:30-13:00 昼休憩
- 13:00-13:30 小林千鶴(中央大学 理工学研究科 M1)・鎌倉稔成(中央大学 理工学部)  
柳本武美(中央大学 理工学部)  
「有意でない試験を継続するベイズ的視点」
- 13:30-14:15 廣瀬英雄（九州工業大学 情報工学研究院）  
「Estimation of the number of failures using ordinary differential equations  
to reliability models, with applications to SARS, A(H1N1), and FMD cases」
- 14:15-14:30 休憩
- 14:30-15:15 内藤貫太（島根大学 総合理工学部）・野津昭文（総合研究大学院大学 D2）  
宇田川 潤（滋賀医科大学 医学部）・大谷 浩（島根大学 医学部）  
「Statistical analysis with dilatation for development process of human fetuses」
- 15:15-16:00 熊澤 蕃（元内閣府原子力安全委員会事務局技術参与）  
「確率的比例効果則とフィードバック効果を組み合わせた生体リスク管理モデル」

以上

## Introduction to the session “Factor Rotation and Sparseness”

Yutaka Kano (Osaka Univ)

The factor rotation problem has a long history (e.g., Harman 1976; Browne 2001). Psychometricians had solved the indeterminacy problem by introducing ‘interpretability,’ which can be called sparseness of factor loadings in modern terminology. Two important developments on factor rotation with interpretability are to present analytical rotation methods (e.g. Kaiser 1958) and to derive analytical standard errors of rotated factor loading estimators (Jennrich 1973).

An alternative solution to the factor rotation problem was given by Mooijaart (1985) who used the nonnormality and independence assumptions of common factors. The approach is equivalent to independent component analysis (ICA).

We should not forget that psychometricians had found the important notions, sparseness and ICA, long time ago. In this session, we would like to review the history of the factor rotation problem, and at the same time, to discuss most recent research on it.

Personally I can't help wondering, why sparseness means interpretability. Later sessions, we will discuss that sparse coding is an actual activity of human brain when one views objects (e.g., Olshausen and Field, 1996). Recent developments in brain science could provide a hint to solve my question above.

### References

- Browne, M. W. (2001). An overview of analytic rotation in exploratory factor analysis. *Multivariate Behavioral Research*, 36, 111-150.
- Harman, H. H. (1976). Modern factor analysis, third edition. Chicago: University of Chicago Press.
- Jennrich, R. I. (1973). Standard errors for obliquely rotated factor loadings. *Psychometrika*, 38, 593–604.
- Kaiser, H. F. (1958). The varimax criterion for analytic rotation in factor analysis. *Psychometrika*, 23, 187-200.
- Mooijaart, A. (1985). Factor analysis for non-normal variables. *Psychometrika*, 50, 323-342.
- Olshausen, B. A. and . Field, D. J. (1996). Emergence of simple-cell receptive field properties by learning a sparse code for natural images. *Nature*, 381(13), 607-609.

# Sparse Component and Factor Analysis Procedures via Directly Constraining the Sparseness of Loadings

Kohei Adachi <sup>1</sup>

Nickolay T. Trendafilov <sup>2</sup>

<sup>1</sup>Graduate School of Human Sciences, Osaka University, Japan

<sup>2</sup>Faculty of Mathematics, Computing & Technology, The Open University, United Kingdom

## 1. Introduction

In the symposium, we proposed a new technique for sparse PCA in which the PCA loss function is simultaneously minimized and the degree of the required sparseness of the loadings matrix is achieved. Further, we extended this approach to exploratory factor analysis (EFA). Our presentation was organized as follows. In section 2, the new sparse PCA was considered in detail. Section 3 outlined the procedures for avoiding local minima and selecting the optimal sparsity based on well known information criteria. In Section 4, a sparse EFA technique was proposed which was formulated in a similar manner to the sparse PCA in Section 2.

## 2. Sparsity Constrained Principal Component Analysis

Let  $\mathbf{X}$  be an  $n$ -observations  $\times$   $p$ -variables column-centered data matrix. Then, the ordinary PCA is formulated as minimization of the following least squares loss function

$$f(\mathbf{F}, \mathbf{A}) = \|\mathbf{X} - \mathbf{FA}'\|^2 = \text{tr}(\mathbf{X} - \mathbf{FA}')'(\mathbf{X} - \mathbf{FA}') \quad (1)$$

over an  $n$ -observation  $\times$   $m$ -components score matrix  $\mathbf{F}$  and  $p \times m$  loadings matrix  $\mathbf{A}$ , subject to normalization constraint

$$\frac{1}{n} \mathbf{F}'\mathbf{F} = \mathbf{I}_m . \quad (2)$$

Here,  $\mathbf{I}_m$  denotes the  $m \times m$  identity matrix with the number of components  $m \leq \min(n, p)$ . To the above ordinary PCA, we add a direct sparsity constraint that

$$q \text{ elements of } \mathbf{A} \text{ should be exact zeros,} \quad (3)$$

i.e., constrain the sparsity of the loadings matrix to be a prescribed constant  $q$ . Thus, our proposed sparse PCA is formulated as minimizing (1) over  $\mathbf{F}$  and  $\mathbf{A}$  subject to the two constraints (2) and (3). The solution is obtained by alternately iterating [1] the minimization of (1) over  $\mathbf{A}$  subject to (2) with  $\mathbf{F}$  kept fixed and [2] that of (1) over  $\mathbf{F}$  subject to (2) with  $\mathbf{A}$  fixed. Further, we can skip [1] to obtain the solution from covariance matrix  $\mathbf{S} = n^{-1} \mathbf{X}'\mathbf{X}$ .

### 3. Treatment of Local Minima and Sparsity Selection

The proposed method is sensitive to local minima. In order to avoid (or at least minimize) the chance of choosing a local minimizer as a final solution, we employ a two-best-solution stopping procedure, in which the starts of the above algorithm are replicated until two solutions termed ‘equivalently best’ are obtained.

The minimization of (1) is equivalent to maximizing the likelihood derived from the model  $\mathbf{X} = \mathbf{FA}' + \sigma\mathbf{E}$  where  $\mathbf{E}$  is a matrix of standard normal errors independent of each other. The equivalence is shown by the fact that the log likelihood is given by  $\log L = \text{const} - 0.5np\log\sigma^2 - 0.5\sigma^{-2}\|\mathbf{X} - \mathbf{FA}\|^2$  and the ML estimate of  $\sigma^2$  is given by  $np^{-1}\|\mathbf{X} - \mathbf{FA}\|^2$ . This fact implies  $2\log L = \text{const} - np\log\{f(\mathbf{F}, \mathbf{A})/(np)\} - np = \text{const} - np\log\{f(\mathbf{A})/(np)\} = \text{const}^\# - np\log f(\mathbf{A})$  with  $\text{const}$  and  $\text{const}^\#$  irrelevant of  $\mathbf{F}$  and  $\mathbf{A}$ . Further, the number of parameters is  $nm+pm-q$ , i.e., the number of the elements of  $\mathbf{F}$  and  $\mathbf{A}$  minus that of the zero elements in  $\mathbf{A}$ . Given  $m$ , we can thus choose the optimal  $q$  using the well known information criteria AIC and BIC, i.e.,

$$\text{AIC}(q) = -2\log L + 2(nm+pm-q) = \text{const}^\# - np\log f(\mathbf{A}) - 2q \quad (4)$$

$$\text{BIC}(q) = -2\log L + (nm+pm-q)\log n = \text{const}^\# - np\log f(\mathbf{A}) - q\log n \quad (5)$$

with  $\text{const}^\#$  irrelevant to the selection of  $q$ .

### 4. Sparse Factor Analysis

Recently, a new EFA technique has been proposed whose least squares function is defined as

$$f(\mathbf{Z}, \mathbf{A}, \mathbf{\Psi}) = \|\mathbf{X} - (\mathbf{FA}' + \mathbf{U}\mathbf{\Psi})\|^2 = \|\mathbf{X} - \mathbf{Z}[\mathbf{A}, \mathbf{\Psi}']\|^2 \quad (6)$$

for the  $n \times p$  data matrix  $\mathbf{X}$ . Here,  $\mathbf{A}$  is the  $p \times m$  matrix of factor loadings,  $\mathbf{\Psi}$  is the  $p \times p$  diagonal matrix including the squared roots of unique variances, and  $n \times (m+p)$  matrix  $\mathbf{Z} = [\mathbf{F}, \mathbf{U}]$  consists of  $\mathbf{F}$  ( $n \times m$ ) and  $\mathbf{U}$  ( $n \times p$ ) that contains common and unique factor scores, respectively, and are constrained as  $n^{-1}\mathbf{F}'\mathbf{F} = \mathbf{I}_m$ ,  $n^{-1}\mathbf{U}'\mathbf{U} = \mathbf{I}_p$ , and  $n^{-1}\mathbf{F}'\mathbf{U} = \mathbf{O}_{m \times p}$  with  $\mathbf{O}_{m \times p}$  the  $m \times p$  matrix of zeros. Those three constraints are summarized into

$$\frac{1}{n}\mathbf{Z}'\mathbf{Z} = \mathbf{I}_{m+p}. \quad (7)$$

The sparsity constraint on  $\mathbf{A}$  expressed as (3) is incorporated into this EFA technique. That is, our proposed sparse EFA is formulated as minimizing (17) over  $\mathbf{Z}$ ,  $\mathbf{A}$ , and  $\mathbf{\Psi}$  subject to (18) and (3). Its solution can be obtained with an algorithm similar to the one for the sparse PCA procedure in Section 2.

# A cluster-based factor rotation and its properties

MICHIO YAMAMOTO  
OSAKA UNIVERSITY

## 1. Introduction

Exploratory factor analysis (EFA) is a popular statistical method used to find the latent structure behind multivariate data in an exploratory way. As is well known, a solution in EFA has the indeterminacy of rotation of a factor loading matrix. For interpretation of a result in EFA, the initial loading matrix is usually rotated such that the corresponding pattern becomes as simple as possible. That is, many rotation methods intend to find perfect simple structure, which is a loading matrix if each row has at most one nonzero element. However, perfect simple structure does not always exist. In addition, there may be some situations where a little more complex structure is interpretable. When each row has more than one nonzero element, existing methods may fail to find such a little more complex structure. Thus, a rotation method, which can produce a little more complex structure than perfect simple structure, is needed. In this paper, for such a goal, a new rotation method is proposed using information about cluster structure of variables.

## 2. The Proposed Method

A new oblique rotation technique is proposed. This is a method which aims at producing a better simple structure of factor loadings, using prior information about clusters of variables. In this paper, the proposed method is referred to as Obliclus (OBLique rotation technique with CLUStering of variables) procedure. Now,  $p, q$  and  $r$  denote the numbers of variables, factors, and clusters of variables, respectively. Let  $\mathbf{A}$  ( $p \times q$ ) be an initial (orthogonal) pattern matrix,  $\mathbf{T}$  ( $q \times q$ ) be a pattern transformation matrix, and  $\mathbf{\Lambda}$  ( $p \times q$ ) be a rotated pattern matrix, i.e.,  $\mathbf{\Lambda} = \mathbf{A}(\mathbf{T}')^{-1}$  which has  $\lambda_{ij}$  for an element  $(i, j)$ . Let  $\mathbf{U}$  ( $p \times r$ ) be a matrix whose  $(i, k)$  element  $u_{ik}$  is one if variable  $i$  is assigned to cluster  $k$  and zero otherwise. This indicator variable  $u_{ik}$  is referred to as cluster parameter. Let  $\mathbf{C}$  denote an  $r \times q$  matrix of centroids or mean values of squared loading values among a particular cluster, i.e., a  $(k, j)$  element  $c_{kj}$  is given by  $c_{kj} = \frac{1}{N_k} \sum_{i=1}^p u_{ik} \lambda_{ij}^2$ , where  $N_k$  is the number of variables which belong to cluster  $k$ .

Let  $\langle \mathbf{X}, \mathbf{Y} \rangle = \text{tr}(\mathbf{X}'\mathbf{Y})$  be the Frobenius product of matrices  $\mathbf{X}$  and  $\mathbf{Y}$  of the same size, where  $\text{tr}(\mathbf{M})$  denotes a trace of a square matrix  $\mathbf{M}$ . Also,  $\|\mathbf{M}\| = \langle \mathbf{M}, \mathbf{M} \rangle^{1/2}$  is the Frobenius



norm of  $\mathbf{M}$ . In addition,  $\mathbf{X}^{(2)}$  denotes an Hadamard product (an element-wise product) of two  $\mathbf{X}$ s, that is,  $\mathbf{X}^{(2)} = \mathbf{X} \odot \mathbf{X}$ . Then, a criterion for Obliclus rotation method is defined as

$$Q(\mathbf{T}, \mathbf{U}) = \langle \mathbf{\Lambda}^{(2)}, \mathbf{\Lambda}^{(2)} \mathbf{D} \rangle + \|\mathbf{\Lambda}^{(2)} - \mathbf{U} \mathbf{C}\|^2 \quad (1)$$

where  $\mathbf{D}$  is a  $q \times q$  matrix with zeros on the diagonal and ones elsewhere. This criterion is the sum of two optimization criteria for Quartimin rotation and the  $k$ -means-like clustering method. Based on this criterion, the rotation problem for Obliclus method is formulated as follows:

$$\min_{\mathbf{T}} Q(\mathbf{T}, \{u_{ik}\}_{i \notin \mathbf{I}_{\text{fixed}}} | \{u_{ik}\}_{i \in \mathbf{I}_{\text{fixed}}}) \quad \text{subject to} \quad \begin{cases} \text{diag}(\mathbf{T}'\mathbf{T}) = \mathbf{I}_q, \\ \sum_{i=1}^p u_{ik} \geq 1, \quad (k = 1, \dots, r), \\ \sum_{k=1}^r u_{ik} = 1, \quad (i = 1, \dots, p). \end{cases} \quad (2)$$

Here,  $\mathbf{I}_{\text{fixed}}$  denotes a set of indices of variables whose cluster parameters are fixed based on prior information. That is, in Obliclus procedure, the criterion will be minimized for a rotation matrix  $\mathbf{T}$  and non-fixed cluster parameters  $\{u_{ik}\}_{i \notin \mathbf{I}_{\text{fixed}}}$  simultaneously.

For minimization of  $Q$ , an alternating least squares (ALS) algorithm (de Leeuw, Young, & Takane, 1976) is developed. The ALS algorithm is implemented as follows:

*STEP1.* Initialize the parameters  $\{u_{ik}\}_{i \notin \mathbf{I}_{\text{fixed}}}$  subject to the restrictions given in Eq.(2).

*STEP2.* Minimize  $Q$  over  $\mathbf{T}$  for fixed  $\mathbf{U}$  using the GP algorithm (Jennrich, 2002).

*STEP3.* Minimize  $Q$  over  $\{u_{ik}\}_{i \notin \mathbf{I}_{\text{fixed}}}$  for fixed  $\mathbf{T}$  and  $\{u_{ik}\}_{i \in \mathbf{I}_{\text{fixed}}}$  using the  $k$ -means algorithm (Lloyd, 1982).

*STEP4.* Go to *STEP2*, or stop.

### 3. Artificial and real data examples

Two artificial data and Thurstone's box data were used to assess the property of Obliclus. The results indicate that Obliclus method can reproduce the perfect simple structure if it exists. And then, it is also indicated that Obliclus can reproduce more types of complicated loading structure than the other rotation methods can do. Moreover, Obliclus succeeded in reproducing the true structure of Thurstone's box data.

### References

- de Leeuw, J. and Young, F. W. and Takane, Y. (1967). Additive structure in qualitative data: An alternating least squares method with optimal scaling features. *Psychometrika*, 41, 471–503.
- Jennrich, R. I. (2002). A simple general method for oblique rotation. *Psychometrika*, 67, 7–19.
- Lloyd, S. P. (1982). Least squares quantization in PCM. *IEEE Transactions on Information Theory*, 28, 128–137.

## **Abstract**

This presentation has two parts. The first is titled:

### **Rotation Algorithms: From the beginning**

This is an historical account of the development of rotation algorithms from graphical methods to current analytic methods. Following the graphical methods, the introduction of rotation criteria led to the development of analytic methods for orthogonal rotation. Early analytic methods used criteria such as quartimax and varimax. Each criterion came with its own rotation algorithm. Each used pairwise methods that involved cycling through pairs of factors. Early criteria were all quartic criteria. The next step was to develop pairwise algorithms for general quartic criteria. Adding a line search later generalized these to arbitrary criteria. Turning to oblique rotation these began with applying rotation criteria to reference structures rather than to loading matrices. It was shown how to minimize the quartimin criterion by modifying the reference structure one column at a time. These methods were extended to the oblimin family and for some time were the only methods of oblique rotation. Harman called these indirect methods. Direct methods apply rotation criteria to loading matrices directly and minimize these over all oblique rotations of an initial loading matrix. A one-parameter representation for the result of rotating one factor in the plane of two was introduced. This was used to produce a pairwise algorithm for the oblimin family. Adding a line search later gave a direct oblique algorithm for arbitrary criteria. An alternative approach used gradient projection methods to give a direct orthogonal and oblique algorithms for arbitrary criteria. These did not require cycling through pairs of factors. The rotation algorithm problem is in a sense solved. At present there are simple very general and reliable algorithms for both orthogonal and oblique rotation. Free computer code for the gradient projection methods may be downloaded from <http://www.stat.ucla.edu/research/gpa>.

The second part is titled:

### **Exploratory bi-factor analysis**

Bi-factor analysis is a form of confirmatory factor analysis originally introduced by Holzinger. The bi-factor model has a general factor and a number of group factors. The purpose of this paper is to introduce an exploratory form of bi-factor analysis. An advantage of using exploratory bi-factor analysis is that one need not provide a specific bi-factor model a priori. The result of an exploratory bi-factor analysis, however, can be used as an aid in constructing a specific bi-factor model. Our exploratory bi-factor analysis is simply exploratory factor analysis using a bi-factor rotation criterion. This is a criterion designed to produce perfect cluster structure in all but the first column of a rotated loading matrix. Examples are given to show how exploratory bi-factor analysis can be used with ideal and real data. The relation of exploratory bi-factor analysis to the Schmid-Leiman method is discussed. An example shows the Schmid-Leiman method can fail seriously while the exploratory bi-factor method works perfectly.

Both orthogonal and oblique methods are considered. Because there are many more oblique rotations of an initial loading matrix than orthogonal rotations, one expects the oblique results to approximate a bi-factor structure better than orthogonal rotations and this is indeed the case. A surprising result arises when oblique bi-factor rotation methods are applied to ideal data.

# A bayesian exploratory factor analysis solution with the simplimax target

Sho Hashimoto \*

## 1 Introduction

In the factor analysis model, there is the following indeterminacy of the loading matrix  $\Lambda$ .

$$\mathbf{x} = \Lambda \mathbf{T} \mathbf{T}^{-1} \mathbf{f}' + \mathbf{e} = \mathbf{B} \mathbf{g}' + \mathbf{e}. \quad (1)$$

Subject to  $\text{diag}(\mathbf{T}^{-1} \mathbf{T}^{-1'}) = \mathbf{I}$ , the new loading matrix  $\mathbf{B}(= \Lambda \mathbf{T})$  and the new score vector  $\mathbf{g}(= \mathbf{f} \mathbf{T}^{-1'})$  are suitable for the factor analysis model. Therefore, we cannot obtain the unique interpretable solution of lambda. For solving this problem, we estimate the initial solution via some constraints and to rotate it to be interpretable.

In this study, we proposed an exploratory BFA model which use the simplimax target as the prior mean matrix without specifying number of zeros.

## 2 Prior Distributions of $\Lambda$

Following Press & Shigemasu (1989) and Shigemasu & Hoshino (2004), we decomposed the prior distribution as  $P(\Lambda, \mathbf{F}, \Psi) = P(\Lambda | \Psi) P(\Psi) p(\mathbf{F})$  and defined  $P(\Lambda | \Psi)$  as follows:

$$P(\Lambda | \Psi) \propto |\Psi|^{-\frac{m}{2}} \exp \left[ -\frac{1}{2} \text{tr} \{ \Psi^{-1} (\Lambda - \Lambda_0) \mathbf{H} (\Lambda - \Lambda_0)' \} \right], \quad (2)$$

Loading matrix  $\Lambda$  is distributed as  $N_{p \times m}(\Lambda_0, \mathbf{H}, \Psi)$ , where  $\Lambda_0$  is the mean matrix,  $\mathbf{H}$  is the row covariance matrix, and  $\Psi$  is the column covariance matrix. Here  $\mathbf{H} = n_0 \mathbf{I}_m$  implies the certainty of the prior knowledge with the hyperparameter  $n_0$  expressing the hypothetical sample size.

## 3 Decomposition of $\Lambda_0$

If a loading matrix is simple, one of its element is sufficiently large or near zero. Then, we consider that  $\Lambda_0$  consists of some zeros and sufficiently large values in the absolute sense. Therefore,  $\Lambda_0$  can be expressed as the Hadamard product of two matrices:  $\Lambda_0 = \mathbf{W} \odot \Lambda^*$ . Binary matrix  $\mathbf{W}(p \times m)$  indicates the positions of zeros in  $\Lambda_0$ . Matrix  $\Lambda^*(p \times m)$  indicates the values of nonzero elements of  $\Lambda_0$ . Noting

---

\*Graduate School of Human Sciences, Osaka University, 1-2 Yamadaoka, Suita, Osaka 565-0871, Japan

the  $\mathbf{\Lambda}_0$ 's simple structure, the prior distributions of each elements of  $\mathbf{\Lambda}^*$  are defined as the following normal distribution:

$$P(\lambda_{ij}^*|w_{ij}, s_{ij}) \propto \exp \left\{ -\frac{n_0^*}{2} (\lambda_{ij}^* - \mu_{ij})^2 \right\}, \quad (3)$$

where  $\mu_{ij} = w_{ij}\mu_{ij}(2s_{ij} - 1)$  and  $s_{ij}$  is the plus/minus sign indicator of  $\lambda_{ij}^*$ . Integer  $\mu$  and  $n_0^*$  are hyperparameters. Because we have no prior information about the positions of zeros and the signs of elements,  $p(w_{ij})$  and  $p(s_{ij})$  are assumed to be discrete uniform. This decomposition is the same as that is used in simplimax, however, we did not specify the number of zeros in  $\mathbf{W}$ , but the hyperprior mean value of a element of the prior mean matrix.

## 4 Simulation Study

To illustrate the effect of the relationship between real data size  $n$  and hypothetical number of subjects  $n_0$  in the parameter recovery and achievements of the simple structure, the simulated data were generated using the normal factor analysis model. Furthermore, to illustrate the effect of the relation between real data size  $n$  and hypothetical number of subjects  $n_0$  in the parameter recovery and achievement of the simple structure, we set  $3 \times 3$  combinations of numbers of real data ( $n = 25, 50$ , and  $100$ ) and number of subjects ( $n_0 = n/2, n$ , and  $2n$ ). The table demonstrates that as compared to rotation methods, true

Table 1: Indices of each dataset and method in the simulation study

sample size	indices	$n_0 = n/2$	$n_0 = n$	$n_0 = 2n$	varimax	simplimax
$n = 25$	CC	0.94	0.94	0.94	0.89	0.90
	LS	0.66	0.70	0.71	0.39	0.43
	SSE	1.18	1.28	1.64	2.30	2.08
$n = 50$	CC	0.98	0.97	0.97	0.93	0.97
	LS	0.78	0.80	0.76	0.50	0.55
	SSE	0.54	0.65	0.71	0.89	0.81
$n = 100$	CC	0.99	0.96	0.84	0.95	0.98
	LS	0.86	0.82	0.45	0.65	0.70
	SSE	0.28	0.72	3.40	0.57	0.52

loadings can be recovered sufficiently in the proposed model.

## References

- Kiers, H. A. L. (1994). SIMPLIMAX: Oblique rotation to an optimal target with simple structure. *Psychometrika*, **59**, 567-579.
- Press, S. & Shigemasu, K. (1989). Bayesian inference in factor analysis. *Contributions to probability and statistics*, 271-287.
- Shigemasu, K. & Hoshino, T. (2004). Bayesian procrustes solution. *Behaviormetrika*, **31**, 29-41.

# Exploratory factor analysis: modern aspects

November 11, 2011

Nickolay T. Trendafilov<sup>1</sup>

Department of Mathematics and Statistics, The Open University,  
Walton Hall, Milton Keynes MK7 6AA, UK  
E-mail: N.Trendafilov@open.ac.uk

This talk is about some modern aspects of exploratory factor analysis (EFA). They are modern in a sense that they consider extension of the classical EFA for analyzing data sets with more variables than observations ( $p > n$ ), which are typical in a number of modern applications as climate, finance and gene expression data analysis.

The talk is composed by two parts. In the first part the classical EFA is revisited shortly. It is pointed out that the classical EFA model presents observable random variables as a linear combination of two types of random variables, called common and unique factors. This EFA definition is restrictive in a sense that it complicates the estimation of all EFA unknowns. This gives a serious motive to propose an alternative definition in which the EFA is formulated as a specific data matrix decomposition. The immediate reward is the ability to construct a procedure for simultaneous estimation of all EFA unknowns by optimizing a least squares loss function. Another benefit from the new EFA definition is that a simple computational algorithm can be constructed which is based on singular value decomposition (SVD) only. This is extremely important feature of the newly defined EFA because it becomes computationally competitive to the very popular and versatile principal component analysis (PCA). In the same time the EFA decomposition provides better fit to the data than PCA.

Next, it is shown that the new definition of EFA can be naturally extended for analyzing data with more variables than observations, i.e.  $p > n$ . This case requires introduction of new more general constraints than those in the classical EFA. Thus, the resulting EFA is called generalized EFA (GEFA) because it is applicable to data of any format and contains the classical EFA as a partial case. It is shown that GEFA can be solved by the same simple alternating algorithm involving SVD only.

---

<sup>1</sup>Invited talk presented at the *Symposium on Life Sciences and Statistics*, 3–5 November 2011, Osaka, Japan, and supported by the Japan Society for the Promotion of Sciences.

The rotational freedom in classical EFA which is frequently used for simplifying the resulting factor loadings is also present in GEFA. However, if  $p \gg n$  the standard rotation methods become computationally unfeasible. It turns out that a possible alternative is doing independent component analysis of the factor scores. This approach is computationally very attractive, but its justification for arbitrary data remains to be studied.

The introduction of GEFA is completed with two well know numerical examples illustrating the two situation when  $n > p$  and  $p > n$ . It is shown that the GEFA solutions are in concordance with the standard EFA solutions.

The classical EFA and PCA are popular techniques for dimension reduction and investigating the structure of the data. However, these fundamentally different techniques are frequently confused, and the differences between them are obscured, because they give similar results in some practical cases.

In the second part of the talk we investigate conditions under which they are expected to be close to each other, by considering EFA as a matrix decomposition so that it can be directly compared with the data matrix decomposition underlying PCA. We propose a QR decomposition as an alternative to the usual SVD for PCA, which can be reasonable when analyzing large data with respect to time consuming and storage space. Correspondingly, we propose an extended version of PCA, called the EFA-like PCA, which mimics the EFA matrix decomposition in the sense that they contain the same unknowns. We provide iterative algorithms for estimating the EFA-like PCA parameters, and derive optimality conditions that have to be satisfied for the two techniques to give similar results. Throughout, we consider separately the cases  $n > p$  and  $p > n$ . Finally, the derived algorithms and matrix conditions are illustrated on the same same well known data sets, considered at the end of the first part of the talk.

Full details, proofs and more examples can be found in the following forthcoming papers.

## References

- Trendafilov, N., & S. Unkel (2011) Exploratory factor analysis of data matrices with more variables than observations, *Journal of Computational and Graphical Statistics*, to appear.
- Trendafilov, N., S. Unkel, & W. Krzanowski (2011) Exploratory factor and principal component analyzes: some new aspects, *Statistics and Computing*, 2011, to appear.

# Exploratory data-fitting factor analysis with clustering of individuals

Hironori Satomura\*

## Abstract

A method for simultaneously performing exploratory factor analysis and k-means clustering is proposed, which is achieved by an extension of factor analysis model with fixed both common and specific factors. In our strategy, it is avoided that tandem analysis problem that makes it impossible for us to interpret the effect of cluster structure. An efficient alternating least squares algorithm is developed. To illustrate the usefulness, some numerical analyses are conducted.

**Keyword:** *factor analysis; k-means clustering; singular value decomposition.*

Recently, exploratory factor analysis (EFA) model with fixed common and specific factors has been presented by de Leeuw (2004), and been elaborated by Unkel and Trendafilov (2010) and Adachi (submitted). The model can be described as

$$\mathbf{X} = \mathbf{F}\mathbf{\Lambda}' + \mathbf{S}\mathbf{\Psi} \quad (0.1)$$

where  $\mathbf{X}$  is a standardized  $n \times m$  data matrix,  $\mathbf{F}$  is an  $n \times p$  ( $p < m$ ) fixed common factor score matrix,  $\mathbf{\Lambda}$  is an  $m \times p$  matrix of common factor loadings,  $\mathbf{S}$  is an  $n \times m$  fixed specific factor score matrix and  $\mathbf{\Psi}$  is an  $m \times m$  diagonal matrix of coefficients called specificity (Unkel and Trendafilov, 2010; Adachi, 2011). It is assumed that

$$n^{-1}\mathbf{F}'\mathbf{F} = \mathbf{I}_p \quad (0.2)$$

$$n^{-1}\mathbf{S}'\mathbf{S} = \mathbf{I}_m \quad (0.3)$$

$$\mathbf{F}'\mathbf{S} = \mathbf{O}_{mp} \quad (0.4)$$

with  $\mathbf{I}_m$  and  $\mathbf{I}_p$ ,  $m \times m$  and  $p \times p$  identity matrix respectively, and  $\mathbf{O}_{mp}$  a  $m \times p$  matrix of zeros. All parameters are obtained by minimizing following least squares loss function

$$f(\mathbf{F}, \mathbf{\Lambda}, \mathbf{S}, \mathbf{\Psi}|\mathbf{X}) = \|\mathbf{X} - \mathbf{F}\mathbf{\Lambda}' - \mathbf{S}\mathbf{\Psi}\|^2 \quad (0.5)$$

under the assumptions described above. This is a matrix approximation problem in which data matrix  $\mathbf{X}$  is decomposed into common part  $\mathbf{F}\mathbf{\Lambda}'$  and specific part  $\mathbf{S}\mathbf{\Psi}$ , and the procedure can be viewed as an extension of principal component analysis (Adachi, 2011), rather than a tool for exploring causal relationships.

Because common and specific factors are treated as fixed parameter, the model is suitable for the case in which data are not randomly sampled from an infinite population, e.g., rows

---

\* Graduate School of Human Sciences, Osaka University, 1-2 Yamadaoka, Suita, Osaka, 565-0871, Japan, E-mail: satomura@hus.osaka-u.ac.jp, Tel: +81-6-6879-8053



of the  $\mathbf{X}$  denotes prefecture or states of a country, specified cigarette brands, and so. We often encounter that types of data in marketing science or sensory analysis. In those areas, researchers often carry out a preliminary factor analysis on  $\mathbf{X}$ , and the cluster analysis is then performed on the estimated factor scores. This two-step procedure has two problems.

The first is the problem of factor indeterminacy. Similar to the random factor model, optimal factor scores  $\widehat{\mathbf{F}}$  and  $\widehat{\mathbf{S}}$  cannot be determined uniquely (Unkel and Trendafilov, 2010; Adachi, 2011). How should we determine factor scores (with combinatorial optimization of choosing and permutating singular vectors that correspond zero singular values)? What kind of scores are suitable for cluster analysis?

The second is the problem called "tandem analysis". Particularly it has been warned by several authors (e.g., DeSarbo, et al., 1990; De Soete and Carroll, 1994; Vichi and Kiers, 2001) that the fact that estimated factor score  $\widehat{\mathbf{F}}$  via ordinary factor analysis procedure do not necessarily define a subspace that is most informative about the cluster structure in the data. Indeed, the problem is more than that, which means that we cannot get contribution ratio of cluster structure to data matrix.

As a solution to those problem, we propose a method for simultaneously performing factor analysis and k-means clustering. The model is described as:

$$\mathbf{X} = \mathbf{U}\mathbf{C}\mathbf{A}' + \mathbf{S}\mathbf{\Psi} + \mathbf{E} \quad (0.6)$$

where  $\mathbf{U}$  is an  $n \times k$  membership matrix and  $\mathbf{C}$  is a  $k \times p$  (assume  $k > p$ ) cluster centroid matrix.

Details of the algorithm for parameter estimation and numerical analysis were reported on the day.

## References

- Adachi, K. (2009). Joint procrustes analysis for simultaneous nonsingular transformation of component score and loading matrices. *Psychometrika*, **74**, 685-703.
- Adachi, K. (2011). A fixed factor analysis procedure as an extension of principal component analysis. *Proceedings of 58th World Statistics Congress of the International Statistical Institute*.
- Adachi, K. (submitted). Some contributions to data-fitting factor analysis with empirical comparisons to covariance-fitting factor analysis.
- de Leeuw, J. (2004). Least squares optimal scaling of partially observed linear systems. In K. van Montfort, J. Oud and A. Satorra (Eds.), *Recent developments of structural equation models: Theory and applications*. Dordrecht: Kluwer Academic Publishers, pp. 121-134.
- DeSarbo, W.S., Jedidi, K., Cool, K. and Schendel, D., (1990). Simultaneous multidimensional unfolding and cluster analysis: an investigation of strategic groups. *Marketing Letters*, **2**, 129-146.
- De Soete, G. and Carroll, J.D. (1994). k-means clustering in a low-dimensional Euclidean space. In: Diday, E., et al. (Eds.), *New Approaches in Classification and Data Analysis*. Springer, Heidelberg, pp.212—219.
- Unkel, S. & Trendafilov, N. T. (2010). Simultaneous parameter estimation in exploratory factor analysis: An expository review. *International Statistical Review*, **78**, 363-382.
- Vichi, M. & Kiers, H. A. L. (2001). Factorial k-means analysis for two-way data. *Computational Statistics and Data Analysis*, **37**, 49-64.

## 1. はじめに

本報告では、シンボリックデータ解析法の概要を報告した後、シンボリックデータの応用例として、放射線治療に関する数理モデルと治療計画について扱った。シンボリックデータ解析は、1980年代後半より、E. Diday によって提唱され、その後、多くの研究成果が公表されている。また、放射線治療は、外科的治療、化学的治療とならび、悪性腫瘍に対する標準的な対処法となっている。対象とする腫瘍に照射する線量および生体への影響を考察するとき、シンボリックデータ解析法における分布値データを考えるのが有効である。

## 2. シンボリックデータ解析とは

従来のデータ解析において取り扱う対象の記述方法としては、ベクトルや行列を代表とする「構造を有する数値の集まり」が中心であった。しかし、解析対象が、個体の集まりである場合や、対象間に論理的関係を有する場合などにおいては、そのような記述方法だけでは不十分である。そこで、解析対象をシンボリックデータと名付け、より広い概念とした。また、その記述方法としては、区間値、ヒストグラム、分布など多くの可能性がある。

## 3. シンボリックデータの定義

シンボリックデータ解析では広範囲なデータを扱うため、シンボリックデータの定義もかなり広い。簡単な定義としては、「解析対象の内部変動も記述したデータ」とされている。より厳密には、個体に対して集まりとしてのコンセプトを想定し、そのコンセプトを解析対象とする。そのとき、個体に関する記述をコンセプトに関する記述にまとめる。逆に、その記述に対応するものをシンボリックオブジェクトとよぶ。より形式的には、シンボリックオブジェクトとは、帰属関数、関係、記述の3つ組で表現される。

## 4. シンボリックデータ解析の手法

従来のデータ解析に対応して、シンボリックデータに対する解析法が多数、開発されている。多次元区間値データに対する主成分分析を例にとって紹介する。もっとも単純な方法は、Symbolic-Numerical-Symbolic とよばれる方法で、シンボリックデータ（ここでは、多次元区間値データ）を通常の多次元データに変換した後、通常の主成分分析を適用する。得られた次元縮小法により、オリジナルの多次元区間値データの次元を縮約する。代表的な手法としては、頂点法(Vertexes PCA)、中心法(Center PCA)などがある。その他、MRPCA(Midpoints Radii PCA)や Spaghetti PCA などの Hybrid approach、IPCA(PCA based on Interval Algebra)などの Symbolic-Symbolic-Symbolic アプローチがある。

## 5. 放射線治療の概要

悪性腫瘍に対する代表的な治療法として放射線の利用がある。外科的治療、化学的治療とも共通するが、根本的な考え方は、正常組織（特に、危険臓器）へのダメージを最小にししながら、腫瘍を削除することである。現在、体幹部定位放射線治療と呼ばれる方法では、呼吸や心拍による身体の動きを監視または制御することで、1mm以下の精度で放射線を照射できるようになっている。

#### 6. 放射線量に対する細胞の生存率

放射線に対する細胞または臓器の生存率のモデルについては、多くの研究がなされている。最も基本的なものは、**LQ model** であり、放射線治療における臨床でも広く利用されている。これは、吸収線量の2次式を引数とした指数関数により生存率を評価するものである。

#### 7. IMRT と分割照射

放射線治療として最近、普及しつつある強度変調放射線治療（**Intensity Modulated Radiation Therapy: IMRT**）では、正常組織への照射量を少なくすることができる。また、照射計画では、1回の照射で治療する方法、ある程度の期間をかけて複数回照射する方法（分割照射。例えば、週に5回、合計30回）などがある。腫瘍および正常組織の両方に関して **LQ model** を仮定した場合、1回の照射と分割照射のどちらが適切であるかは重要な問題である。

#### 8. おわりに

シンボリックデータ解析は、橋の強度、発電所の亀裂、経済活動など関する多くの応用が発表されている。本報告では、シンボリックデータ解析の概要を紹介した後に、放射線治療に関する適用可能性について報告をした。**IMRT** 治療において、**DVH**（**Dose Volume Histogram**）が治療計画における重要な情報となっている。これを分布値データとして扱うことが可能である。さらに、**LQ model** およびそれを拡張したモデルを仮定することで、治療の影響について評価することができる。

# シンボリックデータ解析における MDS について

寺田吉壱

大阪大学 大学院基礎工学研究科

シンボリックデータ解析においては、解析対象は個々の個体ではなくそれらの集まりからなる “concept” である。そのため、対象間の非類似度が単一の値だけでなく、区間値などによって表現される場合がある。例えば、従来の MDS では視覚化が困難な大規模非類似性データを concept ごとに集約する場合、2 つの concept 間の距離として複数の非類似性が得られるため、これらの最小値及び最大値からなる区間値を用いて集約することで区間値非類似性データが得られる。

区間値非類似性データ  $\mathbf{D}_I = \left( \left[ \xi_{ij}^{(L)}, \xi_{ij}^{(U)} \right] \right)_{n \times n}$  に対する MDS は、Denceux and Masson (2000) によって下記の Stress 関数の最小化問題として定式化されている。

$$\sigma^2(\mathcal{R}) = \sum_{i \leq j} \left( \xi_{ij}^{(L)} \min_{\mathbf{x}_i \in R_i, \mathbf{x}_j \in R_j} \|\mathbf{x}_i - \mathbf{x}_j\| \right)^2 + \sum_{i \leq j} \left( \xi_{ij}^{(U)} \max_{\mathbf{x}_i \in R_i, \mathbf{x}_j \in R_j} \|\mathbf{x}_i - \mathbf{x}_j\| \right)^2$$

ここで、 $\mathcal{R} = \{R_1, \dots, R_n\}$  は対象を表現するユークリッド空間上領域の集合である。区間値非類似性データに対する MDS としては、対象が hyperbox で表現されると仮定した I-Scal が挙げられる (Groenen et al, 2006)。

しかし、区間値非類似性データは、非類似度の分布に関する情報を損失するだけでなく、最大値及び最小値を用いて区間を構成するため外れ値などの影響を受けやすい。そのため、下側及び上側  $\alpha_k$  パーセント点 ( $k = 1, \dots, K$ ) を用いた  $K$  個の区間値非類似性データ  $\mathbf{D}_{I_k} = \left( \left[ \xi_{kij}^{(L)}, \xi_{kij}^{(U)} \right] \right)_{n \times n}$  を用いることが考えられる。このとき、 $\alpha_k > \alpha_l$  に対して、 $\left[ \xi_{kij}^{(L)}, \xi_{kij}^{(U)} \right] \subset \left[ \xi_{lij}^{(L)}, \xi_{lij}^{(U)} \right]$  が成り立つ。ここでは、このような非類似性データをパーセンタイル値非類似性データと呼ぶ。パーセンタイル値非類似性データに対する MDS は、下記の Stress 関数の最小化問題として定式化される。

$$\begin{aligned} & \sigma^2(\mathcal{R}_1, \dots, \mathcal{R}_K) \\ &= \sum_{k=1}^K \sum_{i \leq j} \left( \xi_{kij}^{(L)} \min_{\substack{\mathbf{x}_{ki} \in R_{ki}, \\ \mathbf{x}_{kj} \in R_{kj}}} \|\mathbf{x}_{ki} - \mathbf{x}_{kj}\| \right)^2 + \sum_{k=1}^K \sum_{i \leq j} \left( \xi_{kij}^{(U)} \max_{\substack{\mathbf{x}_{ki} \in R_{ki}, \\ \mathbf{x}_{kj} \in R_{kj}}} \|\mathbf{x}_{ki} - \mathbf{x}_{kj}\| \right)^2 \end{aligned}$$

ここで、 $\mathcal{R}_k = \{R_{k1}, \dots, R_{kn}\}$  は  $R_{ki} \subset R_{li}$  ( $k < l$ ) を満たす各対象を表現する  $\mathbb{R}^p$  上の領域の集合である。Groenen and Winsberg (2006) は、各対象が同じ中心点をもつ入れ子状の hyperbox で表現されると仮定したモデルに対するアルゴリズム Hist-Scal を提案している。また、Terada and Yadohisa (2011) では、Hist-Scal のモデルに対して、反復 majorization に基づくより良いアルゴリズム Percen-Scal を提案している。

本発表では、シンボリックな非類似性データの中でも区間値非類似性データ及びパーセンタイル値非類似性データに着目し、それらのデータに対する MDS について述べ、実データに対する適用例を紹介した。

区間値非類似性データに対する MDS としては、各対象が hyperbox で表現されると仮定した hyperbox model と Groenen et al. (2006) で提案された hyperbox model に対するアルゴリズム

“hyperbox model I-Scal”について紹介した．また，シンボリックな非類似性に対しては，有効なクラスタリング方法はこれまでに提案されていないことから，hyperbox model を仮定した MDS とクラスタリング法を組み合わせた方法である hyperbox model SYMCLUS を提案した．具体的には，MDS として hyperbox model，区間値データに対するクラスタリング法として，下記の目的関数を最小化する方法を考えた．

$$\sigma_C(\mathbf{C}; \mathbf{X}, \mathbf{R}, K) = \|\mathbf{X} - \mathbf{C}\mathbf{U}\|^2 + \|\mathbf{R} - \mathbf{C}\mathbf{V}\|^2$$

ここで， $\mathbf{X}$ ， $\mathbf{R}$  はそれぞれ区間値データの中心点と辺長を表す  $n \times p$  行列， $\mathbf{C}$  はクラスターへの帰属を表す  $n \times K$  行列， $\mathbf{U}$ ， $\mathbf{V}$  は，それぞれ各クラスターの中心点と辺長の平均を表す  $K \times p$  行列である．つまり，MDS とクラスタリング法の同時分析として，下記のストレス関数の最小化問題に対する反復 majorization に基づくアルゴリズムの提案を行った．

$$\sigma_{SC}(\mathbf{X}, \mathbf{R}, \mathbf{C}) = \sigma_M(\mathbf{X}, \mathbf{R}) + \lambda \sigma_C(\mathbf{C}; \mathbf{X}, \mathbf{R}, K)$$

ここで， $\sigma_M$  は区間値 MDS のストレス関数である．

さらに，パーセンタイル値非類似性データに対する MDS については，まず，各対象が同一の中心点をもつ入れ子状の複数の hyperbox で表現されると仮定した concentric hyperbox model と Terada and Yadohisa (2011) で提案された concentric hyperbox model に対する反復 majorization に基づくアルゴリズム “concentric hyperbox model Percen-Scal”を紹介した．

しかし，Concentric model では，中心点の同一性を仮定しているため，実際のデータには適さない場合がある．例えば，非類似性の外れ値が非類似性が大きい方向か小さい方向のいずれか一方にしかない場合，影響を受けやすいモデルとなっている．そこで，パーセンタイル値非類似性データに対して，中心点の同一性を仮定しない hyperbox model に対する MDS を提案した．具体的には，2 つの hyperbox が入れ子状になるための必要十分条件を明らかにすることで，(外点) ペナルティ関数及びその関数に対する majorizing 関数を導出し，ストレス関数にペナルティ関数を加えた拡張ストレス関数の最適化を行うための反復 majorization に基づくアルゴリズム “non-concentric hyperbox model Percen-Scal”を提案した．

## 参考文献

- [1] Denceux, T., & Masson, M. (2000). Multidimensional scaling of interval-valued dissimilarity data, *Pattern Recognition Letters*, **21**, 82–92.
- [2] Groenen, P. J. F., Winsberg, S., Rodrguez, O. & Diday, E. (2006). I-Scal: Multidimensional scaling of interval dissimilarities, *Computational Statistics & Data Analysis*, **51**, 360–378.
- [3] Groenen, P. J. F., & Winsberg, S. (2006). Multidimensional scaling of histogram dissimilarities, In Vladimir Bategelj, Hans-Hermann Bock, Anuška Ferligoj, & Aleš Žiberna (Eds.), *Data Science and Classification* (pp. 581–588), Heidelberg, Berlin: Springer-Verlag.
- [4] Terada, Y., & Yadohisa, H. (2011). Multidimensional scaling with hyperbox model for percentile dissimilarities, In Junzo Watada, Gloria Phillips-Wren, Lakhmi C. Jain, & Robert J. Howlett (Eds.), *Intelligent Decision Technologies* (pp. 779–788), Heidelberg, Berlin: Springer-Verlag.

## セッション「統計的正則化理論と神経生理学」の序

大阪大学 基礎工学研究科  
狩野 裕

人や動植物に代表される生命体に関わる現象を理解するために、様々な方法で計測が行われ統計的方法が適用されてきた。同時に、確率的なふるまいをする生命現象への取り組みは統計学の発展を促す大きな要因でもあった。それは個体差を記述するモデルと言ってよいのではないか。生物統計学、生物情報や心理統計学で開発された相関分析、生存時間分析や因子分析モデル(FA)、項目反応理論(IRT)等がその代表であろう。続いて登場したのは、個体内の変化を生物学的・機械学習的に捉えるモデル、即ち、ニューラルネットワーク(NN)であった。NN は脳内神経回路網における情報伝達を模倣することによって非線型の回帰モデルを構築した。

近年、計測機器・方法の進化と個別分野の発展によって、サイエンスは生命体における(因果)システムの解明へと迫っている。その結果、統計学の取り巻く環境が変化し、統計学にとって新たな市場が開けたと言ってよい。新たな市場といっても、それは従来の統計学の連続的な延長上にあるかどうかは分からない。

従来の枠組みでは識別不能・不適切な問題と考えられてきたものを正則化法で解くといったことがある。正則化法は「評価関数(e.g., 尤度) + ペナルティ項」によって推測するアプローチであるがその歴史は意外に古く 100 年近く遡れると云う。典型的には、平滑化やベイズ法は正則化法と見なすことができるし、因子分析における因子回転も同じような考え方をする。

近年、正則化法、特にスパース・コーディングはずいぶん成功しているように見える。その理由はどこにあるのか。正則化法で生みだされるスパースな解に、生物学的・脳科学的な妥当性が与えられる(可能性はある)、つまり、脳内の情報処理にスパース性が利用されているとすれば、どうだろうか。単にゼロが多い解が得られるという数理的な性質だけでなく、この事実は、スパース・コーディングは脳機能を模倣した推測方法であったことになる。近年の正則化法が成功している理由が分かったような気がする。

Olshausen and Field(2004, p.481, Current Opinion in Neurobiology) は、スパース・コーディングの特徴について以下のように述べている：

This strategy, referred to as 'sparse coding', could possibly confer several advantages. First, it allows for increased storage capacity in associative memories; second, it makes the structure in natural signals explicit; third, it represents complex data in a way that is easier to read out at subsequent levels of processing; and fourth, it saves energy.

脳機能は複雑で、スパース・コーディングやその他の情報処理が同時並行的に行われ、それらを集約し判断していると考えられている。たとえば、2次元網膜像から3次元形状を推定することについて、乾(2004)は、(1)両眼視差による立体視、(2)運動からの構造復元、表面輪郭からの表面方向の推定、(4)表面テクスチャーからの構造復元、(5)陰影からの形状復元、が並列独立に処理されそれらが統合される、と述べている。ここで、陰影からの形状復元において正則化法が用いられている。

統計科学は、脳の情報処理機能に学ぶことが多いのではないかと予想している。脳は如何にしてスパース・コーディングを実現しているのか、多くの情報をどのように統合しているのか、莫大な情報処理になぜ（大量の）エネルギーを必要としないのか、こういったことから、統計科学や機械学習は多くのことを学ぶことができ、それらを一般化させることができるのではないだろうか。

正則化は必ずしもスパース性を意味せず、たとえば、先に述べた「陰影からの形状復元」はスパース性とは無関係である。脳機能における正則化と言っても様々であり、研究テーマに事欠かない。正則化法に並び注目されている多変量解析法にカーネル法がある。正則化法はカーネル法を適切に適用するためにも登場する（赤穂，2008）。

Olshausen and Field (2004). Sparse coding of sensory inputs. *Current Opinion in Neurobiology*, 14, 481-487.

乾 敏郎（2004）「視覚計算とマルコフ確率場」『統計科学のフロンティア第2巻』（岩波書店）

赤穂昭太郎(2008). カーネル多変量解析. 岩波書店.

# カーネル法と正則化

赤穂 昭太郎  
産業技術総合研究所

2011 年 11 月 4 日

## 1 カーネル法と正則化

### 1.1 カーネル法とは

カーネル法は従来の有限次元の線形多変量解析手法を、一般には無限次元の再生核ヒルベルト空間上に拡張した物であり、その性能の高さから多くの応用に用いられるようになってきている [1, 3] . 具体的には、再生核  $k: \mathcal{X} \times \mathcal{X} \rightarrow \mathcal{R}$  を用いて

$$f(x) = \sum_y \alpha_y k(y, x) \quad (1)$$

のパラメータ  $\{\alpha_y\}$  を何らかの規準で最適化する .

### 1.2 カーネル法と正則化

カーネル法は一般に無限個の基底を持つことから、解くべき問題が往々にして「不良設定問題」となり、解が不安定となる . それを安定化させるための手段が正則化であるが、基本となるのは Tikhonov 正則化である . サンプル集合  $x_1, \dots, x_n$  を  $f$  で処理した結果の損失関数  $R(f(x_1), \dots, f(x_n))$  に対し、ある条件を満たす  $\Omega(f)$  を重み付きで加算した

$$\min_f R(f(x_1), \dots, f(x_n)) + \lambda \Omega(f) \quad (2)$$

という最小化問題が安定となり、更に多くの場合  $n \rightarrow \infty$  のとき、適切に  $\lambda \rightarrow 0$  とすれば  $f$  が consistent になることが示される .

カーネル法と正則化の関係では「Representer 定理」が本質的に重要な役割を果たす .  $\Omega(f)$  が再生核ヒルベルト空間のノルム  $\|f\|$  の単調増加関数であれば、最小化問題の解はサンプルのカーネル関数を基底として

$$f(x) = \sum_i \alpha_i k(x_i, x) \quad (3)$$

と表現できることを主張する . これにより、無限次元の最適化が実質的には有限次元の最適化問題に帰着される .

### 1.3 正則化のいろいろなタイプ

正則化は空間への制約であり、Tikhonov 正則化以外にもいろいろなやり方がある . 例えば、ノルム  $\|f\|$  を一定にするというような条件下で損失関数を最小化する問題であり、Lagrange の未定乗数法の形で書けば

$$R(f(x_1), \dots, f(x_n)) + \lambda \|f\| \quad (4)$$

の極値問題を解くことに帰着され、一見 Tikhonov 正則化に類似しているが、 $R$  の形によっては Representer 定理が必ずしも成立しない .

### 1.4 計算的側面

カーネル法は Representer 定理が成り立つ場合でもサンプル数程度のパラメータの自由度があるため、最適解を求める手法が効率的であることが重視される . その第一の柱となるのが凸性であり、損失関数や正則化項がパラメータを凸関数に限定することによって最適化の効率を確保する (伝統的な多変量解析は幸いほとんど凸である) .

さらに、正則化においては正則化パラメータ  $\lambda$  の決定が重要な問題となる . そのためには複数のモデルからモデル選択規準に従って選ぶことになるが、その際一般に数多くの最適化問題を解く必要がある .

代表的なモデル選択規準として Cross-validation があるが、重回帰分析のように解が従属変数の線形関数として与えられる場合には、全サンプルを用いた最適化を一度行うだけですべての場合の値を計算できるというメリットがある .

サポートベクトルマシン (SVM) などでは一般にそこまで単純にはいかないが、最近注目されているのが「パス追跡」[4] と呼ばれる手法であり、 $\lambda$  を 0 から (あるいは  $\infty$  から) 順に変化させていったときに、解が区分線形関数で変化することに着目し、すべての  $\lambda$  に対する解を効率的に求めることができる .



## 1.5 スパース性

Representer 定理により有限のパラメータ数に抑えることはできても、一般にそれではまだパラメータ数が大きすぎて最適化も  $f(x)$  の計算も計算量が多くなるという問題がある。そこで、多くの  $\alpha_i$  が 0 となるようなスパースな解を求めることが求められている。

SVM では正則化項ではなく、損失関数を区分線形関数にすることによりスパース性を確保している。この場合、クラスの主要な部分をなすサンプル (つまり誤差が出ないようなサンプル) は  $f$  に全く寄与しなくなる点が関数の解釈などを困難にすることもある。

一方、正則化によってスパース性を確保するという研究の流れが近年活発に行われている。その元となるものは lasso [6] であり、これは  $L_1$  ノルム  $\sum_i |\alpha_i|$  を正則化項として用いるもので、非常にスパースな解が得られることが知られている。ただし、これは本来 Representer 定理の仮定からはみだすことには注意する必要がある。

なお、 $\{\alpha_i\}$  をスパースにするということはサンプルを選ぶということであるが、カーネルの場合はこれが基底にもなっているので変数選択や特徴選択といった次元縮約の手法とも関連している。

$L_1$  正則化はスパースな解を求めるが、スパースになりすぎるという欠点があった。これを解決する方法として  $L_1$  正則化と通常の Representer 定理の成り立つ  $L_2$  正則化を折衷し、 $L_2$  ノルムの 1 乗と 2 乗を重み付けて足した  $a\|f\| + b\|f\|^2$  といった正則化項が考案され (elastic net[9])、性能のよいものとしてこのバリエーションが多く研究されている。

lasso そのものの拡張では、サンプル間に類似性などの構造がある場合にそれを積極的に正則化のモデルに取り入れる試みが多数行われている。代表的な物として、何らかの意味で隣接するサンプル間の  $\alpha_i$  の値を近くするために  $|\alpha_i - \alpha_j|$  といった項を加える fused lasso[7] や、グループ単位でスパースな構造を取り出すためにグループ毎の  $L_2$  ノルムの 1 乗和を正則化項とする group lasso[5] などがある。

## 2 カーネル劣成分分析の諸問題

空間中の点に曲線や曲面を当てはめ場合に、非線形写像で移した空間での線形当てはめを利用する方法が用いられる。これは、主成分分析が共分散行列の固有値の大きい軸を取り出すのに対して、小さくなる軸を取り出すため劣成分分析と呼ばれる。有限次元の場

合は主成分分析の補空間として劣成分分析をとらえることができるが、カーネル劣成分分析は Representer 定理が成立しないことと関連して不思議な振る舞いを示すことが示されている。

## 3 部分空間制約によるカーネル行列の補完

未計測のカーネル行列を補完する問題は広い応用分野を持つが、ここでは補助情報を用いた手法を紹介する [8]。一般にこうした最適化問題は半定値計画問題 (SDP) という、効率的に解ける物の中では計算量の多い問題に帰着される。しかしながら、正定値行列の幾何の導くいくつかの部分空間を制約条件として入れ込むと、簡単な繰り返し法によって解を求めることができることが示される。

## 参考文献

- [1] 赤穂昭太郎. カーネル多変量解析 — 非線形データ解析の新しい展開. 確率と情報の科学. 岩波書店, 2008.
- [2] H.D. Bondell and B.J. Reich. Simultaneous regression shrinkage, variable selection, and supervised clustering of predictors with oscar. *Biometrics*, Vol. 64, No. 1, pp. 115–123, 2007.
- [3] 福水健次. カーネル法入門—正定値カーネルによるデータ解析—. 多変量データの統計科学 8. 朝倉書店, 2010.
- [4] T. Hastie, S. Rosset, R. Tibshirani, and J. Zhu. The entire regularization path for the support vector machine. *Journal of Machine Learning Research*, Vol. 5, pp. 1391–1415, 2004.
- [5] L. Meier, S. van de Geer, and P. Bühlmann. The group lasso for logistic regression. *J. R. Stat. Society B*, Vol. 70, No. 1, pp. 53–71, 2008.
- [6] R. Tibshirani. Regression shrinkage and selection via the lasso. *J. R. Statist. Soc. B*, Vol. 58, No. 1, pp. 267–288, 1996.
- [7] R. Tibshirani, M. Saunders, S. Rosset, J. Zhu, and K. Knight. Sparsity and smoothness via the fused lasso. *J. R. Stat. Society B*, Vol. 67, No. 1, pp. 91–108, 2004.
- [8] K. Tsuda, S. Akaho, and K. Asai. The em algorithm for kernel matrix completion with auxiliary data. *Journal of Machine Learning Research*, Vol. 4, pp. 67–81, 2003.
- [9] H. Zou and T. Hastie. Regularization and variable selection via the elastic net. *J. R. Stat. Society B*, Vol. 67, No. 2, pp. 301–320, 2005.

# 正則化法に基づくスパース推定と調整パラメータの選択

廣瀬 慧

大阪大学大学院基礎工学研究科

## 1. はじめに

生命科学，システム工学など諸科学の様々な分野で応用されている  $L_1$  型正則化法に基づくスパース推定は，モデリングの過程においてペナルティの強さを調節するパラメータ（チューニングパラメータ）の推定が本質的となる．チューニングパラメータの選択はモデルの評価・選択と捉えることができる．特に，線形回帰モデルにおいて，一般化自由度 (Ye, 1998) に基づくモデル評価基準は，標本数に関する漸近理論を用いることなく導出されているため，変数の次元がサンプル数を超えた場合においても理論的に適用可能である．しかしながら， $L_1$  型正則化法では，正則化項に微分可能でないパラメータの絶対値が含まれるため，一般化自由度を解析的に求めることが困難となる．本報告では，この問題に対処するため，線形回帰モデルにおける正則化法の推定アルゴリズムである GPS アルゴリズム (Generalized Path Seeking; Friedman, 2008) を拡張し，一般化自由度を数値的に計算するアルゴリズムを提案する．提案する計算アルゴリズムは，その推定値（解）を求めると同時に，一般化自由度も計算することができるという特徴を有する．数値実験を通して提案手法の特徴と有効性を検証する．

## 2. 設定

$p$  次元説明変数ベクトル  $(x_1, \dots, x_p)^T$  と目的変数  $y$  に関する  $N$  組のデータ  $\{y_i, x_{i1}, \dots, x_{ip}\}_{i=1}^N$  が得られたとする．このとき，次のガウス型線形回帰モデルを考える．

$$y_i = \beta_0 + \sum_{j=1}^p \beta_j x_{ij} + \varepsilon_i, \quad i = 1, \dots, N.$$

以下，説明変数は基準化され，目的変数は中心化されているとし，一般性を失うことなく  $\beta_0 = 0$  とする．このモデルを  $L_1$  型正則化法

$$\hat{\beta} = \arg \min_{\beta} \hat{R}(\beta) \quad \text{s.t.} \quad P(\beta) \leq t \quad (1)$$

によって推定する．ただし， $\beta = (\beta_1, \dots, \beta_p)^T$ ， $\hat{R}(\beta) = \sum_{i=1}^N (y_i - \sum_{j=1}^p \beta_j x_{ij})^2 / N$ ， $P(\beta)$  はペナルティ項とし，たとえば，lasso ペナルティ (least absolute shrinkage and selection operator, Tibshirani, 1996) は  $P(\beta) = \sum_{j=1}^p |\beta_j|$  で与えられる．また， $t$  はチューニングパラメータであり，この値を次節の一般化自由度によって推定する．

## 3. 一般化自由度

モデリングプロセス  $m(\cdot)$  が与えられたとき，真の平均構造をモデル  $\hat{\mu} = m(y)$  によって推定することを考える．ここで， $\hat{\mu} = (\hat{\mu}_1, \dots, \hat{\mu}_N)^T$  は  $L_1$  型正則化法によって推定されたモデルである．このとき，一般化自由度 (GDF) は次式で与えられる．

$$\text{GDF} = \sum_{i=1}^N \frac{\text{cov}(\hat{\mu}_i, y_i)}{\tau^2} \quad (2)$$

ただし， $\tau^2$  は  $y_i$  の真の分散とする．一般化自由度は，モデルの複雑さを表し，たとえば赤池情報量規準や Mallows の  $C_p$  基準における第 2 項目のパラメータ数に対応する．この一般化自由度を導

出するためには、推定量  $\hat{\mu}$  を解析的に求める必要がある。しかしながら、 $L_1$  型正則化法では、正則化項にパラメータの絶対値が含まれるため、一般化自由度の解析的な導出が困難となる。そこで、GPS アルゴリズムを拡張して一般化自由度を数値的に推定するアルゴリズムを提案する。

#### 4. GPS アルゴリズムに基づく一般化自由度の推定

GPS アルゴリズム (Friedman, 2008) は、(1) 式の解を逐次的に求める高速なアルゴリズムである。まず、推定値  $\hat{\beta}$  の初期値を  $\hat{\beta} = \mathbf{0}$  と設定する。  $\Delta t$  を十分小さい正数としたとき、チューニングパラメータが  $t$  のときの推定値  $\hat{\beta}(t)$  から、  $t + \Delta t$  の推定値  $\hat{\beta}(t + \Delta t)$  を求める。GPS アルゴリズムでは、次式のように 1 回の更新ステップでパラメータベクトルの要素の 1 つのみを更新することにより、解を近似的に求める。

$$\begin{aligned}\hat{\beta}_k(t + \Delta t) &= \hat{\beta}_k(t) + \Delta t \cdot \text{sign}(\hat{\beta}_k(t)), \\ \{\hat{\beta}_j(t + \Delta t) &= \hat{\beta}_j(t)\}_{j \neq k}.\end{aligned}$$

ただし、

$$k(t) = g_k(t)/p_k(t), \quad g_k(t) = -[\partial \hat{R}(\beta)/\partial \beta_k]|_{\beta=\hat{\beta}(t)}, \quad p_k(t) = [\partial \hat{P}(\beta)/\partial |\beta_k|]|_{\beta=\hat{\beta}(t)}$$

とし、  $k = \arg \max_j \{|\hat{\beta}_j(t)|\}$  とする。ただし、厳密にはこの  $k$  の定義はパラメータの絶対値が単調増加するという条件の下で成り立つもので、一般には成り立たない (Friedman, 2008)。

GPS アルゴリズムは、あらゆるチューニングパラメータ  $t$  に対する推定値の近似値を求めることを可能とする。しかしながら、それらの解の候補から、適切なモデルを選択することはできない。そこで、(2) 式の一般化自由度を次のように逐次的に更新するアルゴリズムを提案する。

$$M(t + 1) = M(t) + \frac{2\Delta t}{N|g_k(t)|} \mathbf{x}_k \mathbf{x}_k^T \{I - M(t)\}$$

ここで、  $I$  は単位行列、  $\mathbf{x}_k = (x_{1k}, \dots, x_{Nk})^T$ 、  $M(t) = \text{cov}(\hat{\mu}(t), \mathbf{y})/\tau^2$  である。一般化自由度から導出されるモデル評価基準を用いて、適切なモデルを選択する。

#### 5. パッケージ msgps

msgps (Model Selection criteria via Generalized Path Seeking) は、 $L_1$  型正則化法における一般化自由度を求め、様々なモデル評価基準によって選択された解を出力する統計パッケージである。このパッケージは統計ソフトウェア R (R Development Core Team, 2010) によって実装され、<http://www.keihiro.se.com/research/msgps.html> から入手可能である。本パッケージは、サンプルサイズが大きい場合においても高速に計算することができるよう、本稿で述べた計算アルゴリズムをさらに改良している。詳細については、Hirose *et al.* (2011) を参照されたい。

#### 参考文献

- Friedman, J. (2008). Fast sparse regression and classification. Technical report, Department of Statistics, Stanford University, 2008.
- Hirose, K., Tateishi, S. and Konishi, S. (2011). Efficient algorithm to select tuning parameters in sparse regression modeling with regularization. arXiv:1109.2411 (arXiv).
- R Development Core Team (2010). R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL <http://www.R-project.org/>.
- Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *J. Roy. Statist. Soc. Ser. B*, **58**, 267–288.
- Ye, J. (1998). On measuring and correcting the effects of data mining and model selection. *J. Amer. Statist. Assoc.*, **93**, 120–131.

## 学習神経回路における活動同期現象について 運動学習と強化学習における誤差信号の神経生理学的探索

大阪大学大学院・生命機能研究科・社会経済研究所・JST さきがけ・小林康

脳における学習理論の研究の原点は運動制御誤差学習理論にあります。われわれは運動イメージと実現された運動の差、「運動制御誤差」を最小化するように神経回路を変更させることにより最適な運動を学習します。この学習の座が小脳にあることが理論的にも実験的にも確立されてきました。

わたしたちは京都大学の河野憲二先生らと共同で「視覚像が動いたときの視野の安定化」に役立っていると思われる、広い視野の像の動きによって像と同方向に誘発される「追従眼球運動」時に、サル小脳皮質腹側傍片葉よりプルキンエ細胞のニューロン活動を記録しました。実験の結果、小脳でのシナプス学習に必要な「運動誤差」が延髄下オリーブ核由来の登上線維入力によって与えられていることが明らかになりました。さて現在、脳における学習理論研究のトピックは強化学習に移ってきています。手がかり刺激に対して行動の報酬を予測し、過去に形成された予測と実際に得られた報酬との差、「報酬予測誤差」を最小化するように予測を更新することにより、行動を最適化するのが強化学習理論です。現在、中脳にありますドーパミン作動性神経細胞（ドーパミンを伝達物質として放出する細胞、ドーパミン細胞）が報酬予測誤差を表現しているということがほぼ確立されていますが、計算理論の重要な鍵となる誤差信号の計算メカニズムについてはその実体が明らかにされていません。

中脳ドーパミン細胞の機能を知るために、解剖学的研究で知られている既知の神経回路網を眺めてみますと、中脳ドーパミン細胞は大脳基底核などから抑制性、大脳皮質などから興奮性の入力を受けますが、中脳のアセチルコリン作動性の脚橋被蓋核（PPTN）がドーパミン細胞に対してもっとも強力な興奮性入力を送っています。このことは、すでに知られている神経回路構造から見てPPTNが中脳ドーパミン細胞による報酬予測誤差計算の中心であることを示唆します。わたしたちはこの神経回路構造に基づく仮説に基づいて、サルを用いて報酬予測課題中のPPTNからのニューロン活動記録実験を行いました。

実験の結果、脳内報酬予測の更新に必要な、記憶された予測報酬の情報と実際に得られた報酬の情報が、それぞれ分離独立した形でサルPPTNに表現されているということが明らかになりました。この研究によって、報酬予測誤差計算にPPTNが重要な役割を果たしており、さらに中脳ドーパミン細胞での報酬予測誤差計算過程で、短期記憶された報酬予測情報と実報酬情報がPPTNでそれぞれ同時表現され、最終的にドーパミン細胞で予測と現実の誤差が計算されるという

ことが明らかになりました。最後にこれらの2つの「一見異なる学習システム」の話を総括し、神経細胞による情報表現、随所に顔をのぞかせるニューロン活動の同期現象についてお話しできればと思います。

#### 参考文献

Okada K, Toyama K, Inoue Y, Isa T and **Kobayashi Y**

Different Pedunculopontine Tegmental Neurons Signal Predicted and Actual Task Rewards. *The Journal of Neuroscience* **29**: 4858-4870, 2009

PPTN が報酬予測誤差の計算に関係していることを示した神経生理学的研究

**Kobayashi Y**, Inoue Y, Yamamoto M, Isa T and Aizawa H

Contribution of Pedunculopontine Tegmental Nucleus Neurons to Performance of Visually Guided Saccade Tasks in Monkeys. *J Neurophysiol* **88**: 715-731, 2002.

PPTN が運動制御、覚醒などに関係していることを示した神経生理学的研究

Yamamoto K, **Kobayashi Y**, Takemura A, Kawano K, and Kawato M

Computational Studies on Acquisition and Adaptation of Ocular Following Responses Based on Cerebellar Synaptic Plasticity. *J Neurophysiol* **87**: 1554-1571, 2002

小脳による眼球運動制御を神経生理データに基づき計算機シミュレートした研究

**Kobayashi Y**, Kawano K, Takemura A, Inoue Y, Kitama T, Gomi H, Kawato M

Temporal Firing Patterns of Purkinje Cells in the Cerebellar Ventral Paraflocculus During Ocular Following Responses in Monkeys. II. Complex Spikes. *J. Neurophysiol* **80**: 832-848, 1998

一般化線型モデルなどを用いて小脳による運動制御機構を示した神経生理学的研究

# 予測的処理の神経生理学

乾 敏郎

京都大学大学院情報学研究科

## 1. 脳内での予測的処理

運動指令が出てから手が動くまでの時間は0.1秒以上ある。その手が動いたという確認は、自分の手を見ればよい。だが、視覚情報は中枢に到達するまでにやはり遅れがある。おそらく、視覚においても予測的に見ているのだろう。運動指令が出てから自分の感覚フィードバックを使えるまでに200–300ミリ秒位の遅延があり、さらに、さまざまなノイズが神経系には存在する。したがって、予測を正しく行う、予測制御が重要となる。自分の運動指令を頭の中でコピーして自分の手のモデルを利用し、手はこう動くはずだということをどこかで無意識にモニターしているとする。しかし、それは自分の身体のモデルだけ使っているので、ノイズなどの影響で正確には予測できない。そこで、実際に手が動く結果生じた感覚フィードバックでチェックする。この感覚フィードバックは遅れて伝わってくるので、少し過去に戻って差分をとる。この誤差を利用し、予測を更新する。

以上のような考えを背景に、Ogawa & Inui(2007)は、自己運動と外部運動の予測・推定における神経基盤の違いを検討した。その結果、前補足運動野は自己、他者運動に依存しない視覚運動の内的なイメージ化に関連し、後部頭頂皮質における左下頭頂小葉(IPL)、右SPL/IPLがそれぞれ自己、他者の内的な運動推定に関与することが示唆された。左IPLでは、状態予測、および体性感覚野からの自己受容感覚フィードバックに基づく状態推定が行われると考えられる。また我々の先行研究から、この状態推定と視覚フィードバックとの誤差が右頭頂間溝(IPS)において評価され、右側頭頭頂接合部(TPJ: Temporo-Parietal Junction)が視覚的運動誤差と状態推定との適切な統合に関連していることが示唆されている(Ogawa, Inui, & Sugio, 2006; Ogawa, Inui, & Sugio, 2007)。右のTPJは視点変換にも重要な部位であり、ここが障害を受けると様々な視点変換を伴う身体の錯覚（たとえば、幽体離脱や自己幻視）が生じることが知られている。またこれらの結果は、我々がこれまで行ってきた腕や手指の模倣や道具使用のイメージ化の結果とよく対応している(Inui, et al., 2000; Tanaka, et al., 2001; Tanaka & Inui, 2002; 中村ら, 2003; Imazu, et al., 2007)。また到達把持運動の実験とモデルともよく対応する。

## 2. 到達把持運動における予測制御

到達把持運動では、運動中の視覚遮断により指が通常より大きく開く(Fukui & Inui, 2006)。つまり、到達把持はオンラインで制御されているが、運動指令が実際の運動として効果器(手や指)に現れるまでや、運動結果が視覚・自己受容感覚としてフィードバックされるまでには遅延やノイズが存在する。これを補償するために、我々の脳内では運動指令の結果としての現在の効果器の状態について予測が行われ、感覚フィードバックが得られたときに予測誤差を計算することで予測を修正し、運動予測と感覚が統合される。このようにして予測(推定)された効果器の情報は、運動指令の遠心性コピーにより再び更新され、次の状態予測となる。この枠組みはカルマンフィルタ(予測誤差にカルマンゲイン $K$ をかけて予測を修正する)を用いてモデル化され、現在・未来の

状態推定・予測および推定・予測精度（誤差分散情報）がオンライン制御に利用される制御モデルを提案した (Fukui, Takemura, & Inui, 2006; Takemura, Inui, & Fukui, 2007).

### 3. 大脳視覚皮質の計算理論・階層的ベイズ推定・誤信念

視覚系の大きな役割は網膜像から世界の状態を推定することである。この過程は、ベイズ推定の枠組みで定式化され、脳内の神経回路で実現する枠組みが川人と乾(1990)によって提案され、その後 fMRI 実験などによって、部分的に実証されている。この理論では、大脳皮質の低次から高次への結合により、近似逆変換がなされ、高次から低次への結合により、順変換を行うことにより、仮説を検証するという計算を行っている。そして、トップダウン信号とボトムアップ信号の誤差を最小化することにより、正しい推定が可能になることが示されている。脳の計算は、基本的に階層的ベイズ推定の枠組みでとらえることが可能だと考えられる(乾, 2004)。ベイズ推定は (事後確率)  $\propto$  (事前確率)  $\cdot$  (条件付確率) と書ける。事前確率は先験的知識である。事後確率はある仮説が成立つ確率を意味している。この枠組の基づき、誤った知覚から世界に対する誤った信念が生じることが指摘されており、これがドーパミン系の働きによることも指摘されている。そして右前頭前野背外側に信念を書き換える働きをする部位が同定されている。統合失調症の妄想スコアが高い場合、この部位が障害されており信念を書き換えることが困難となる。

### 4. 心的回転の脳モデル

我々は、自己中心座標で表現された物体の心的イメージが、自己の仮想的な運動によって回転される認知過程が心的回転(mental rotation)であると仮定し、神経メカニズムの解明を試みている。最近の研究では、頭頂葉の CIP および AIP と呼ばれる領域が、それぞれ対象の自己中心的立体表現、視覚と運動の統合に重要な役割を果たしていると考えられている。多くのイメージング研究で、心的回転課題時の運動関連領域の活動が示されていることから、心的回転は運動制御における内部モデルと関連付けて考えられている。ここで内部モデルとは、運動指令信号を入力とし、運動に対する環境フィードバックの予測を出力とするモジュールを指す。すなわち、自己の仮想的運動によって内部モデルが駆動され、視覚フィードバックの予測が形成される処理が、心的回転のようなイメージ変換課題と関係していると考えられる(Hinkle & Connor, 2001)。我々は、これらの知見を踏まえた神経回路モデルを提案し、提案モデルが三次元心的回転を部分的に再現可能であることを示した(Inui and Ashizawa, 2009)。同時に、CIP 層に立体勾配に選択的なユニットが、そして AIP 層に立体勾配に選択的な visual-dominant ユニットや visual-motor ユニットが学習によって創発した。これらのユニットと同様の振る舞いをするニューロンが、実際に CIP と AIP に存在することが知られている(Nguyenkim and DeAngelis, 2003; Sakata, et al., 1993)。我々の結果は、心的回転時に、CIP では対象の自己中心的三次元構造が表現され、AIP では CIP で表現される情報が統合されるという我々の仮説を支持する。さらに最近になって、我々のグループは、心的回転中に、右頭頂部の電極における  $\beta$  帯域 (14.7Hz) の脳波パワーと心的回転の回転量に有意な負の相関を見いだした。この電極における 14.7Hz の脳波パワーと BOLD 信号(blood oxygenation level dependent signals)が相関を示す脳部位を調べた結果、右前・後部頭頂間溝 (AIP, CIP)、左 IPL および左運動前野において有意な負の相関が見られた。この結果は我々のモデルを強く支持するものである。

# Dual saddlepoint equalities in model averaging

九州大学・経済学研究院 大西 俊郎  
中央大学・理工学部 柳本 武美

## 1. 問題設定

Bayesian model averaging の枠組み (Hoeting *et al.*, 1999) で次の Bayes モデルを考える.

$$p(x; \theta, k) = p_k(x; \theta), \quad \pi(\theta, k) = \lambda_k \pi_k(\theta).$$

ただし,  $k$  は離散的であり  $1, \dots, K$  の値をとるものとする ( $K \geq 2$ ). これは  $K$  個の Bayes モデル

$$p_k(x; \theta), \quad \pi_k(\theta) \quad (k = 1, \dots, K)$$

において  $k$  番目のモデルに  $\lambda_k$  の重みをつけることによって得られる ( $\sum_{k=1}^K \lambda_k = 1$ ).

この Bayes モデルの下で予測問題

$$\min_{q(y|x)} E[D(q(y|x), p(x; \theta, k)) \mid p(x; \theta, k)\pi(\theta, k)] \quad (1)$$

を議論する. ここで,  $D(q(y|x), p(x; \theta, k))$  は Kullback-Leibler ダイバージェンス型の損失であり,  $e$ -ダイバージェンス損失と呼ばれることがある. ベースになった  $K$  個の Bayes モデルでの予測問題は

$$\min_{q(y|x)} E[D(q(y|x), p_k(x; \theta)) \mid p_k(x; \theta)\pi_k(\theta)] \quad (k = 1, \dots, K) \quad (2)$$

である. (1) および (2) の最適解はそれぞれ,

$$\begin{aligned} q^{(e)}(y|x) &\propto \exp \{E[\log p(y; \theta, k) \mid \pi(\theta, k|x)]\}, \\ q_k^{(e)}(y|x) &\propto \exp \{E[\log p_k(y; \theta) \mid \pi_k(\theta|x)]\} \end{aligned} \quad (3)$$

で与えられる (Corcuera and Giummole, 1999). これは  $e$ -mixture 予測分布と呼ばれる.

上の問題と並行して, Bayes 予測問題

$$\begin{aligned} \min_{q(y|x)} E[D(p(x; \theta, k), q(y|x)) \mid p(x; \theta, k)\pi(\theta, k)] \\ \min_{q(y|x)} E[D(p_k(x; \theta), q(y|x)) \mid p_k(x; \theta)\pi_k(\theta)] \end{aligned} \quad (4)$$

を議論する. ここの損失は, (1) および (2) で採用された損失と双対であり,  $m$ -ダイバージェンス損失と呼ばれることがある. 最適解はそれぞれ

$$\begin{aligned} q^{(m)}(y|x) &= E[p(y; \theta) \mid \pi(\theta, k|x)], \\ q_k^{(m)}(y|x) &= E[p_k(y; \theta) \mid \pi_k(\theta|x)] \end{aligned} \quad (5)$$

で与えられる (Aitchison, 1975). これはいわゆる分布の mixture であるが,  $e$ -mixture との対比で  $m$ -mixture 予測分布と呼ばれることがある.

本発表には 2 つの目的がある. 1 つは,  $e$ -ダイバージェンス損失と  $m$ -ダイバージェンス損失の下での議論における著しい類似性を明らかにすることである. 対数尤度と  $e$ -ダイバージェンス損



失がペアとなり、同様に Shannon エントロピーと  $m$ -ダイバージェンス損失がペアとなる等式がキーとなる。もう 1 つは、Bayes 予測問題に内在する双対性であることを明らかにすることである。ルジャンドル変換に着目すると、Bayesian model averaging の枠組みにおける予測問題 (1) および (4) はそれぞれ等価な制約つき最小問題に書き換えられる。

## 2. Saddlepoint equality

指数型分布族では MLE に対して次の等式が成り立つことが知られている (Kullback, 1959)。

$$\log \frac{p(x; \hat{\theta}_M)}{p(x; \theta)} - D(p(y; \hat{\theta}), p(y; \theta)) = 0 \quad \text{for any } x \text{ and } \theta. \quad (6)$$

望大項と望小項がバランスするという意味で (6) を saddlepoint equality と呼ぶことにする。

Yanagimoto & Ohnishi (2009) は Bayes モデル  $p(x; \theta)\pi(\theta)$  において  $e$ -mixture 予測分布  $p^{(e)}(y|x)$  が満たす等式

$$\mathbb{E} \left[ \log \frac{p^{(e)}(x|x)}{p(x; \theta)} - D(p^{(e)}(y|x), p(y; \theta)) \mid \pi(\theta|x) \right] = 0 \quad \text{for any } x \quad (7)$$

を導いた。期待値記号内の第 1 項は対数尤度比の拡張と考えられるので、この等式も saddlepoint equality と呼ぶ。

一方、 $m$ -mixture 予測分布も saddlepoint equality と呼ぶべき等式を満たすことを示せる。確率密度  $p(y)$  のエントロピーを  $H[p(y)] = \mathbb{E}[-\log p(y) \mid p(y)]$  と書くと、

$$\mathbb{E} \left[ H[p^{(m)}(y|x)] - H[p(y; \theta)] - D(p(y; \theta), p^{(m)}(y|x)) \mid \pi(\theta|x) \right] = 0 \quad \text{for any } x \quad (8)$$

が成り立つ。対数尤度比の代わりに Shannon エントロピーの差が現れていることに注意されたい。2 つの等式 (7) および (8) は、 $e$ -ダイバージェンスと対数尤度のペア、 $m$ -ダイバージェンスと Shannon エントロピーのペアの間に存在する類似性を示している。

本発表では、以上の内容を第 1 節で述べた Bayesian model averaging の問題設定の下で議論する。特に、最適予測分布が満たす (7) および (8) のタイプの saddlepoint equality を導く。また、対数尤度最大化によって得られる予測分布が (6) のタイプの saddlepoint equality を満たし、その  $m$ -ダイバージェンス版の等式を満たす予測分布が Shannon エントロピー最大化によって得られることを示す。これらは、2 つのダイバージェンス損失の下での議論で見られる著しい類似性である。

最適予測分布である  $e$ -mixture 予測分布 (3) および  $m$ -mixture 予測分布 (5) に対して、その事後リスクの上限を統一的な方法で求めることができる。具体的には、それぞれの損失の下である凸関数を定義し、それを最小化することによって求められる。

## 3. 予測問題に内在する双対性

予測問題は最小問題であるが、それと等価な制約つき最小問題に書き換えることができる。ここで等価性とは最適解が同一という意味である。第 2 節の最後で述べた凸関数およびそれによって引き起こされるルジャンドル変換がキーとなる。予測問題に内在する双対性に着目することにより、 $e$ -ダイバージェンス損失・対数尤度の世界、 $m$ -ダイバージェンス・Shannon エントロピーの世界のそれぞれを簡明に理解することができる。

# Bayesian モデル $e$ -統合：複数の証拠と意見

中央大学 柳本 武美  
九州大学 大西 俊郎

## 0. 要約

複数のベイズモデルから、統合した母数の事後密度の構成法についての一つの提案を行う。提案はベイズモデルの  $e$ -混合のアイデアに基づいている。この研究が、共通の標本密度を証拠と見なし事前分布を専門家の意見と見なすことによって、現在成長しつつある重要な分野への適用の展開が望めることを指摘する。

## 1. 既存の統合法

複数のベイズモデルがあって、これらを統合した母数の事後密度を構成することは一つの典型的な問題である。 $K$  通りのベイズモデル  $p_k(x|\theta)\pi_k(\theta)$ ,  $k = 1, \dots, K$  を考える。更にモデル  $k$  が正しいと見なされる確率を  $w_k > 0$ ,  $\sum w_k = 1$  とする。この問題で通常に用いられる統合法は、ベイズモデル平均法として、

$$\pi_C(\theta|x) = \frac{\sum w_k m_k(x) \pi_k(\theta|x)}{\sum w_k m_k(x)}. \quad (1.1)$$

但し  $m_k(x)$  は周辺密度  $\int p_k(x|\theta)\pi_k(\theta)d\theta (= E\{p_k(x|\theta)|\pi_k(\theta)\})$  である。この方法は、ベイズモデルをごく素朴に理解すればこのモデル平均法が事後密度として得られる。

## 2. 提案する方法

しかし、「モデル  $k$  が正しいと見なされる確率」の解釈は必ずしも自明ではない。また、Genest and Zidek (1986) がレビューしているように、複数の事前分布の統合では、通常の  $m$ -混合ではなくて  $e$ -統合の方が理論的な整合性がとれるとの指摘もある。既存の方法 (1.1) に対して提案する統合法は

$$\pi_e(\theta|x) \propto \exp\{\sum w_k \log \pi_k(\theta|x)\}. \quad (2.1)$$

である。この統合法は通常に用いられる  $m$ -混合ではなくて、双対な関係にある  $e$ -混合の良さ (Yanagimoto and Ohnishi, 2009) に動機づけられている。

提案する方法を評価するためには、次の統合法

$$\pi_m(\theta|x) = \sum w_k \pi_k(\theta|x). \quad (2.2)$$

が興味深い。Cox (1972) の部分尤度法、共通母数の推定 (例えば Yanagimoto, 1990) は各層の主要な情報を抽出して、抽出した情報 (この場合は各層の事後密度) を統合する。

提案法は以下に示すように、別の二つの定義が可能である。

**Proposition 2.1.** 1) Let  $\tilde{\pi}(\theta|x)$  be  $c(x) \exp \sum w_k \log \{p_k(x|\theta)\pi_k(\theta)\}$  with  $c(x)$  being the normalizing constant. Then  $\tilde{\pi}(\theta|x)$  is the same as  $\pi_e(\theta|x)$  in (2.1).

2) Suppose that the sampling densities are common through  $k$ . We consider the weighted the  $e$ -mixture prior densities  $\pi_e(\theta) \propto \exp \sum w_k \log \pi_k(\theta)$ . Then the posterior density is equivalent with  $\pi_e(\theta|x)$ .

### 3. 性質

統合した事後密度は、ある統合した事前密度から誘導される方が望ましい。このために、標本密度  $p(x|\theta)$  が  $k$  に依存しないで共通である場合を扱う。

**Proposition 3.1.** The two aggregated prior densities,  $\pi_e(\theta|x)$  and  $\pi_C(\theta|x)$ , are induced from the following two aggregated prior densities, respectively.

$$\pi_e(\theta) \propto \exp\{\sum w_k \log \pi_k(\theta)\}. \quad (3.1)$$

$$\pi_C(\theta) = \sum w_k \pi_k(\theta) \quad (3.2)$$

提案法は、事前密度を  $e$ -混合して統合事前密度を仮定した場合に対応する。この事前密度の統合法は、Genest and Zidek (1986) でも興味深い方法と評価されている。次に、単調性を調べる。

**Proposition 3.2.** Suppose that  $\pi_k^*(\theta)$  has the increasing likelihood ratio in  $\theta$  than  $\pi_k(\theta)$  for every  $k$ . Then it follows that 1)  $\pi_e^*(\theta)$  has the increasing likelihood ratio in  $\theta$  than  $\pi_e(\theta)$ , 2)  $\pi_e^*(\theta|x)$  has the increasing likelihood ratio in  $\theta$  than  $\pi_e(\theta|x)$ , and 3)  $\pi_m^*(\theta|x)$  is stochastically larger than  $\pi_m(\theta|x)$ .

一方、既存の統合化法は単調性を持たない。つまり、標本密度と異なる事前密度には小さなウェイトをおくからである。この事実は、情報量の非常に小さい事前分布を無視することを示している。情報量の小さい事前密度の排除は、Bayes 推論にとって大きな欠陥である (Yanagimoto and Ohnishi, 2011)。

### 4. 複数の証拠と意見

始めに述べたように、この問題は複数の証拠と意見の統合の問題として定式化できる。ベイズモデルでは標本密度  $p(x|\theta)$  で標本  $x$  は固定して議論される。標本  $x$  は固定すると尤度関数となり、これは証拠に基づいた効用関数  $U(\theta)$  と見なことができる。標本との直接的な関係は無視することにより統合的な証拠に基づいた効用関数を利用することも可能になる。証拠に基づいて第三者 (例えば事務局) が提示した、母数  $\theta$  に対する効用の評価とする。事前分布はパネラーあるいは専門家の意見と見なす。歴史的に見ても事前分布は当事者の主観的な意見の記述として現れている。もし効用関数があれば専門家の意見の統合であり、事前分布の統合問題になる。

提案した統合事後分布の特徴として二つが挙げられる。1) 専門家に全く意見がなければ、improper な無情報事前分布として記述される。既存法では全く無視されてしまう。提案法では全体の意見を位置は変えないで結論の強さを変える。2) ある専門家が確率的に大きい意見を提示すれば、提案法では大きい結論に誘導する (単調性)。一方既存法では効用関数との乖離が大きい場合には当該専門家の意見を無視する。

二つの特徴は、多くの場合に意見の統合法として望ましい性質である。

### References

- 1) Aitchison, J. (1975). *Biometrika*, **62**: 547-554.
- 2) Cox, D.R. (1975). *Biometrika*, **62**, 269-276.
- 3) Genest, C. and Zidek, J.Z. (1986). *Statistical Science Page*, **1**, 114-135
- 4) Yanagimoto, T. (1990). *J. Statist. Plan. Inf.*, **25**, 187-198.
- 5) Yanagimoto, T. and Ohnishi, T. (2009). *J. Statist. Plann. Inf.*, **139**, 3064-3075.
- 6) Yanagimoto, T. and Ohnishi, T. (2011). *J. Statist. Plann. Inf.*, **41**, 1990-2000.

# 有意でない試験を継続するベイズ的視点

中央大学大学院 理工学研究科 小林 千鶴  
中央大学 理工学部 鎌倉 稔成・柳本 武美

## 1. はじめに

統計的検定は、拙速に最適解を求める手法とは異なり、誤った仮説の否定を介して支持する仮説を正当化するための検証的推論形式として広く定着している。一方で、その推論の枠組みに様々な改善すべき点を抱えていることも事実である。その一つとして、Cornfield (1966) が提起した、統計的検定の問題と Bayes 法の適用に関する基本的な問題がある。そこで、この問題について、予測信用区間 (Yanagimoto and Ohnishi, 2009a) の視点からの提案を試みる。

## 2. 扱う問題と動機

Cornfield (1966) は、ある研究者が一度試験を実施して有意でない結果を得た場合を例示した。その結果に不満な研究者が更に研究を進めようとしても、有意水準のインフレーションを招くため、それ以降の試験が実施できないという奇妙な推論の枠組みになっていることを指摘した。つまり、検定を複数回繰り返すときは、指定した水準を遵守する配慮が必要であり、DeMets ら (1994) の  $\alpha$  消費関数法はそのために発展されてきた。一方で、有意でないという結果を得た場合の結論は、対立仮説が棄却されたということで、帰無仮説が肯定されたわけではない。

Yanagimoto and Ohnishi (2009a) は、Lindley paradox を批判的に検証する過程で予測信用区間を提案した。この視点が上記の問題に有用であることを示唆している。そこで、この問題に対しての性能を検証すると共に、可能な改善法を探る。予測信用区間の概説は 5 節で行う。

## 3. 既存の方法

前の試験を解析して対立仮説が棄却されたときの対応策としては次の 3 方法がある。

- 1) 追加試験は行わない。(水準遵守)
- 2) 二つのデータを統合して解析する。
- 3) 追加試験の方策を探る。

## 4. 提案法

この問題に対する一つの提案は、既存のデータから事前分布を構成して新しいデータを加えて予測分布を導きその上で信用区間を求めることである (Yanagimoto and Ohnishi, 2009a)。この方式は Bayes 法の論理の継続性に着目した方式である。この発表の目的は、直感的には妥当と期待される方式の実際的な性能を探ることにある。

### 4.1. 1 標本正規分布の場合

信用区間は、

$$I_\alpha = \{\mu \mid \sqrt{n}(\mu - \bar{z}) > -z_\alpha\}$$

この区間に  $\mu = \mu_0$  が含まれないことは、

$$\sqrt{n}(\bar{z} - \mu_0) > z_\alpha$$

と書ける。これは、

$$R(y) = \left\{ x \mid \sqrt{n}(\bar{x} - \mu_0) > \frac{1}{n} \{(n+m)z_\alpha - \sqrt{nm}(\bar{y} - \mu_0)\} \right\}$$

とも書ける。

ここで述べた方法は、通常のベイズ推測方式（例えば Bolstad, 11 節, 2007）において、事後密度を予測子型事後密度と置き換えたの推測方式に一致する。その性能を調べているが、検出力が対立仮説に応じて増加していることが確かめられた。

#### 4.2. 1 標本 2 項分布の場合

同様に計算し、 $\hat{\eta} = E\{\log(p/(1-p)) | \text{Be}(t+1/2, N-t+1/2)\}$ ,  $\tilde{p} = \exp \hat{\eta} / (1 + \exp \hat{\eta})$  とおくと、信用区間は以下ようになる。

$$I_{1-\alpha} = \left\{ p \left| 2 \left[ \tilde{p} \log \frac{\tilde{p}}{p} - (1-\tilde{p}) \log \frac{1-\tilde{p}}{1-p} \right] < c_\alpha \right. \right\}$$

ここで  $c_\alpha$  は信用区間  $I_\alpha$  の予測子型事後密度の下での確率が指定の水準  $1-\alpha$  と一致するように選ぶ。この場合予測子型事後密度はベータ分布  $\text{Be}(n\tilde{p}+0.5, n(1-\tilde{p})+0.5)$  に従う。対立仮説が片側の場合は適当な  $p_l(\alpha)$  を求めて  $I_{1-\alpha} = (p_l(\alpha), 1)$  と表される。Beta 分布は広く用いられていてソフトウェアも整備されていて本法の適用により格別の計算困難を生じることはない。

この区間に  $p = p_0$  が含まれないときに帰無仮説が「棄却」される。 $p = p_0$  が含まれないことは、区間  $(p_0, 1)$  の予測子型事後確率が  $1-\alpha$  より大きいとも言い換えられる。

正規分布が仮定できる場合と異なり、2 項分布を仮定した場合には性能を評価するために様々な工夫が求められる。その一つが、最適予測子が自然母数  $\eta$  の事後平均を用いて表現されることがある。通常に用いられる推定量  $\hat{p}_M = t/N$  が現れないことである。この問題は Jeffreys' prior を仮定が功を奏する。この事実は、Jeffreys' prior の性質から想像され、Yanagimoto and Ohnishi (2009b) でも解析的に評価されているが、近似は大変良いことが確かめられた。

これを拡張し、2 標本 2 項分布の場合についての解析も試みた。

以下に結果を示すが、これらは、 $(m_0, y_0) = (50, 20)$ ,  $(m_1, y_1) = (50, 26)$ ,  $n_0 = n_1 = 50$ ,  $p_1 = p_l(0.05)p_u$ ,  $p_l = 0.4$ ,  $p_u = 0.7$  とした場合である。(図 1)(図 2)

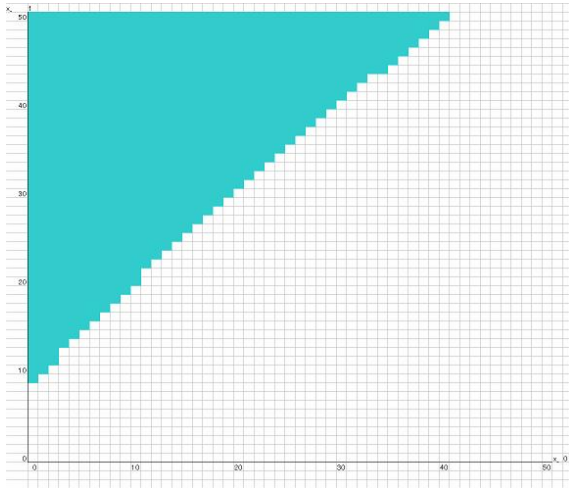


図 1: 提案法の棄却域

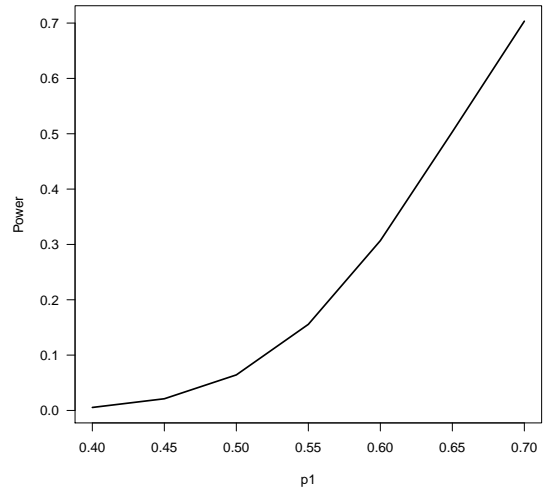


図 2: 提案した検定法の検出力

# Estimation of the Number of Failures Using Ordinary Differential Equations to Reliability Models, with Applications to SARS, A(H1N1), and FMD Cases

Hideo Hirose

*Department of Systems Design and Informatics,  
Kyushu Institute of Technology,  
Fukuoka 820-8502, Japan*

---

## Abstract

In estimating the number of failures using grouped truncated data, we often encounter cases that the estimate is smaller than the true one when we use the likelihood principle to conditional probability. In infectious disease spread predictions, the SIR model described by simultaneous ordinary differential equations is commonly used, and it can predict reasonably well the number of infected patients even when the size of observed data is small. It was attempted here if the ordinary differential equation model can estimate the number of failures more accurately than does the likelihood principle under the condition of grouped truncated data. The positive results are obtained in the Weibull model, similarly to the cases of the SARS, A(H1N1), and FMD.

**Keywords:** Weibull model; type II right truncated grouped data; number of failures; SIR model; differential equation, best-backward solution;

---

## 1. Introduction

We consider a problem that failures occur day by day, first increasingly but later decreasingly, and failures per day will finally cease. We want to know how many failures will occur and when they will cease in the midst of failure occurring. For example, when claims of some products are increasingly reported, we have to make a decision whether we accept recall or not. In other cases, we may want to know how many vaccines we have to provide when serious infectious diseases are propagating.

Many approaches are dealt with to each type of problems. In reliability analysis, we often assume a Weibull model for failures and obtain its parameters using the likelihood principle. In the case of product failures, the truncated models [2], the limited failure model [3], and the truncated model [1] are used based on statistical modeling. In infectious disease analysis, the SIR model described by simultaneous ordinary differential equations are often used, and recently the agent-based models are also used for more precise computation [4].

Here, we deal with the estimation problem for the number of failures in the Weibull model via a different approach from the conventional likelihood principle approach; that is, when failures are observed day by day, we estimate the final number of failures using the ordinary differential equation.

## 2. Differential Equation Approach in Weibull Models

As indicated by the cases of pandemics, the estimate for the total number of failures (deaths) may become smaller than the true one when we use the truncated model and the number of observed data is not large. On the contrary, the differential equation models may provide reasonable (at least for the danger side) estimate. Therefore, we apply the differential equation model to the incomplete data cases. For the Weibull model, the differential equation can be expressed as,

$$W'(t) = \left(\frac{b}{a}\right) \exp\left(\frac{t}{a}\right)^{b-1} \{N - W(t)\}, \quad (t > 0), \quad (1)$$

where,  $W(t)$  denotes the cumulative number of failures by time  $t$ , and it can be written by

$$W(t; a, b) = N \left[1 - \exp\left\{-\left(\frac{t}{a}\right)^b\right\}\right], \quad (a, b > 0). \quad (2)$$

The unknown parameters are  $a$ ,  $b$ , and  $N$ , similar to the truncated model [1].

### 3. Simulation Studies

We use a Weibull case mimicking the observed data in real world [1]. The Weibull parameters are  $a = 57.2, b = 3.1, N = 175$ , and we provide data during  $0 \leq t \leq 100$ , as shown in Figure 1.

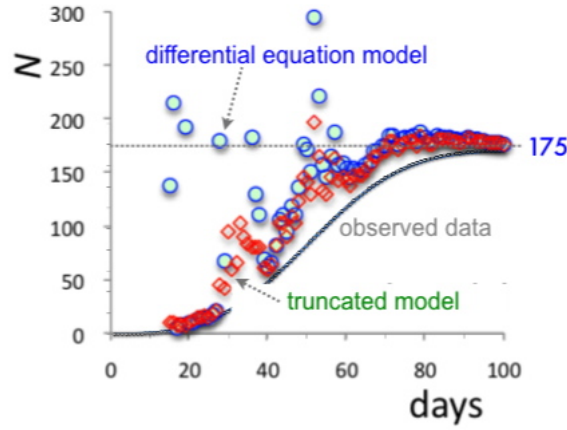


Figure 1: Comparison of a simulation case for the estimated total number of failures between the differential equation model and the truncated model.

Figure 1 shows the estimated  $\hat{N}$  for the proposed method and the truncated model method for a simulation case. From the figure, we can see that these two methods provide similar estimates in later stages, typically  $T > 80$ . Using the truncated model, when  $T = 50$ , which suggests that we observe data only before the inflection point of the cumulative distribution function, we have failed to obtain the estimates in many cases. Even if the estimates are obtained, the estimated values are smaller than the true one. On the contrary, the differential equation method can obtain the estimates even in earlier stages although the estimates vary much. This property that we can obtain the information about the final stages from the early stages is considered to be crucial when we deal with a very important prediction such as the pandemic analysis or the failure analysis for extremely valuable products. Therefore, we have found that the proposed method has an ability to predict the final stages with rather small samples, i.e., with the data in earlier stages.

### 4. Concluding Remarks

In estimating the number of failures using the truncated data for the Weibull model, we often observe that the estimate is smaller than the true one when we use the likelihood principle to conditional probability. To overcome this problem, we have proposed to use the differential equation method, because we have experienced that the SIR model described by simultaneous ordinary differential equations can predict the final stage condition, i.e., the total number of infected patients, even when the size of observed data is small, unlike the truncated model approach. We have investigated, in this paper, whether the number of failures in the Weibull model can be estimated accurately using the ordinary differential equation, and have found that the proposed method has the ability to predict the final stages even with a small number of samples in the early stage.

### References

- [1] H. Hirose, The truncated model and its applications to lifetime analysis: unified censored and truncated models, *IEEE Transactions on Reliability*, 54, pp.11-21 (2005).
- [2] J.P. Klein and M.L. Moeschberger, *Survival Analysis*, New York: Springer, 1997.
- [3] W.Q. Meeker, Limited failure population life tests: application to integrated circuit reliability, *Technometrics*, 29, pp.51-65 (1987).
- [4] Y. Toyosaka and H. Hirose, Pandemic simulations by MADE: the hybrid method of multi-agent and differential equations, *Asia Simulation Conference*, pp.1-5 (2009).

## Statistical Analysis with Dilatation for Development Process of Human Fetuses

島根大・総合理工	内藤貫太
総合研究大学院大学	野津昭文
滋賀医科大	宇田川潤
島根大・医	大谷 浩

はじめに：ヒト胎児の発生過程については未だ多くの未解明な問題が存在するものの，臓器間にある種の調和的關係が観察されることが知られている．このような知見は直感的，経験的なものであり，調和的発生を説明する理論は未だ乏しい．本研究においては，等角写像からの乖離を測る測度である歪曲度を用いて，調和的発生の概念の数理的理解を与えたい．

調和度:  $\mathbb{R}^m$  におけるサイズ  $N$  の対応のある 2 つの点群  $X^T = [x_1, \dots, x_N]$  と  $Z^T = [z_1, \dots, z_N]$  を考える．対応があるとは， $x_i$  と  $z_i$  は同一個体に関する異なる 2 つの  $m$  次元変数であることを意味する．具体例としては例えば  $m = 2$  として，ヒトの (右大腿長, 右上腕長)(=  $x$ ) と同一人の (左大腿長, 左上腕長)(=  $z$ ) が挙げられる．ここで議論する問題は，この 2 つの点群  $X^T$  と  $Z^T$  の類似性，“調和度”，をどのように考えるかということである．

歪曲度: この 2 つの点群を対応づける変換写像  $f$  の存在を仮定する：

$$z = f(x) .$$

我々のアイデアは，点群  $X^T$  と  $Z^T$  の調和度はこの変換写像  $f$  の性質を通して議論しようというものである．特に，擬等角写像論において等角写像からの乖離を測る尺度である歪曲度と呼ばれる指標をもって調和度とすることを提案する．適当な設定のもとで歪曲度は

$$d = d(f) = \max_x \kappa(x)$$



と定義される．ここで，

$$\kappa(x) = \frac{\|Df(x)\|^m}{J_f(x)},$$

であり，

$$Df(x) = \left( \frac{\partial f_i(x)}{\partial x_j} \right)_{1 \leq i, j \leq m}, \quad J_f(x) = \det(Df(x)),$$

$\|B\| = \sup\{|Bx| : x \in \mathbb{R}^m, |x| = 1\}$  は作用素ノルム， $|w|$  は  $w \in \mathbb{R}^m$  のユークリッドノルムを表す． $\kappa(x) \geq 1$  であり， $\kappa(x) \equiv 1$  は  $f$  が等角写像であることと同値である．我々の定式化は

$$\text{点群 } X^T \text{ と } Z^T \text{ は調和} \iff f:\text{等角} \iff \kappa(x) \equiv 1$$

となる．

**Linear Mapping:** 変換写像に線形性  $f(x) = Ax$  を仮定すると， $m \times m$  行列  $A$  の推定値  $\hat{A}$  から， $\kappa(x)$  は  $\hat{A}^T \hat{A}$  の固有値で表現される．そのようにして得られる  $\kappa(x)$  の有意性を議論する際には， $f$  を等角と仮定した上での  $A$  の推定が必要となり，そこに Procrustes Analysis が現れる．

等角写像の仮定のもとで， $\hat{d} = d(\hat{f}) = d(\hat{A})$  の極限分布を導出した．極限分布を用いた  $p$  値の計算アルゴリズムについて解説した．

**Radial Mapping:** 変換として，パラメータ  $a$  を含む Radial Mapping

$$f_a(x) = |x|^{a-1}x, \quad x \in \mathbb{R}^m,$$

を用いたときの歪曲度，およびその推定値の等角写像の仮定のもとでの漸近挙動に関する結果を導出した．

ヒト胎児形態計測データへの適用: 京都大学大学院医学研究科附属先天異常標本解析センターにて蒐集された，世界的にも大変貴重な数万体におよぶ日本人ヒト胎児標本のうち，島根大学医学部発生生物学教室に移管され，管理されている胎児標本の形態を計測したものがヒト胎児形態計測データである．データの内容としては，個々の胎児標本について，頭殿長 (CRL)，体重などの外表形態と共に，肺，副腎，肝臓，腎臓など，解剖によって得られた 100 以上の臓器部位の測定値が含まれている．

シンポジウムにおいて，このヒト胎児形態計測データに対し歪曲度を用いた調和度解析を適用した結果を報告した．

# 確率的比例効果則とフィードバック効果を組み合わせた生体リスク管理モデル

熊澤 蕃（元内閣府原子力安全委員会事務局技術参与）

Biological risk control model based on the law of randomly proportional effect with feedback  
Shigeru KUMAZAWA (former technical counselor, Secretariat of the NSC, Cabinet Office)

Abstract-- This paper presents a biologically adverse risk control model based on the law of randomly proportionate effect with feedback mechanisms of reducing the risk, as a basic form of risk control for covering from microscopic to macroscopic levels. The hybrid scale is inherent to attain the reasonable risk control.

Key Words: health risk control, law of proportionate effect, feedback, hybrid scale (unification of linear & log)

## 1. はじめに

生命現象には、「環境と体内に遍在するリスク源に遭いながらも、将来に向けた活力と自然に潜む恵みに与る機能」が必須である。生命系に流入の物質・エネルギー・情報に伴う益と害に係る自力応答には、不確実な状況で未来価値に向けた「リスク管理」の進化的合理性が形成されるとして、本発表では、この確率的リスク管理モデルを比例効果則(law of proportionate effect)とフィードバック生成則(law of creation)から導き、これがリスク感覚尺度を与えることを示す。

## 2. リスク管理の定式化

生命系とは、未来が現在の行動を決定する自己組織化系であり、この自己組織化系は、必要な物質・エネルギー・情報の流れを摂取しつつ、そこに内在する益と害に自力応答し、正味益を達成する存在である。このため、生命系は、「益と害」に係る知性(Intelligence)に基づくアクセル機能とブレーキ機能を働かせるリスク管理の合理的な仕組みを獲得する、と想定する。

第1のアクセル機能は、自然現象（物質拡散・相互作用等）を記述する比例効果則に従い、生命系の摂取（物質・エネルギー・情報）に伴う「益と害」に係るリスク増分を $\Delta R = \varepsilon R$ と置く。放射性物質壊変 $\Delta N = -\lambda N$ 又は $\Delta N/N = -\lambda$ 、刺激 $S$ と感覚 $I$ の $\Delta I = k \Delta S/S$ 、状態数 $W$ とエントロピー $S$ の $\Delta S = k \Delta W/W$ 、生起確率 $P$ と情報量 $I$ の $\Delta I = -\Delta P/P$ など、その例は数多い。

第2のブレーキ機能は、サイバネティックス又は自然のフィードバック生成則に従い、リスク増加係数を $\varepsilon' = \varepsilon - \tau \Delta R$ と低減、 $\Delta R = \varepsilon' R = (\varepsilon - \tau \Delta R) R$ に代入し導かれる $\Delta R = \varepsilon R / (1 + \tau R)$ と置く。生命系の知性が思い描く将来価値を実現する上で、無生物に近い自然生命現象を含め、フィードバックは必須の機能である。

生命系のリスク管理は、管理不要のリスク事象を圧倒的にし、管理有効のリスク事象を少数で効果的とし、より大きい管理飽和のリスク事象は極少とすることが、知性意欲確保上、好ましい。

上記のアクセル機能とブレーキ機能で構成されるリスク管理は、リスク $\tau R \ll 1$ で、 $\Delta R \approx \varepsilon R$ ゆえ「管理不要領域」、 $0.1 < \tau R < 5$ で、 $\Delta R = \varepsilon R / (1 + \tau R)$ ゆえ「管理有効領域」、 $\tau R \gg 1$ で、 $\Delta R \approx \varepsilon / \tau$ ゆえ「管理飽和領域」を有するため、各領域でのリスク事象頻度を、それぞれ、大半、少数、極少にできる。それゆえ、比例効果則とフィードバックを組合せたリスク低減管理モデルは、次のように定式化される。

$$\Delta R = \varepsilon R / (1 + \tau R) \quad (1)$$

## 3. リスク低減管理モデル式の性質

式(1)を $\varepsilon$ につき解くと、 $\varepsilon$ はリスク $\tau R$ の相加的増分と相乗的増分の和に等しいことが導かれる。

$$\varepsilon = \Delta \tau R + \Delta R / R = \Delta (\tau R + \ln [\tau R]) \quad (2)$$

混成関数(hybrid function)を $\text{hyb}(x) = x + \ln(x)$ 、その逆関数をサイブ関数 $\text{cyb}(x)$ と表記すると、リスク増分の区間 $[0, T]$ における多段階の確率増加過程における全リスク $R_T$ は次式で与えられる。

$$\varepsilon = \Delta \text{hyb}(\tau R) \text{ or } \tau R_T = \text{cyb} \left( \int_0^T \varepsilon dt \right) \quad (3)$$

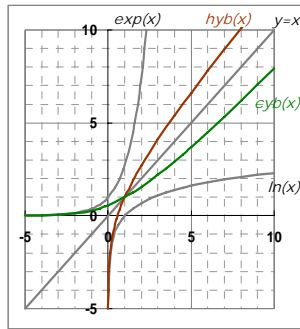


図 1 諸関数(相加的,相乗的,混成的)

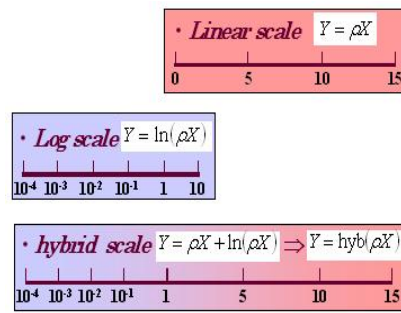


図 2 混成目盛(hybrid scale)

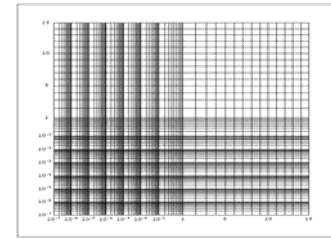


図 3 両混成用紙(hyb-hyb paper)

式(3)から、式(1)のリスク低減管理モデルにおいて、 $\varepsilon$ はリスクの混成的増分 $\Delta \text{hyb}(\tau R)$ に等しく、また、全リスク $R_T$ は $\text{cyb}(A)/\tau$ に等しいことが分かる。ここで、 $A$ は $\varepsilon[t]$ の区間 $[0, T]$ の積分とする。

図 1 から、混成関数  $\text{hyb}(x)$ は、 $x \ll 1$  の対数関数から  $x \gg 1$  の線形関数に変わる連続関数、また、サイブ関数  $\text{cyb}(x)$ は、 $x < 2.2$  の指数関数から  $x > 6.6$  の線形関数に変わる連続関数、という特性を有する。混成関数  $\text{hyb}(x)$ を目盛で表すと、図 2 に示すように、 $x < 0.1$  で対数目盛に近い一方、 $x > 5$  で線形目盛に近く、かつ、その中間を連続接続する「混成目盛」を与える。

式(3)の $\Delta \text{hyb}(\tau R)$ は、 $\tau R < 0.1$  で $\Delta \ln(\tau R)$ の相乗的增加、 $\tau R > 5$  で $\Delta \tau R$ の相加的増加、その中間で混成的増加となる特性がある。また、全リスク $R_T$ は、 $A < 2.2$  で $\exp(A)/\tau$ の指数増加、 $A > 6.6$  で $A/\tau$ に並行な線形増加、その中間でサイブ的增加となる特性がある。

式(3)で、確率変数 $\varepsilon$ が区間 $[0, T]$ で仮想的に $n$ 等分した各段階でマルティンゲール中心極限定理を満たす場合、 $n \rightarrow \infty$ において、 $A$ は正規分布に漸近することが示される<sup>(1)</sup>。それゆえ、 $\text{hyb}(\tau R_T)$ も正規分布に漸近、従って、 $R_T$ は混成対数正規分布に漸近する。これから、式(1)のリスク低減管理モデルに基づくリスクの混成的増分過程(時間過程又は多経路空間過程)によるリスク累積 $R_T$ も混成対数正規分布を示すことが導かれる。これは、物理学的にエントロピー最大又は自由エネルギー最小の法則性を満たす確率過程である。

アクセルとブレーキから成る基本構造は、相乗的作用と相加的作用を連続に接続する HS 特性を示すが、この特性は、図 1 に示す混成目盛の特性そのものである。混成目盛を用いた 2 次元混成空間 (hybrid-hybrid paper) を導入すると、さらに体系的にリスク低減管理状態を記述できる。

#### 4. 適用例

本発表では、分子生物学又は DNA レベルでの影響機序に立ち入らず、マウス細胞レベルでの HS 特性を検討した。第 1 は、電離性放射線分割照射における細胞生残曲線に見られる損傷修復効果の差異の HS モデル(HS 特性)による解析結果、修復能低下は同モデルの損傷修復フィードバック係数の低下となること、第 2 は、精源細胞の染色体相互転座における線量-反応関係の HS モデルによる解析結果、生残細胞当たりの線量-反応関数と細胞生残率関数の積に分離した両関数がリスク管理モデルを反映した HS 特性を示し、その積は元データを再現すること、を示した。

このほか、ヒトの思考に基づく文章の長さ分布(words/sentence)、静止画像並びに動画像注視中の眼球速度分布(角度/秒)、サッカー試合中の前半/後半中特定選手の走行速度分布(m/s)、胎児を含むヒトの成長分布 (m/y) などにも、長い文章の頻度低減化、高速眼球運動の頻度低減化、高速移動の頻度低減化、過大な身長の高頻度低減化などに HS 特性が見られる。なお、生命系以外にも、自然現象、経済現象などについても、広く HS 特性が確認にされる。

#### 5. まとめ

本稿では、生命系が必要とする物質・エネルギー・情報の流れに伴う「益と害」に対して正味益達成のリスク低減管理モデルの成立する基盤及び拠り所とする法則性を示した後、生命系の知性に基づいて達成可能な合理的なリスク管理尺度は、混成尺度であることを示した。混成目盛は、2 次元の両軸に採用することによって、様々な数量変動の特性変換を統一的に行う解析法とするほか、この性質が、生命系のリスク低減管理の解析に広く適用できる可能性のあることを示した。

#### 参考文献

- (1) 熊澤、大橋 “混成対数正規分布とその応用” 応用統計学会 Vol.15, No.1, 1-14(1986).