科研費シンポジウム

「多様な分野における統計科学の教育・理論・応用の新展開」報告書

日時: 2014年10月24日(金)~10月26日(日) 場所: 新潟大学駅南キャンパスときめいと 講義室 A

開催責任者:蛭川潤一(新潟大学),和泉志津恵(大分大学)

本シンポジウムは科学研究費・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」

研究代表者:谷口正信(早稲田大学基幹理工学部)

の助成により開催された

Program

Friday, 24, October

Opening 13:10-13:15 Junichi HIRUKAWA (Niigata University)

Afternoon Session I 13:15-15:15

Chair Koji Yonemoto (Kurume University)

1. 13:15-13:55 阿部 俊弘 (Toshihiro Abe)

東京理科大学 (Tokyo university of Science)

「長さと角度を変量に持つ分布族」(A family of cylindrical distributions)

2. 13:55-14:35 濱岡 豊 (Yutaka Hamaoka)

慶應義塾大学商学部 (Faculty of Business and commerce, Keio University)

「放射線被曝関連データの再分析」(Re-analysis of Radiation and Health Effect Data)

3. 14:35-15:15 新村 秀一 (SHUICHI SHINMURA)

成蹊大学経済学部 (Faculty of Economics, Seikei University)

「新しい判別分析の紹介 (New World of Discrminant Analysis

-Comparison of Linear Discriminant Function by K-fold Cross Validation-)

Afternoon Session II 15:30-17:30

Chair Yoshihiko Maesono (Kyushu University)

4. 15:30-16:10 吉田 知行 (Tomoyuki YOSHIDA)

北海道大学 (Hokkaido University)

「鉛同位体法の統計的研究」(Statistical Studies on Lead Isotope Method)

5. 16:10-16:50 神戸 翼 (Tsubasa Kambe)

慶應義塾大学大学院 健康マネジメント研究科 (Graduate school of health management, Keio University)

「中規模自治体病院における経営に関する要因分析 ~ 健全な経営の施策は? 総務省地方行財政 データから見えるモノ~」

(Factor analysis about the management in medium-sized and local government Hospital. \sim Important indices which are visible from the local administration and finance data of the Ministry of internal Affairs and Communications \sim)

6. 16:50-17:30 刈屋 武昭 (Takeaki KARIYA)

明治大学グローバル・ビジネス研究科 (Graduate School of Global Business, Meiji University)

An efficiency of a GLSE in regression with inversely linear covariance structure

Saturday, 25, October

Morning Session 9:20-12:00

Chair Toshihiro Abe (Tokyo university of Science)

7. 9:20-10:00 小泉 健太 (Kenta Koizumi)

慶應義塾大学大学院理工学研究科 (Graduate School of Science and Technology, Keio University)

Statistical Estimation for Optimal Dividend Barrier

8. 10:00-10:40 劉 言 (Yan LIU)

早稲田大学基幹理工学研究科 (Graduate School of Fundamental Science and Engineering, Waseda University)

Minimum contrast estimation for spectral densities based on exotic disparity

9. 10:40-11:20 汪 金芳 (Jinfang Wang), 萩原 学 (Manabu Hagiwara), 山本 光晴 (Mitsuharu Yamamoto)

千葉大学 大学院理学研究科 (Graduate School of Science, Chiba University)

「Coq/SSReflect を用いた条件付き独立性の形式的証明について」

(Formalization of statistical conditional independence relations using Coq/SSReflect)

10. 11:20-12:00 島谷 健一郎 (Kenichiro SHIMATANI)

統計数理研究所 (Institute of Statistical Mathematics)

Circular statistics for animal/plant behaviors and movement ecology for circular statistics

Afternoon Session I 13:10-15:10

Chair Kengo Kamatani (Osaka University)

11. 13:10-13:50 和泉 志津恵 (Shizue Izumi)

大分大学工学部 (Faculty of Engineering, Oita University)

「2段階サンプリングに基づく研究デザインと解析方法」

(Study design and methods based on the two stage sampling)

12. 13:50-14:30 Yasutaka Shimizu

(Waseda University)

LSE-type estimation for stochastic processes with small Lévy noise

13. 14:30-15:10 前園 宜彦 (Yoshihiko Maesono)

九州大学大学院数理学研究院 (Faculty of Mathematics, Kyushu University)

「ハザード関数のカーネル型推定量の漸近表現とその応用」

(Asymptotic representation of a kernel type hazard function estimator and its application)

Afternoon Session II (Guest Speakers Session) 15:25-17:40

Chair Junichi HIRUKAWA (Niigata University)

14. 15:25-16:25 Ngai Hang CHAN

(Department of Statistics, Chinese University of Hong Kong)

Adaptive Quantile Regression and Risk Bounds

15. 16:40-17:40 Cathy W. S. Chen

(Department of Statistics, Feng Chia University, Taiwan)

Bayesian Inference for Time Series of Counts Models

Sunday, 26, October

Morning Session 9:20-12:00

Chair Yasutaka Shimizu (Waseda University)

16. 9:20-10:00 鎌谷 研吾 (Kengo Kamatani)

大阪大学大学院基礎工学研究科 (Graduate School of Engineering Science, Osaka University)

Markov chain Monte Carlo method for high-dimensional heavy-tailed distribution

17. 10:00-10:40 竹内 文乃 (Ayano Takeuchi)

慶應義塾大学医学部衛生学公衆衛生学

(Department of Preventive Medicine and Public Health, School of Medicine, Keio University)

「ケースクロスオーバーデザインでアウトカムが再発イベントの場合の解析」

(Analysis of recurrent event data under the case-crossocer design)

18. 10:40-11:20 湯澤 あや (AYA YUZAWA)

東芝メディカルシステムズ株式会社 (Toshiba Medical Systems Corporation)

「全国ウツタイン統計を用いた院外心停止例の類型化」

(Classifying in patterns of out of hospital cardiac arrest, Using the Latent Class Analysis)

19. 11:20-12:00 Yusuke Saigusa, Kouji Tahata, Sadao Tomizawa

(Graduate School of Science and Technology, Tokyo University of Science;

Faculty of Science and Technology, Tokyo University of Science;

Faculty of Science and Technology, Tokyo University of Science)

Decompositions of symmetry using models based on f-divergence for square contingency tables

Afternoon Session 13:10-14:30

Chair **Hiroshi Shiraishi** (Keio University)

20. 13:10-13:50 張 元宗 (Yuan-Tsung Chang), 篠崎 信雄 (Nobuo Shinozaki)

目白大学 社会学部 (Faculty of Studies on Contemporary Society, Mejiro University)

慶應義塾大学 (Keio University)

「Pitman nearness 基準の下で平均と分散の双方に順序制約がある場合の 2 つの正規母集団の平均の推定とその応用」

(Estimation of Two Ordered Normal Means with Ordered Variances under Pitman nearness Criterion)

21. 13:50-14:30 明石 郁哉 (Fumiya AKASHI)

(Graduate School of Fundamental Science and Engineering, Waseda University)

Frequency domain GMM estimators for stable processes and its asymptotic optimality

Closing 14:30-14:35 Junichi HIRUKAWA (Niigata University)

A linear-circular model as an extension of exponential and von Mises distributions and its properties

Toshihiro Abe^{1,*}

¹ Tokyo University of Science, Tokyo, Japan

In natural science, we meet various observations which include length and angle. Examples include the strengths and directions of wind, lengths and directions of tree logs. Some literatures related to the models on cylinder are Johnson and Wehrly (1978), Mardia and Sutton (1978) and Kato and Shimizu (2008). Some of cylindrical data indicate that the concentration of direction tends to be high as length becomes longer. Kato and Shimizu (2008) compared their model with the one of Mardia and Sutton (1978), and showed the flexibility of their model. While their model is flexible, the normalization constant is complicated.

For the flexibility of the distribution and its good fit to real data, we propose the distribution whose conditional distributions are the Gamma and sine-skewed von Mises (Abe and Pewsey, 2011) distributions. Its pdf is given by

$$f(x,\theta) = C_{p,\gamma,\kappa} \{ 1 + \lambda \sin(\theta - \mu) \} x^{p-1} e^{-\gamma x \{ 1 - \tanh(\kappa/p)\cos(\theta - \mu) \}}, \tag{1}$$

where $(x,\theta) \in [0,\infty) \times [0,2\pi)$, p>0 is a shape parameter, $\gamma>0$ is a scale parameter, $\kappa\geq 0$ controls the concentration of the direction and $-1\leq \lambda\leq 1$ is a skewing parameter. We denote the distribution (1) by Gam-SSVM $(p,\gamma,\mu,\kappa,\lambda)$. Clearly, the distribution (1) reduces to the Johnson-Wehrly's distribution when $p=1, \gamma=b, \kappa=\tanh^{-1}(a/b)$ and $\lambda=0$, however, our parametrization has no restriction by other parameters.

One of the advantages of the distribution (1) is that its normalizing constant is given by known special functions. The normalizing constant of the pdf (1) is explicitly given by

$$C_{p,\gamma,\kappa} = \frac{\gamma^p}{2\pi\Gamma(p)\cosh^p(\kappa/p)P_{p-1}(\cosh(\kappa/p))}$$
$$= \frac{\gamma^p}{2\pi\Gamma(p)_2F_1(p/2,(p+1)/2;1;\tanh^2(\kappa/p))},$$

where $\Gamma(\cdot)$ is the Gamma function and $P_{p-1}(\cdot)$ is the associated Legendre function of the first kind of degree p-1 and order 0. Here ${}_2F_1(a,b;c;z)$ is the Gauss hypergeometric function defined as

$$_{2}F_{1}(a,b;c;z) = \frac{1}{B(b,c-b)} \int_{0}^{1} \frac{t^{b-1}(1-t)^{c-b-1}}{(1-tz)^{a}} dt$$

or, alternatively, as

$$_{2}F_{1}(a,b;c;z) = \sum_{r=0}^{\infty} \frac{(a)_{r}(b)_{r}z^{r}}{(c)_{r}r!}, \quad |z| < 1,$$

where

$$(a)_r = \begin{cases} a(a+1)\cdots(a+r-1), & r \ge 1, \\ 1, & r = 0, \end{cases}$$

is referred to as Pochhammer's symbol.

The distribution reduces to the Gamma and von Mises distribution denoted by Gam-VM (p, γ, μ, κ) when $\lambda = 0$, Gamma and circular uniform distribution denoted by Gam-CU (p, γ) when $\kappa = 0$. Moreover, the Gam-VM distribution ($\lambda = 0$ in (1)) maximizes the entropy over all angular-linear distributions subject to E[X], $E[\log X]$, $E[X\cos(\Theta)]$ and $E[X\sin(\Theta)]$ taking specified values consistent with expectation.

The marginal density of pdf (1) with regard to angle is the sine-skewed Jones–Pewsey (SSJP) distribution (Abe and Pewsey, 2011) with $1/\psi = -p$:

$$f_{\Theta}(\theta) = C_{p,\gamma,\kappa} \int_{0}^{\infty} (1 + \lambda \sin \theta) x^{p-1} e^{-\gamma x \{1 - \tanh(\kappa/p) \cos \theta\}} dx$$
$$= \frac{(1 + \lambda \sin \theta) \{1 - \tanh(\kappa/p) \cos \theta\}^{-p}}{2\pi \cosh^{p}(\kappa/p) P_{p-1}(\cosh(\kappa/p))}.$$

The marginal density of the pdf (1) with regard to length is an extended version of the one in Johnson and Wehrly (1978):

$$f_X(x) = C_{p,\gamma,\kappa} \int_0^{2\pi} (1 + \lambda \sin \theta) x^{p-1} e^{-\gamma x \{1 - \tanh(\kappa/p)\cos(\theta - \mu)\}} d\theta$$
$$= \frac{\gamma^p x^{p-1} e^{-\gamma x} I_0(\gamma x \tanh(\kappa/p))}{\Gamma(p) \cosh^p(\kappa/p) P_{p-1}(\cosh(\kappa/p))}.$$

Then the conditional density of pdf (1) given angle is the Gamma distribution $\Gamma(p, \gamma(1 - \tanh(\kappa/p)\cos\theta))$:

$$f(x|\theta) = \frac{f(x,\theta)}{f(\theta)} = \frac{\gamma^p \{1 - \tanh(\kappa/p)\cos\theta\}^p}{\Gamma(p)} x^{p-1} e^{-\gamma x \{1 - \tanh(\kappa/p)\cos\theta\}}.$$

The mode of this distribution is taken at $x = (p-1)/\{\gamma\{1-\tanh(\kappa/p)\}\}\$ if p > 1.

The conditional density of pdf (1) given length is the sine-skewed von Mises distribution $SSVM(\mu, \gamma x \tanh(\kappa/p), \lambda)$ (Abe and Pewsey, 2011):

$$f(\theta|x) = \frac{f(x,\theta)}{f(x)} = \frac{(1+\lambda\sin\theta)e^{\gamma\tanh(\kappa/p)x\cos(\theta-\mu)}}{2\pi I_0(\gamma x\tanh(\kappa/p))}.$$

The distribution is unimodal if $D \ge 0$ and bimodal if D < 0, where D is given by

$$D = (1 - \lambda^2)(\gamma \tanh(\kappa/p)x)^6 + \lambda^2 (16\lambda^4 - 20\lambda^2 + 3)(\gamma \tanh(\kappa/p)x)^4 + \lambda^4 (8\lambda^2 + 3)(\gamma \tanh(\kappa/p)x)^2 + \lambda^6.$$
 (cf. Abe and Pewsey, 2011).

Therefore, it is easy to generate random numbers from the distribution with pdf (1).

References

- [1] Abe, T. and Pewsey, A. (2011). Sine-skewed circular distributions. *Statistical Papers*, **52**, 683–707.
- [2] Johnson, R.A. and Wehrly, T.E. (1978). Some angular-linear distributions and related regression models. *Journal of the American Statistical Association*, **73**, 602–606.
- [3] Kato, S. and Shimizu, K. (2008). Dependent models for observations which include angular ones. *Journal of Statistical Planning and Inference*, **138**, 3538–3549.
- [4] Mardia, K.V. and Sutton, T.W. (1978). A model for cylindrical variables with applications. Journal of the Royal Statistical Society: Series B, 40, 229–233.

放射線被曝関連データの再分析 濱岡豊 (慶応大学商学部)

1. 研究の背景と目的

本研究の目的は、以下の3点である。低線量被曝での健康影響がどのように検証されているのか分析方法について批判的に検討すること。可能ならば再分析を行うこと。統計学の放射線低線量被曝研究への貢献のありかたを検討すること。

2.「放射線と健康影響」データの再分析

1) 原爆被爆者データ

広島・長崎の被爆者データの分析については、被曝線量など連続量のカテゴリ化、個人レベルのデータの層別集計による情報損失、モデル選択のあいまいさ、被曝線量の一次項と二次項を導入したモデルにおける多重共線性等の問題がある。

公開されている LSS14 データを用い、固形ガンについて再推定した 1 。線量-応答関数は0zasa et al. (2012) で用いられたモデルおよび閾値を直接推定するモデルも推定した。多重共線性や収束など、推定に問題のあるモデルを除くと、全域 L もしくは閾値を $1^{\sim}20$ mGy とした (手動) 閾値モデルのあてはまりが最良となった(表)。なお、閾値そのものを推定したところ、-23mGy となり 10%水準で有意ではなく (t=-0.0876, p>0.1)、これよりも線形モデルの方があてはまりは良好であった。

このデータはもともと個人レベルのデータについて、線量などの連続量をカテゴライズし、層別に集計したものである。線量間でのサンプルの分布をみると、0.005Gy (5mGy)以下に 5 万人が含まれている。低線量での影響を評価するには、低線量部分のカテゴリを細分化するか、個人レベルでの分析を行うべきであろう。

2) 米国核関連 3 施設従業者調査

Hanford 他米国核関連 3 施設従業者調査(平均被曝量 21-35mSv)について、Gilbert et al. (1993) は個人データを層別に集計してトレンド検定やポアソン回帰を行い、放射線被曝量とがん死には有意な関係はないとした。これについては個票データが公開²されているので、「死因毎の 2 項ロジット・モデル」「生存を基準とした多項ロジット・モデル」「タイミングを考慮した比例ハザードモデル」という個人レベルのモデルで推定した。その結果、いずれも被曝量は正で有意となった。

3) 放射線生物関係のデータ

Oghiso et al. (1994)は、ラットにプルトニウム霧を吸引させ、死亡するまで飼育した。このようにして収集したデータについて log (内部被曝量)と生存期間をプロットしたところ 1Gy を越えると生存期間が急に短くなるようにみえ閾値の存在を示唆している。ただし、横軸に(対数ではなく)線量をプロットすると平均寿命は線量に対して線形で減少しており閾値のような現象は見られないことがわかった。公開されている個票データを用いても同様であった3。

Nomura et al. (2013) は老化の早い klotho 変異マウスに対して 4 水準の γ 線を照射し、死亡するまで飼育した。このデータに \log rank 検定を行い、5%水準で帰無仮説が棄却されたとし、低線量被曝による寿命延長効果が観察されたとしている。ただし、生存曲線をみると最後の数匹が大きく影響している。グラフからデジタイザでデータを読み取り、各時点での生存数を重みとして考慮するWilcoxon 検定を行うと、被曝量群間で生存曲線が異なるという帰無仮説を棄却できなかった (χ^2 (3) =3.7、p= 0.291)。

4) 福島県甲状腺検査データ

福島県で行われている甲状腺検査について、公開されている市町村レベルのデータ(二次検査確定数)を用いて分析した。検査時年齢、被曝時年齢、2次検査結果確定率および log(1+被曝量)を用いてポアソン回帰を行った。

WHO による甲状腺被曝量 $(15^{\sim}95\text{mSv})$ は「悪性もしくは疑い」については有意とはならなかったが (係数=-0.410, t=-0.635, p=0.324)、結節については有意となった。被曝量の係数の t 値は、5.1mm 以上の結節 (係数=0.287, t=1.73, p=0.089) よりも、5mm 以下の結節の方が高くなった (係数=0.622, t=6.02, p<0.001)。結節が成長するには時間がかかることを考えると、整合的な結果ともい

¹放影研「寿命調査 第 14 報 がんおよびがん以外の死亡率データ、1950-2003」 http://www.rerf.or.jp/library/dl/lss14.html ²米国エネルギー省 Comprehensive Epidemiologic Data Resource (CEDR)。 https://www3.orau.gov/CEDR/

³ 放医研「プルトニウム内部被曝に関する動物実験病理データベース」 http://www.nirs.go.jp/db/anzendb/PuDB/PuDB.html.

える。被曝量と結節には正の相関があること、結節がある者の方がその後の甲状腺ガンの発症確率が高いことを示す先行研究も存在する(Nagataki et al. 1994)。

この分析は公開されている市町村レベルのクロスセクションデータを用いた ecological study であり、因果関係を主張することはできない。個人レベルの例えば、ケースコントロール分析によって、被曝量との関係を分析すべきである。もしくは研究者が分析できるよう匿名化データを公開すべきである。もちろん、検査の継続、健康対策が必要であることはいうまでもない。

3. 低線量被曝問題への統計学の貢献可能性

原爆被爆者データの分析は、放射線疫学でもっとも重要な研究であり、線量のカテゴリ化、個人データの層別集計、ポアソン回帰によるリスク評価という手順は、Chernobyl 周辺住民、Mayak 従業員、Techa 川流域住民、国内の原発従業員、原発従業員の国際比較などでも用いられている。このような手法は計算機のパワーが低い時代には正当化されたが、現在はそのような制約は存在しない。個人レベルデータを用いた再分析を行うべきである。そのためには、時間共変量および潜伏期間を導入したハザードモデルが有効であろう。このようなモデルの開発および、当該分野の専門家への統計モデリングについての教育も必要である。なお、このような再分析が可能になったのは、データが公開されているからである。匿名化データの公開を促進する必要がある。

さらに、「100mSv 以下にサンプルを限定すると、回帰係数=0 という帰無仮説が棄却されなかった」という結果を、「100mSv 以下では健康への影響がない」と解釈する例もみられる。これは仮説検定の意味が理解されていないためである。専門家の統計的知識の向上とあわせて、このような情報を伝えるメディア関係者、さらには一般の方々への統計教育も重要である。

謝辞等

・広島、長崎被曝者データ

用いたデータは広島および長崎の放射線影響研究所(放影研)から入手したものである。放影研は、日本の厚生労働省(厚労省)ならびに米国のエネルギー省(DOE)により資金提供を(後者については、その一部を米国学士院に対するDOE 研究助成金 DE-HS0000031 を通じて)受けている公益財団法人である。この報告書に示した結論は著者のものであり、必ずしも放影研またはその資金提供機関の判断を反映するものではない。

・米国核施設労働者データ

Access to nuclear worker data was granted by the US DOE CEDR project. The protocol and results of this study were not reviewed by the DOE. The results and conclusions do not necessarily reflect those of the US Government or DOE.

		適合度			推定值	と有意水準			推定上の問題、
モデル		AIC	BIC	関値.境界値 (mGy)	L1	Q1	LもしくはL2	QもしくはQ2	方法
全域	L	18307.0	18317.9		L2と共通		0.423***		
	LQ	18308.2	18321.8		L2と共通		0.361***	0.038	多重共線性
	Q	18330.7	18341.6		L2と共通			0.218***	
(手動)閾値	0+L2	18307.0	18317.9	1に固定	0に固定		0.423***		
	0+L2	18306.8	18317.7	5に固定	0に固定		0.423***		
	0+L2	18306.9	18317.9	10に固定	0に固定		0.422***		
	0+L2	18307.2	18318.1	20に固定	0に固定		0.420***		
	0+L2	18308.2	18319.2	50に固定	0に固定		0.416***		
	0+L2	18309.4	18320.3	100	0に固定		0.412***		
線量カテゴリダミ・	一変数	18318.1	18380.9						
線型スプライン	L1+L2	18308.9	18322.5	1に固定	20.430		0.426***		
	L1+L2	18305.2	18318.9	5に固定	-22.160**		0.420***		収束せず
	L1+L2	18308.8	18322.4	10に固定	-2.146		0.420***		
	L1+L2	18308.8	18322.5	20に固定	1.209		0.427***		
	L1+L2	18308.5	18322.2	50に固定	0.884		0.427***		
	L1+L2	18308.7	18322.3	100に固定	0.645		0.426***		
2Gyで屈折	L1+L2	18308.8	18322.4		0.398***		0.433***		
	L1Q1+L2Q2	18306.6	18325.7		0.626	-0.089	0.211**	0.181**	多重共線性
	L1Q+L2	18304.8	18321.2		0.213**	0.181**	0.385***		多重共線性
	Q1+Q2	18309.2	18322.8			0.135***		0.330***	
閾値推定		33286.9	33781.6	-23.0			0.417***		P.O
全域	L	33285.0	33759.8				0.414***		— Rのoptimで推定

表 被爆者データ(LSS14: 固形ガン)の再推定結果

注)***:1%水準で有意 *:5%水準で有意 *:10%水準で有意。

- ·Ozasa et al. (2012)と同様、線量の修飾項には性別、被曝時年齢、到達年齢を導入し、ベースラインに都市、性別、被曝年齢カテゴリ、到達年齢カテゴリを導入した。
- ·L、Q はそれぞれ線量の1次項、2次項を導入したことを意味する。
- ・推定値の欄 L1、Q1 は閾値(境界値)よりも低い方、L2、Q2 は高い方の ERR の推定値である。
- ・線形スプライン、(手動) 閾値モデルについては閾値(もしくは境界値)を所与として推定した。
- ・閾値推定モデルは R の optim 関数で対数尤度を最大化することによって推定した。それ以外は AMFIT で推定した。これらは AIC、BIC の 定義が異なるので比較はできない。最下段の全域 L モデルは閾値推定モデルと比較するため R の optim 関数で推定した結果である。

新しい判別分析の紹介

成蹊大学 経済学部 新村秀一

判別分析には、次の4つの問題がある.

- 1) 誤分類数最小化 (Minimum Number of Misclassifications, MNM) 基準による最適線形判別関数 (**改定 IP-OLDF**) 以外の判別 関数は、判別超平面上のケースを正しく判別できないので、正しい誤分類数が求まらない.
- 2) ハードマージン最大化 SVM (H-SVM) と**改定 IP-OLDF** 以外の判別関数は、線形分離可能なデータ (MNM=0) を認識できないことが多い. そして公表している実証研究で Fisher の線形判別関数 (Fisher の LDF) は、線形分離可能な 18 個の分析結果で、誤分類確率の範囲が[0.17,0.23]であることが分かった。すなわち、過去の医学診断等で例え誤分類確率が 23%であっても、実際は線形分離可能なことがあるので、過去の重要な判別分析を用いた研究を見直す必要がある.
- 3) 変数値が一定値をとる場合, Fisher の LDF と 2 次判別関数 (QDF) は分散共分散行列が計算できないので分析からこれらの変数を省く統計ソフトが多い. 一般化逆行列を用いればこれに対応できるが, 一方の群に属する変数値が一定値をとる場合に QDF と正則化判別分析で深刻な問題を起こすことが分かった.
- 4) 判別分析を重回帰分析と同じ推測統計学的手法と間違って考えている人が多い. しかし, 重回帰分析のように判別係数や誤分類確率の 95%信頼区間がなく, 重回帰分析のような洗練されたモデル選択法がない. このため, 「小標本のための k-重交差検証法」を考えた. これによって, 改定 IP-OLDF を Fisher の LDF, ロジスティック回帰, H-SVM (MNM=0 のデータのみ), ソフトマージン最大化 SVM (S-SVM), 改定 LP-OLDF, 改定 IPLP-OLDF の 7 個の線形判別関数を, 学習標本と検証標本の平均誤分類確率で比較検討した. 検証に用いたデータは, Fisher のアイリスデータ, スイス銀行紙幣データ, CPD データ, 学生の成績データ, 18 種類の統計入門の 10 卓 100 間の合格水準を 10%, 50%, 90%の 3 水準による合否判定, 日本車 44 車種の 2 群判別データである. これら全てで, 改定 IP-OLDF の判別結果が良く, また判別係数の 95%信頼区間の検討を行った. 次の表は合否判定データによる 100 重交差検証法の結果である (M1 と M2 は学習標本と検証標本の平均誤分類確率、M1Diff. と M2Diff. は 7 手法と改定 IP-OLDF の差である).

	M1	M2	Diff.		
MNM	0.00	0.07	0.07		
33s	0.00	0.00	0.00		
4, 2	0.00	0.01	0.01		
	0.00	0.00	0.00	M1Diff.	M2Diff.
HSVM	0.00	0.81	0.81	0.00	0. 73
32s	0.00	1. 15	1. 15	0.00	1. 15
4, 0	0.00	0.90	0.90	0.00	0.89
	0.00	0.90	0.90	0.00	0. 90
SVM4	0.00	0.81	0.81	0.00	0. 73
31s	0.00	1. 15	1. 15	0.00	1. 15
4, 0	0.00	0.89	0.89	0.00	0.88
	0.00	0.90	0.90	0.00	0. 90
SVM1	0.00	0.81	0.81	0.00	0. 73
31s	0.72	1. 59	0.87	0.72	1. 59
1,0	0.13	0.98	0.85	0. 13	0. 97
	0.76	1.71	0.95	0.76	1. 71

IPLP	0.00	0.56	0.56	0.00	0.49
45s	0.00	0.63	0.63	0.00	0.63
4, 0	0.00	0.49	0.49	0.00	0.48
	0.00	0.47	0.47	0.00	0. 47
LP	0.00	0.96	0.96	0.00	0.89
28s	0.00	1. 10	1.10	0.00	1. 10
4, 0	0.00	0.95	0.95	0.00	0.94
	0.00	0.80	0.80	0.00	0.80
Logistic	0.00	0.96	0.96	0.00	0.89
20m50s	0.00	1. 10	1. 10	0.00	1. 10
4, 0	0.00	0.95	0.95	0.00	0.94
	0.00	0.80	0.80	0.00	0.80
LDF	9. 64	10. 54	0.90	9. 64	10.47
4m50s	9.89	10.55	0.66	9.89	10. 55
0, 0	9. 48	10.09	0.61	9.48	10.08
	9. 54	9. 91	0.37	9. 54	9. 91

これらに加えて改定 IP-OLDF の研究を通して、1)判別係数と誤分類数の関係、2)MNM の単調減少性、3)最適凸体、という 3 個の判別分析における新知見が分かった。

本研究結果は、以下の5種のデータ(Fisher のアイリスデータ、スイス銀行紙幣データ、CPD データ、学生の成績データ、日本車44車種の2群判別データ)でも確認したが、紙面の都合で合否判定にとどめた。特に線形分離可能なデータの研究は行われていないが、著しい成果が得られている。2014年には英語論文に詳細を発表している[35-37].以下はその一部である.

- 1) スイス銀行紙幣データでは2変数 (X4, X6) で MNM=0 であり,6 個の説明変数の全ての組み合わせ63 個のうち16 個で MNM=0 になる.そして,8 種類の線形判別関数を比較し,H-SVM と改定 IP-OLDF 以外は MNM=0 になる判別モデルで,線形分離可能を認識できない事例を数多く確認した。また学習標本と検証標本の平均誤分類確率の比較結果がより明確になることを確認した.
- 2) 普通車 29 車種と小型車 15 車種を 6 個の性能で判別する「車の判別データ」は、総排気量と座席数の各 1 変数で MNM=0 になるので、63 個の判別関数のうち 48 個が MNM=0 になる、判別をこの 2 種類で比較すると明確な結論が得られることを確認した。

判別分析は、医学診断、ゲノム診断、パターン認識などで重要な手法である。すでに述べたとおり既存の判別関数には問題があり判別結果も悪い。例えば医学診断で誤分類確率が大きいので、間違った結論や研究を諦めたものも多いと考え、重要な研究課題は見直す必要がある。

共同プロジェクトの提案

判別分析にかかわる研究を見直すために、学会あるいは医学部と共同研究を立ち上げることを提案したい. 筆者は、判別分析にかかわる計算を引き受ける、あるいはそれに関する技術を提供することで協力したいと考えている. このようなテーマを持っておられる関係者を知っておられる場合は、ぜひ紹介してください.

謝辞

本研究で利用した JMP の Fisher の LDF とロジスティック回帰の 100 重交差検証法プログラムは、SAS ジャパン㈱の JMP 事業部の協力を得て開発した。また、8 種の数理計画法による線形判別関数の開発は、LINDO Systems Inc. の協力で開発した。

参考文献

- 12) J. Sall (新村秀一訳): SAS による回帰分析の実践, 朝倉書店, 東京, 1986.
- 13) J.P. Sall, L. Creighton, & A. Lehman (新村秀一監修): JMP を用いた統計およびデータ分析入門 (第 3 版). SAS Institute Japan (株), 2004.
- 14) L. Schrage: LINDO -An optimization modeling system-. The Scientific Press, 1981. [新村秀一,高森寛,実践数理計画法.朝倉書店,1992.]
- 15) L. Schrage: Optimizer Modeling with LINGO. LINDO Systems Inc, 2003.
- 21) 新村秀一: 数理計画法を用いた最適線形判別関数. 計算機統計学. 11/2,89-101,1997.
- 24) 新村秀一: JMP による統計学とっておき勉強法, 講談社, 2004.
- 27) 新村秀一: Excel と LINGO で学ぶ数理計画法, 丸善, 2007.
- 30) 新村秀一:最適線形判別関数.日科技連出版社,2010.
- 31) 新村秀一: 数理計画法による問題解決法. 日科技連出版社, 2011.
- 33) 新村秀一: SAS/JMP との歩み, SAS Technical Report, 13/16, 2012.
- 34) S. Shinmura: End of Discriminant Functions based on Variance Covariance Matrices. ICORER, 5-14, 2014.
- 35) S. Shinmura, "Improvement of CPU time of Linear Discriminant Functions based on MNM criterion by IP," Statistics, Optimization and Information Computing, vol. 2, June 2014, pp 114-129.
- 36) S. Shinmura, "Three Serious Problems and New Facts of the Discriminant Analysis" Operations Research and Enterprise Systems, ICORES 2014, Revised Selected Papers, in Press.
- 37) S. Shinmura (2014). Comparison of Linear Discriminant Functions by K-fold Cross Validation. Data Analytic 2014, 1-6, 2014.

鉛同位体法の統計的研究

吉田 知行 (北大)

概要: 鉛同位体法では、古代青銅器に含まれる鉛の同位体比鉛同位体比を測定することによって、青銅器に含まれる鉛の産地や青銅器そのものがどこで作られたかの明らかにする. ここでは、鉛の混合の過程を推定するための数学的方法とその問題点を紹介する.

キーワード: 鉛同位体法,鉛の混合,散布図,椿井大塚山古墳,鉛インゴット

1 はじめに

この研究は 2010 年に熊本大学でおこなわれた統計の研究集会での講演内容を発展させたものである. その後, いくつかの進展があったものの, 新たな困難 (無視できない数のデータの間違い) の発生により, 研究は足踏み状態だった. ようやく, 解決の見通しが立ったので, 研究の現状を紹介したい.

鉛同位体比の散布図には、いたる所に鉛の凝集、直線状の鉛の並び、長円状の鉛の分布があるく.とくに、直線状の鉛の並びが生じる原因として、鉛の混合の可能性が指摘されてきた(馬淵、久野など).そもそも散布図内の点の数が多すぎて、全部が鉱山の鉛から来ているとは考えられない.鉛の混合の可能性を考えるのは自然である.ところが鉛の混合があると、単純な鉛同位体法は使えない.

講演では、椿井大塚山古墳出土鏡のデータ (馬淵 1996 科研費報告集) を例にして、鉛同位体法への統計学の応用を考える。

鉛の混合を考えるときは、同位体比 (率) の精度が問題になる。ところが鉛同位体比 (率) の持つ次の性質 (表 1 参照) のため、データの分析が困難になる。

- (1) 存在率のアンバランス. Pb204 だけが 1.35% 程度と少ない.
- (2) 同位体比の有効桁数の不一致. Pb207/Pb206 が 4 桁, Pb206/Pb204, Pb208/Pb206 は 5 桁.

(3) 同位体比測定精度のアンバランス. Pb206/Pb204 の測定誤差は 0.06%, 他は 0.02% と言われる.

さらに、鉛同位体比のデータは以下の問題を抱えている:

- (4) 再測定が不可能である. 鉛同位体法は破壊検査であり、国宝を破壊するわけにはいかない.
- (5) 精度が悪い $(0.02 \sim 0.06 \, \text{パーセントと言われる})$.
- (6) 信頼できない数値が数多くある.

結局鉛同位体法では、このような精度の悪く信頼性に不安を抱える古い測定結果を扱わざるを得ない.

2 鉛の混合

ここでは鉛を同位体率の 4 次元ベクトルと見なす. 2 次元か 3 次元空間に射影した散布図で考える. 鉛の混合に関しては次の原理が成り立つ.

基本原理:鉛Xが鉛 A,B,\cdots の混合比p:q: の混合なら, $X=pA+qB+\cdots$.

したがって、3次元散布図内での平面状の鉛の分布、2次元あるいは3次元散布図内での直線状の鉛の並びは、鉛の混合を示唆する.

基本問題:鉛同位体率 (あるいは比) のデータから鉛の混合の過程を明らかにせよ.

例. 椿井大塚山古墳の2,34,5号鏡の鉛同位体率は

- $\begin{bmatrix} 2 \end{bmatrix} = (0.013597, 0.246250, 0.212858, 0.527295)$
- $\boxed{34} = (0.013619, 0.247108, 0.212908, 0.526365)$
- 15 = (0.013607, 0.246641, 0.212876, 0.526875)
- (2)と(34)を 5:4 で混ぜた鉛 (5/9)(2)+(4/9)(34)

(0.013607, 0.246631, 0.212880, 0.526881)

は、5号鏡と区別できない、差は

 $(0.0^400, 0.0^410, -0.0^404, -0.0^406)$

で丸め誤差の範囲内である. つまり,5号鏡の鉛は,2号鏡と34号鏡の鉛を5:4で混ぜてできた.

3 平面状配置の探索

椿井大塚山の鉛を例に説明する.

作業 2-1: 鉛同位体率の 3 次元散布図内の 4 点以上 からなる平面状配置をさがす.

作業 2-2: 各平面の生成点をさがす. 生成点とは, その平面に属する鉛の源になった3つの鉛である.

その分析と確率論的考察から次のことが分かる.

結論 2-1:31 の鉛のうち 27 の鉛がひとつの平面上 にある (残差 20ppm 未満).

結論 2-2 (精度仮説): 鉛同位体比の測定精度は極め て高い. データの最後の桁まで信用できる (表 1). 結論 2-3 (3 つの親仮説): 椿井大塚山古墳出土鏡の 鉛は中国の 3 系統の鉛の混合で得られた.

平面状分布を探し出すのに回帰分析の繰り返しを 用いた.これほど多くの鉛がひとつの平面上にある ことは偶然とは考えられない.鉛の混合の結果と考 えるのが自然であろう.

残差の分析から、鉛同位体率の偶然誤差は、従来いわれてきた誤差よりも桁違いに小さい。質量依存誤差 (系統誤差) は、分散、定数項を除く回帰係数、直線関係、平面関係、ユークリッド距離、マハラノビス距離などに影響しない。

椿井大塚山の場合、3つの親 X,Y,Z の有力候補は、山東省香奇鉱山(X)、湖南省瑪瑙山鉱山(Z)、そ

れに系統不明の鉛 (Y) である。(Y) の候補のひとつは 37 号鏡の鉛だが,その近くに鉱山の鉛がないので,(Y) は鉱山の鉛と考えるより,殷や中国四川省三星堆遺跡の青銅器の鉛のスクラップと考えるのが自然である。

4 直線状配置の探索

作業 3-1: 散布図内の直線状配置の探索.

作業 3-2:直線状配置での混合比の計算.

作業 3-3:整数混合比の探索.

3つの鉛 AXB の直線度を、直線 AB と X の「距離」として定義する.「距離」としては、ユークリッド距離などが考えられる。また、グラスマン多様体 G(4,2) 内の点が集積している所は直線状の並びを示す。このような代数幾何的アプローチも考えられる。

混合比 p:q が整数混合比になっているかは、p/q (または q/p) の連分数展開か、ファレイ数列を使えば分かる.

椿井大塚山の場合, 2-15-34(=19)-36, 7-3-25-14-24, 5-31-20, 2-12-6, 1-5-12 等が直線状並びである.

椿井大塚山の鉛の製造に関する仮説:

- ・直線 37–2–1 はライン D とよばれる. この順番 に新しくなる. ライン D の鉛は, 三星堆などの古 い青銅器のスクラップと北方の鉛の混合で得られた (新井説).
- ・残りの鏡 (すべて三角縁神獣鏡) の鉛はライン D 上の鉛と江南の鉛の混合の繰り返しで得られた.
- $\cdot 2 + 36 \rightarrow 15,35, \ 2 + 6 \rightarrow 12; \ 1 + 12 \rightarrow 5;$ $7 + 24 \rightarrow 3,25,14; \ 5 + 20 \rightarrow 31.$
- \cdot 2 や 34 などで鉛インゴットが使われた可能性がある.

中規模自治体病院における 経営に関する要因分析

〜健全な経営の施策は? 総務省地方行財政データから見えるモノ〜

慶應義塾大学大学院 健康マネジメント研究科 神戸 翼

1. はじめに

地域医療の中核である自治体病院の実に 6 割以上 が赤字¹⁾と言われている今日、自治体病院の効率化と 経営改善が叫ばれている。しかしながら、自治体病 院の中には、不採算医療や僻地医療の問題を抱え、 国や自治体からの補助金等を除いた医業収支単体で の黒字化は困難な状況が見受けられる。このような 状況を踏まえ、病院経営の実態を適時把握し、的確 な意思決定を行う必要がある。そこで本研究では、 全国の自治体病院の内、経常利益率が低い 200 床~ 399 床までの市町村立病院(以下、中規模自治体病院 という)を対象として、医業経営の指標である「医 業収支」に与える要因の分析を行う事とした。加え てその分析を通して、中規模自治体病院の黒字化に 向けた策と可能性を検討する事を目的とした。尚、 医業収支とは入院収益や外来収益などの医業活動か ら生じる医業収益と、給与費、材料費、経費などの 医業活動に要する医業費用との差額金額を指す。

解析に当たっては JMP11.0 を使用し、変数増減法で重回帰式を作成した。回帰診断ではテコ比と t 値の吟味を行い、必要な場合にはデータのマスクを行った。また、VIF を基に説明変数間の多重共線性に注意し、大きな問題がないことを確認した上で各変数の係数や寄与率の解釈を行っている。

2) 概念図

健全な医療経営を遂行するために、「社会」「個人」 「経営」「医療」の4つのフィールドより可視化した。 概念図として図1に示す。医療経営において重要視 される、「質の高い医療」と「黒字経営」の2点を主 軸に、各フィールドにおいて影響する要因を検討した。

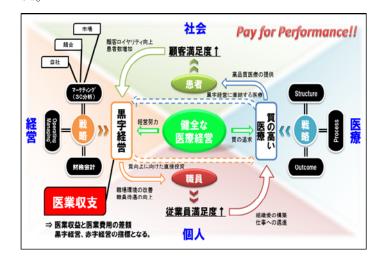


図1 概念図(医療経営)

2. 方法

1) 前提条件

表1に前提条件を示す。

表 1 前提条件

マクロ								
時間	エリア							
2012年	全国 47 都道府県							
ミク	ロ							
種類	病床規模							
市町村立病院	200~399 床							

2) 特性要因図と変数の候補

本研究では、図2に示す特性要因図を用いた。「医療の質」と「経営の質」の観点から検討し、Structure、Process、Outcome、マーケティング3C(自社、競合、市場)、財務会計、Operating Management を上位概念として、変数候補を選択した。

4) 重回帰分析に用いたデータ

中規模自治体病院 151 施設のデータを総務省地方 公営企業年鑑、各自治体運営の医療情報ネット、及 び2次医療圏データベース WELLNESS より収集した。 6,7,8)

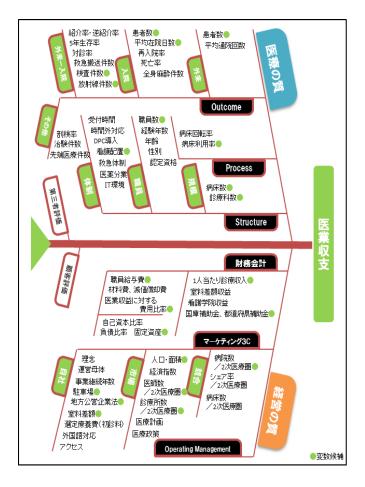


図 2 特性要因図

3. 結果と考察

1) 結果の要約

(1) 自由度調整済寄与率: 0.63

(2) 残差の標準偏差:308806

(3) Yの平均:-351966

(4) 重回帰式: データ数: 151(マスクデータ: 21)

Y	=	1295347.6		<定数項>
	+	10930.098	X7	病床利用率
	-	64.674	X15	室料差額(個室・最低額)
	-	0.902	X27	都道府県補助金
	+	18.132	X31	看護職員1人1日当たり診療収入
	-	3.771	X34	職員給与費(看護職員)
	-	1.644	X35	職員給与費(事務職)
	-	0.760	X36	職員給与費(その他)
	-	15.750	X39	1床当たり固定資産

2) 考察

得られた回帰式において、各々の係数の符号と絶対値を見ると、常識の範囲内で納得のできる回帰式であると言える。各々の変数の係数を見てみると他の変数の値を固定した上で、図 3 に示すモデル図が導き出された。

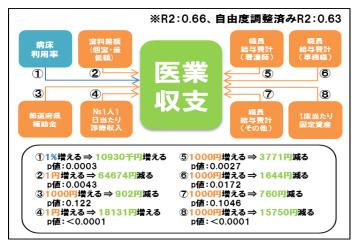


図3 モデル図

<u>4. おわりに</u>

中規模自治体病院は 200~399 床の病床を抱え、 急性期医療と慢性期医療を提供するという非常に 難しい立場にある。これは外来診療のみだけでな く、入院診療も重要な要素であることを示唆して いる。本研究にて「病床利用率」1%の違いで10930 千円の医業収支への影響がある事が分かっており、 中規模自治体病院にとって重要な指標であること が示唆された。また、「病床利用率」を如何に高め ていくかが、今後の経営改善のカギであると考え られる。尚、今回の研究では、病床数を 200~399 床の規模の自治体病院に焦点を当てている。その ため、規模の異なる病院や私立病院においても同 様の分析を行い、結果の比較検討を行う事が求め られる。これにより、更なる医療経営における問 題点と改善点の糸口が見つかるのではないかと考 えられる。

【参考文献】

- 1) 厚生労働省 (2010): 医療施設調査
- 2) Donabedian A (1966) : Evaluating the quality of $\,$ medical care. Milbank Mem Fund Q,44(3) : 166-203
- 3) Donabedian A (1980): The definition of quality and approaches to its assessment. Exploration s in quality assessment and monitoring (Volume1) .Health Administration Press, Ann Arbor, Michigan
- 4) 大坪徹也 (2008): 自治体病院の医業収支推移に関する規模 別要因分析:日本公衆衛生誌,55-11
- 5) 足立泰美 (2013): 自治体病院経営の効率性-医療機関の機能分化と地域医療連携-: 会計検査研究,47
- 6) 総務省: 地方公営企業年鑑 2012: http://www.soumu.go.jp/main_sosiki/c-zaisei/kouei_kessan.html
- 7) 厚生労働省、各自治体: 医療情報ネットワーク: http://www.mhlw.go.jp/stf/seisakunitsuite/bunya/kenkou_iryou/iryou/teikyouseido/
- 8) 2 次 医 療 圏 デ ー タ ベ ー ス WELLNESS : http://www.wellness.co.jp/siteoperation/msd/

A Finite Sample Efficiency of a GLSE in Regression with Inversely Linear covariance Structure

Takeaki Kariya, Meiji University and H. Kurata, University of Tokyo

In this paper, we revisit the upper bound problem for a GLS with inversely linear covariance structure with respect to the Nonlinear Version of the Gauss-Markov Theorem that gives the Lower bound for the covariance matrices of GLSEs in terms of the ordering of nonnegative definiteness. More specifically in a regression model,

$$y = X\beta + \varepsilon$$
 with $P \equiv \mathcal{L}(\varepsilon) \in \mathcal{P}_n(0, \sigma^2 \Sigma)$ with

$$\mathcal{P}_n(0,\Omega) = \{P: E(\varepsilon) = 0 \mid Cov(\varepsilon) = E(\varepsilon\varepsilon') = \Omega \in \mathcal{S}(n)\}$$

we are interested in deriving an upper bound for the covariance matrix of a GLSE

OLSE
$$b(I_n) = (X'X)^{-1}X'y$$
 belongs to C_0
GLSE $b(\Sigma(\hat{\theta})) = (X'\Sigma(\hat{\theta})^{-1}X)^{-1}X'\Sigma(\hat{\theta})^{-1}y$
when $\Omega = \sigma^2\Sigma(\theta)$

where θ is estimated by the OLS residual e = y - Xb(I) and

$$\Omega = \sigma^2 \phi(\theta) \Sigma(\theta) = \sigma^2 \phi(\theta) (I + \sum_{j=1}^p \lambda_j(\theta) C_j)^{-1}$$

$$(\theta \in R^q), C_j \text{ 's are known.}$$

As for the lower bound of the GLSE, Kariya (1995) proved the nonlinear Gauss Markov Theorem;

$$Cov(b(\Sigma)) \leq Cov(\hat{\beta})$$
 for any $\hat{\beta} \in C_0$ and $P \in \mathcal{P}_n(0, \sigma^2 \Sigma)$
 $C_0 = {\hat{\beta} = Cy \mid C \text{ is a } k \times n \text{ matrix such that } CX = I_k}$
 $Cov(b(\Sigma)) = \sigma^2(X'\Sigma^{-1}X)^{-1}$

Here when the model of error terms follow AR(2) with

$$\Omega = \tau^2 \Phi \quad \text{with} \quad \Phi = \Phi(\theta) = \frac{1}{1 - \theta^2} \begin{pmatrix} \theta^{|i-j|} \end{pmatrix}$$

$$C_i \quad C_i \quad C$$

or equivalently

$$\Phi^{-1} = (1 - \theta)^{2} [I + \lambda_{1}(\theta)C_{1} + \lambda_{2}(\theta)C_{2}]$$
$$\lambda_{1}(\theta) = \frac{\theta}{(1 - \theta)^{2}}, \quad \lambda_{2}(\theta) = \frac{\theta}{(1 - \theta)}$$

we derive an upper bound for a GLSE under the linearly inverse covariance structure relative to this lower bound.

$$\begin{split} R(b(\hat{\Sigma}),\beta) &= Cov(b(\Sigma)) + E\left\{ \left(b(\hat{\Sigma}) - b(\Sigma)\right) \left(b(\hat{\Sigma}) - b(\Sigma)\right)'\right\} \\ &= \sigma^2 \phi(\theta) A^{-1} + \sigma^2 \phi(\theta) A^{-1/2} E(\Delta \Delta') A^{-1/2} \\ \Delta &= (\bar{X}' \bar{\Sigma}^{-1} \bar{X})^{-1} \bar{X}' \bar{\Sigma}^{-1} \bar{Z} \eta_2 : k \times 1 \\ \text{Want E}(\Delta'\!\Delta) &\leq \text{y/ for some y} \\ a' \Delta &= a' J^{-1} L \eta_2 \\ J &= J(\hat{\theta}) &= \bar{X} \ '\bar{\Sigma}^{-1} \bar{X} = \bar{X} \ '\bar{\Sigma}(\hat{\theta})^{-1} \bar{X}, \\ L &= L(\hat{\theta}) &= \bar{X} \ '\bar{\Sigma}^{-1} \bar{Z} \end{split}$$

In the inversely linear model of error terms, we have

Evaluation of
$$L = \overline{X} \, \overline{\Sigma} (\hat{\theta})^{-1} \overline{Z}$$
:

$$L = \overline{X} \, \Sigma^{1/2} \hat{\Sigma}^{-1} \Sigma^{1/2} \overline{Z}$$

$$= \overline{X} \, \Sigma^{1/2} [I + \sum_{j=1}^{p} \lambda_j C_j + \sum_{j=1}^{p} (\hat{\lambda}_j - \lambda_j) C_j] \Sigma^{1/2} \overline{Z}$$

$$= \sum_{i=1}^{p} (\hat{\lambda}_j - \lambda_j) \tilde{C}_j \quad with \quad \tilde{C}_j = \overline{X} \, \Sigma^{1/2} C_j \Sigma^{1/2} \overline{Z}$$

and

$$\begin{split} J &= \overline{X} \, {}^{\mathsf{L}} \Sigma^{1/2} \hat{\Sigma}^{-1} \Sigma^{1/2} \overline{X} = \overline{X} \, {}^{\mathsf{L}} \Sigma^{1/2} [I + \sum\nolimits_{j=1}^{p} \hat{\lambda}_{j} C_{j}] \Sigma^{1/2} \overline{X} \\ &= \overline{X} \, {}^{\mathsf{L}} \Sigma^{1/2} \{I + \sum\nolimits_{j=1}^{p} \lambda_{j} C_{j} + \sum\nolimits_{j=1}^{p} (\hat{\lambda}_{j} - \lambda_{j}) C_{j} \} \Sigma^{1/2} \overline{X} \\ &= I + \sum\nolimits_{j=1}^{p} [\hat{\lambda}_{j} - \lambda_{j}] \overline{C}_{j} \quad \textit{with} \quad \overline{C}_{j} = \overline{X} \, {}^{\mathsf{L}} \Sigma^{1/2} C_{j} \Sigma^{1/2} \overline{X} \end{split}$$

Using these evaluation, we obtain an explicit upper bound for the AR(2) error model, though it is necessary to investigate the sharpness of the upper bound through a numerical analysis.

Kariya, T. and Kurata, H. (2004) Generalized Least Squares, Wiley

Statistical Estimation for Optimal Dividend Barrier

慶應義塾大学理工学研究科 小泉 健太 慶應義塾大学理工学部 白石 博

保険会社の破産リスクに関する理論(risk theory, ruin theory)の応用として、会社の余剰資本(サープラス, surplus)がある境界(barrier)を上回ったときに、その部分を株主に返還する配当(dividend)の問題がある。本研究では、伝統的な複合ポアソンモデルを想定し、最適な境界(optimal dividend barrier)の推定問題を考える。

配当が無い場合の時刻 $t \geq 0$ におけるサープラス過程 $\{U(t)\}$ を

$$U(t) := u + ct - S(t); \quad S(t) = \sum_{i=1}^{N(t)} X_i$$

と定義する。ここで、 $u\geq 0$ は初期サープラス;c>0 は単位時間あたりの保険料率;N(t) は時刻 t までのクレーム(保険金請求)頻度; X_i を第 i 回目におけるクレーム金額;S(t) は時刻 t までのクレーム累積額とする。本研究では、伝統的なモデルとして、 $\{N(t)\}$ を強度 $\lambda>0$ を持つポアソン過程; X_i を i.i.d. で確率密度関数 f(x) を持つ正値確率変数とし、 $\{S(t)\}$ が複合ポアソン過程に従うと仮定する。また、 $b\geq u$ を配当境界(dividend barrier)とし、サープラス過程 $\{U(t)\}$ が b を上回る部分を配当として株主に返還すると仮定する。このときの区間 (0,t] における配当総額を D(t) とすると、微小区間 (t,t+dt] における配当額 dD(t) は

$$dD(t) = D(t + dt) - D(t) = \begin{cases} 0 & \text{if } U_b(t) < b \\ cdt & \text{if } U_b(t) = b \end{cases}$$

と表すことができる。ここで、 $\{U_b(t)\}$ とは、配当がある場合のサープラス過程であり

$$U_b(t) = U(t) - D(t)$$

と表すことができる。最適配当問題(optimal dividend problem)として、破産時刻 $T=\inf\{t|U_b(t)<0\}$ までに支払われる配当の現在価値の累積の期待値

$$V(u,b) = E\left[\int_0^T e^{-\delta t} dD(t)\right]$$

が最大となる $b(b^*$ とする。)を最適な境界とする。ここで、 $\delta > 0$ は瞬間利子率とする。

本研究では、n 組のクレーム金額およびクレーム発生間隔 $\{(X_1,T_1),(X_2,T_2),\dots,(X_n,T_n)\}$ が観測されたときの b^* を推定する。まず初めに、V(u,b) に関する次の Pollaczek-Khinchin 公式を導出する。

定理 1. $(Pollaczek ext{-}Khinchin\ 公式)\ |\lambda/c|<1\$ および f,V に関する適当な有界条件を仮定する。このとき、V(u,b) は次式で表現することができる。

$$V(u,b) = \alpha e^{-(b-u)/\alpha} \left\{ 1 - \frac{\lambda}{\lambda + \delta} \sum_{j=0}^{\infty} \left(\frac{\lambda}{c}\right)^{j+1} G^{(j)}(b) \right\}^{-1} \sum_{j=0}^{\infty} \left(\frac{\lambda}{c}\right)^{j} F^{(j)}(u,b)$$

 $\zeta \zeta \tau$, $\alpha = c/(\lambda + \delta)$,

$$F^{(0)}(u,b) = 1, \quad F^{(1)}(u,b) = \int_{u}^{b} \int_{0}^{s} f(x)e^{-x/\alpha}dxds,$$

$$F^{(j)}(u,b) = \int_{u}^{b} \int_{0}^{s_{1}} \int_{s_{1}-x_{1}}^{b} \int_{0}^{s_{2}} \cdots \int_{s_{j-1}-x_{j-1}}^{b} \int_{0}^{s_{j}} \prod_{k=1}^{j} f(x_{k})e^{-\sum_{h=1}^{j} x_{h}/\alpha}dx_{j}ds_{j} \cdots dx_{2}ds_{2}dx_{1}ds_{1} \quad j \geq 2,$$

$$G^{(j)}(b) = \int_{0}^{b} e^{-x/\alpha}f(x)F^{(j)}(b-x,b)dx \quad j \geq 0$$

とする。

次に、 $F^{(j)}$ を再生過程 (renewal process) の確率として表現する。

補題 1. $\{Y_i\}$ を以下の $\tilde{F}(y)$ を確率密度関数にもつような独立同一分布に従う確率変数の列とする。

$$\tilde{F}(y) := \frac{1}{C_Y} \int_0^y f(x) e^{-x/\alpha} dx$$

ここで、f は X_i の確率密度関数、 $C_Y = \int_0^\infty \int_0^y f(x) e^{-x/\alpha} dx dy$ である。このとき、 $F^{(j)}(u,b)$ は、

$$F^{(j)}(u,b) = \begin{cases} C_Y P(u \le Y_1 \le b) & \text{if } j = 1\\ C_Y^j \left\{ P(u \le Y_1 \le b) \prod_{l=2}^j P(0 \le Y_l \le b) - P(0 \le Y_1 + \dots + Y_j \le b) \right\} & \text{if } j \ge 2 \end{cases}$$

と表すことができる。

更に、U-statistics のクラスとしての $F^{(j)}$ の推定量 $F_n^{(j)}$ を用いて b^* の推定量を以下のように定義する。 定義 1.

$$\hat{b}_n^* := \hat{b}_n^*(u) := \arg \max_{b \ge u} V_n(u, b)$$

ここで、

$$V_n(u,b) := \hat{\alpha}_n e^{-(b-u)/\hat{\alpha}_n} \left\{ 1 - \frac{\hat{\lambda}_n}{\hat{\lambda}_n + \delta} I_n(b) \right\}^{-1} \sum_{j=0}^n \left(\frac{\hat{\lambda}_n}{c} \right)^j F_n^{(j)}(u,b),$$

$$I_n(b) := \sum_{j=0}^n \left(\frac{\hat{\lambda}_n}{c} \right)^{j+1} G_n^{(j)}(b), \quad G_n^{(j)}(b) := \frac{1}{n} \sum_{i=1}^n I(X_i < b) e^{-X_i/\hat{\alpha}_n} F_n^{(j)}(b - X_i, b)$$

 $\mathrm{Frees}(1986)$ を参考にすると、上記の V_n に関する一致性(一様収束性)を示すことができる。

参考文献

- 1. 清水泰隆 (2011) 危険理論における Gerber-Shiu 関数と統計的推測 『統計数理』統計数理研究所, 59(1), 105-124.
- 2. Dickson D. C. M.(2005) *Insurance risk and ruin*. International Series on Actuarial Science. Cambridge University Press, Cambridge.
- 3. Frees, E. W. (1986) Nonparametric Renewal Function Estimation. The Annals of Statistics. 14(4), 1366-1378.

Minimum contrast estimation for spectral densities based on exotic disparity

Yan LIU¹ (Waseda University)

Abstract

Minimum contrast estimation is a fundamental method for parameter estimation in time series analysis. There are several results for minimum contrast estimation based on minimizing the prediction error, interpolation error and some functions of them. In this talk, we extend the class of the disparity function to a much broader class and call it the exotic disparity. We investigate the properties of the class of the exotic disparity and propose a procedure for estimation. The consistency and asymptotic Gaussianity is shown for the class. Furthermore, the robustness against the fourth order cumulant of the class is also shown. Asymptotic efficiency only holds when the disparity is equivalent to the prediction error up to some constant multiples.

keywords: minimum contrast estimation, prediction error, interpolation error, exotic disparity

In time series analysis, one remarkable result for parameter estimation of the model is the method of minimizing the prediction error represented in the spectral density form. The concept of the spectral density function for a stationary linear process has a long history. Whittle (1952) systematically investigated the parameter estimation by the means of the spectral density for the first time after he found that it was difficult to derive the inverse matrix of parametrized variance matrix of Gaussian stationary process explicitly. He proposed the method to approximate the matrix by the spectral density and next estimate the parameters in the spectral density. In the case, Whittle suggested minimizing the functional $\int_{-\pi}^{\pi} I_{n,X}(\omega)/f_{\theta}(\omega)d\omega$ for the Gaussian stationary process $\{X_t; t \in \mathbb{Z}\}$.

The method has been generalized to be a minimization problem of a certain criterion (or disparity measure) $D(f_{\theta}, \hat{g}_n)$ in Taniguchi (1979). Here, the nonparametric estimator \hat{g}_n is substituted for the periodogram $I_{n,X}(\omega)$ because of non-consistency in the case of nonlinear integral functional of $g(\omega)$ in the criterion. The approach minimizing criterion D between the parametrized spectral density and the nonparametric estimator is called *minimum contrast estimation*. The asymptotic properties of the approach for time series case has been considered for a long time. Taniguchi (1981) proposed $D(f_{\theta}, \hat{g}_n) = \int_{-\pi}^{\pi} [\Phi(f_{\theta}(\omega))^2 - 2\Phi(f_{\theta}(\omega))\Phi(\hat{g}_n(\omega))]d\omega$ as the disparity measure between $f_{\theta}(\omega)$ and $\hat{g}(\omega)$ with a bijective function $\Phi(\cdot)$, and showed the asymptotic properties of the estimator. As another direction, Taniguchi (1987) proposed $D(f_{\theta}, \hat{g}_n) = \int_{-\pi}^{\pi} K(f_{\theta}(\omega)/\hat{g}_n(\omega))d\omega$ as the disparity measure with a sufficiently smooth contrast function $K(\cdot)$ whose minima exists uniquely at 1, and showed the asymptotic normality based on the contrast function $K(\cdot)$.

We mainly focus on the minimum contrast estimators based on the exotic disparity, which is

¹This work was supported by Grant-in-Aid for JSPS Fellows: 26·7404 (Liu, Y.)

defined by

$$D(f_{\theta}, I_{n,X}) = \int_{-\pi}^{\pi} a(\theta) f_{\theta}(\omega)^{\alpha} I_{n,X}(\omega) d\omega, \quad \alpha \neq 0,$$
(0.1)

where $a(\boldsymbol{\theta})$ is given by $a(\boldsymbol{\theta}) = \left(\int_{-\pi}^{\pi} f_{\boldsymbol{\theta}}(\omega)^{\alpha+1} d\omega\right)^{\beta}$. The disparity includes the integration functional form of prediction error and interpolation error up to some constant multiples.

- (i) the prediction error is given when $\alpha = -1$;
- (ii) the interpolation error is given when $\alpha = -2$ and $\beta = -2$.

The consistency of the disparity is guaranteed by the condition that $\beta = -\frac{\alpha}{\alpha+1}$. Under this condition, the extreme value of the exotic disparity in the parameter space will maximize the disparity if $\alpha > 0$ and vice versa, if the parametrized spectral density of the stationary process is contained in the parameter family. The property is shown by Hölder's inequality. In this talk, we suppose $\alpha < 0$ and take the procedure to choose the minimizer of the exotic disparity.

To investigate the asymptotic properties of the statistical procedure, we define the statistics based on the functional of the periodogram of the stationary process. The statistics are usually called the minimum contrast estimator since the properties between the fitted spectral density and the periodogram. The estimator is asymptotically normal from the central limit theorem for the linear process in Hosoya and Taniguchi (1982) based on assumptions that the innovation process can be approximated by some martingale processes. The Gaussianity then is shown by the martingale central limit theorem. If the assumption on the innovation process can be made stronger that the fourth order cumulants only exist as the simultaneous case, then the process is asymptotically normal and the asymptotic variance is robust against the fourth order cumulant if the true spectral density in contained in the parameterized spectral density family. Furthermore, the property can be made stronger if the considered process is 1-dimension, that is, the asymptotic variance is robust against the fourth order cumulant even if the fitted spectral density cannot attain the true density. As for the efficiency of the estimation, the Fisher information of the estimation in time series analysis is defined in a limiting form. However, only the case $\alpha = -1$ leads to the efficient result, that is, the asymptotic variance of the estimation is the inverse of fisher information in time series analysis. The result is shown by so called Kholevo's inequality. Even though, the exotic disparity has another robust property for estimation. We can first estimate the parameter by the exotic disparity for some initial value in Newton-Raphson's iteration, and then estimate the true value by non-iterative method based on the disparity of $\alpha = -1$.

References

- [1] Hosoya, Y. and M. Taniguchi (1982) "A central limit theorem for stationary processes and the parameter estimation of linear processes," *The Annals of Statistics*, Vol. 10, pp. 132–153.
- [2] Taniguchi, M. (1979) "On estimation of parameters of Gaussian stationary processes," *Journal of Applied Probability*, Vol. 16, No. 3, pp. 575–591.
- [3] Taniguchi, M. (1981) "An estimation procedure of parameters of a certain spectral density model," *Journal of the Royal Statistical Society. Series B*, Vol. 43, pp. 34–40.
- [4] Taniguchi, M. (1987) "Minimum contrast estimation for spectral densities of stationary processes," *Journal of the Royal Statistical Society. Series B*, Vol. 49, pp. 315–325.
- [5] Whittle, P. (1952) "Some results in time series analysis," *Scandinavian Actuarial Journal*, Vol. 1952, No. 1-2, pp. 48–60.

Formalization of statistical conditional independence relations using Coq/SSReflect

Jinfang Wang, Manabu Hagiwara and Mitsuharu Yamamoto Graduate School of Science, Chiba University

1 Introduction

Several axiomatic systems have been proposed for studying probabilistic conditional independence relations. Among such systems, the most well known ones are the graphoid of Pearl and Paz (1987) and the separoid of Dawid (2001), these axiomatic systems being based on the fundamental properties satisfied by PCI. For instance, symmetry is an intrinsic property for the ternary CIP relation $X \perp\!\!\!\perp Y \mid Z$, that is, $X \perp\!\!\!\perp Y \mid Z$ holds if and only if $Y \perp\!\!\!\perp X \mid Z$. Another axiomatic system called *cain* (Wang, 2010), however, takes a quite different approach. The cain algebra axiomatizes the most fundamental algebraic properties concerning the probability density (mass) functions, In addition to other properties, for instance, the collection of probability density (mass) functions forms an abelian group with respect to the usual product for real numbers. The cain enjoys a great advantage over either the graphoid or the separoid in that all the relation concerning PCI are expressed in equal forms. To demonstrate the possible advantage of this approach, we shall formalize the theories of cain using theorem-prover Coq/SSReflect.

Coq is an interactive theorem prover, which is based on the theory of the calculus of inductive constructions. Coq allows the expression of mathematical assertions and mechanically checks proofs of these assertions. The Curry-Howard isomorphism expresses a direct relationship between computer programs and mathematical proofs. According to this theory, propositions are types and (correct) proofs are programs. SSReflect (small scale reflection) is an extension of coq developed in the path to formalize the proof of the Four Color Theorem (2004), The proof of the Feit Thomson Theorem (Feit and Thompson, 1962, 1963) has also been formalized using SSReflect. The Feit Thomson Theorem states that *every finite group of odd order is solvable*. This can be inputted into SSReflect in the following.

```
Theorem (gT:finGroupType) (G:{group gT}):odd |G| \rightarrow solvable G.
```

The actual proof involved more than 170,000 lines of codes with more than 15,000 definitions and 4,200 theorems. Most recently, the formalization of the Kepler Conjecture has also been completely. SSReflect has also been partially used in this formations.

2 Computer Verified Conditional Independence Relations

The following codes define a bounded lattice L with bottom bot. To formalize the cainoid, we only need to introduce the join operator.

```
Parameter L : eqType.

Parameter bot : L.

Parameters meet join : L-XL-L.

Axiom join_idempotent : idempotent join.

Axiom join_commutative : commutative join.
```

```
Axiom join_associative : associative join. Definition ge x y := join x y = x. Axiom bot_minimum : \forall x, ge x bot.
```

The following first introduce the set of coins, which is of the type called <code>choiceType</code>. The lattice $\mathbb L$ is closed under the binary dot operator dot defined on the direct product of $\mathbb L$. The mixed coin mix is also a binary dot operator dot defined on the direct product of $\mathbb L$.

```
Parameter coins : choiceType.

Parameter dot : coins \rightarrow coins \rightarrow coins.

Parameter mix : L\rightarrowL\rightarrowcoins.
```

The following are special cases of mix, with bob, up x and down x corresponding to $1, \mathbb{T}^x$ and \mathbb{T}_x respectively.

```
Definition bob := mix bot bot.

Definition up x := mix x bot.

Definition down x := mix bot x.
```

Now the axioms of a cainoid C1 through C5 may then be specified as follows.

```
Axiom dotC : commutative dot. 

Axiom dotA : associative dot. 

Axiom bob_unitL : left_id bob dot. 

Axiom up_down_unitL : \forall x, dot (up x) (down x) = bob. 

Axiom mix_up_down : \forall x y, x \neq bot \rightarrow 

mix x y = dot (up (join x y)) (down y).
```

The following give the cain-algebraic analogy of the Bayes' Theorem and the corresponding formalization in SSReflect.

```
THEOREM 2.1 (Bayes' Theorem). If x > \emptyset, y > \emptyset, then \mathbb{T}_y^x = \mathbb{T}_x^y \mathbb{T}^x \mathbb{T}_y
```

```
Theorem Bayes: \forall x y, (x \neq bot) \land (y \neq bot) \rightarrow mix x y = dot (mix y x) (dot (up x) (down y)).
```

The formal proof of this theorem can be obtained by basically applying the axioms of a cain in direct fashion.

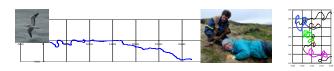
```
Proof.
  move > x y; case > xneb yneb.
  rewrite [in RHS]mix_up_down; last exact.
  by rewrite dotA -![ (dot (dot (up _) _) _)]dotA [dot (_ x) (_ x)]dotC
    up_down_unitL bob_unitL join_commutative mix_up_down. Qed.
```

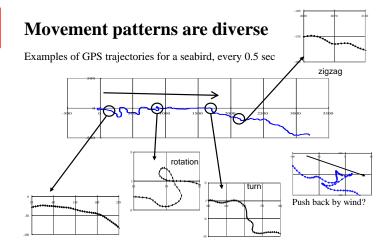
Circular statistics for animal/plant behaviors and movement ecology for circular statistics



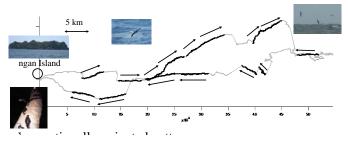
Kenichiro Shimatani

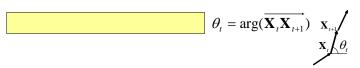
(The Institute of Statistical Mathematics)

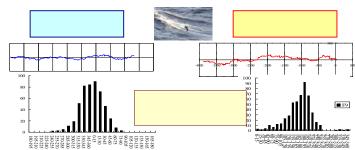




Among various patterns, oriented movements are most commonly seen.







2. Kato's circular auto-regression

$$\theta_t = \arg(\overrightarrow{\mathbf{X}_t \mathbf{X}_{t+1}})$$
 :heading direction

Kato's circular auto-regression (Kato 2010)

 $\theta_t = \alpha + 2\arctan\{w \tan((\theta_{t-1} - \alpha)/2))\} + e_t$



 $Y=wX+e_t$

 e_t ~normal $(0,\sigma^2)$

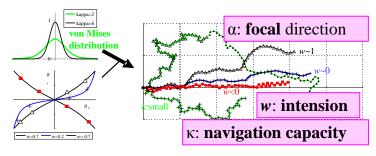
1. Insert link function tan(x/2) for transforming $[-\pi, \pi)$ to **R** tan(Y/2)=w tan(X/2)

2. Change normal distribution to circular distribution.

3.Add reference angle α (unlike linear variables, 0 is not specific on a circle) $Y=\alpha+\beta+2\arctan\{w\tan((X-\alpha)/2)\}+e$ Downs and Mardia (2002)

2. Kato's circular auto-regression $\theta_t = \alpha + 2\arctan\{w \tan((\theta_{t-1} - \alpha)/2))\} + e_t$

Concentration to α might sound unsuitable as regression model, but this property is suitable for modeling oriented movements. Parameters are ecologically interpretable.



	i auto-regression		True	values	Kato's circular auto-regression Random walk			m walk	model by AIC		
	Model	Probability distribution		kappa /r	von Mises	wrapped Cauchy		wrapped Cauchy	w	kappa	alpha
	Kato's circular	von Mises	0.7	5	-33.1	-39.0	-37.2		0.64	4.89	-0.29
	auto-regression	von Mises	0.2	2	-50.2	-51.1	-64.4		0.06	2.77	-0.16
		von Mises	-0.7	5	-35.6	-42.5	-82.7		-0.72	4.47	0.03
	Random walk		1	8	-18.5	-25.5			1	7.88	
	Kato's circular	wrapped Cauchy	0.7	0.9	-28.2	-19.5		-32.2	0.65	0.87	0.04
	auto-regression	wrapped Cauchy	0	0.6	-54.4	-53.1		-64.1	0.35	0.71	-0.09
15 г	Random walk	wrapped Cauchy	-0.7 1	0.9	-33.9	-9.9		-74.1 -7.9	-0.71	0.91	0.01
13	Kandom walk	wrapped Cauchy	1	0.9	-14.2	-7.3	-	-/.9	1	0.90	
5	N vM		****			\ \ \ \ \ \ \ \ \ \ \ \ \ \ \ \ \ \ \	:				
→ vM (0.7, 5)	→ vM (0.2,	(2)	X.	*							
→ vM (-0.7, 5)	→ vM (1, 8)		-	x - ~4	•••					
→ wC (0.7, 0.9)	→ wC (0, 0	0.6)									
→ wC (-0.7, 0.9)	wC (1, 0	1.9) 20		30		40					

r auto-regression

Maximum log-likelihood

2.5 What statistical models are desirable?

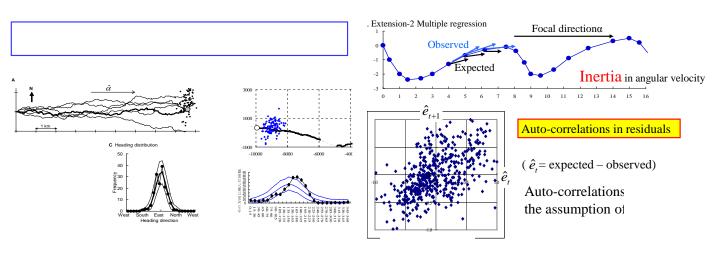
What is more important for the basic model is,

"3. a broad range of movement patterns seems to be explained by this model but **actually**, **not**".

The basic model should be rejected as a **null model**.

Goodness-of-fit (summary statistics) are crucial for understanding what aspects cannot be explained by the null model?

Goodness-of-fit for the null model should clearly show insufficient fitting aspects, then, we may extend the null model to more applicable models.



4. Concluding remark

The movement model using Kato's circular auto-regression satisfies requirements for a **null model**.

- 1. have an **ecological basis** and the parameters have **ecological interpretations**.
- 2. mathematically/statistically tractable.
- 3. Can explain a **broad range** of movement patterns, and can be flexibly **extended**.

Rejection of a null model suggests effects of some (often, more) essential factor(s).

If a null model is inappropriate, its rejection gives us nothing (or wrong conclusion).

"Movement ecology will advance in parallel to developments in circular statistics, and the development of circular statistics will be promoted by the practical demands made from movement ecology."

(Shimatani, K. Yoda, K., Katsumata, N., and Sato, K. "Toward the Quantification of a Conceptual Framework for Movement Ecology Using Circular Statistical Modeling." PLoS One (2012))

Circular statistics for animal/plant behaviors and movement ecology for circular statistics

Aakala, T., Shimatani, IK, Abe, T., Kubota, Y., Kuuluvainen, T. (2014) Tree crown plasticity to solar radiation and neighborhood structure at high latitudes. (under review)



On-going collaborated work with Dr. K. Yoda, Dr. K. Sato, Mr. Y. Goto, Dr. T. Noda, Dr. S. Kato, Dr. T. Abe, Dr. T. Aakala, Dr. T. Kuuluvainen, Dr. Y. Kubota, and more.

2段階サンプリングに基づく研究デザインと解析方法

Study design and methods based on the two stage sampling

和泉 志津恵, Shizue Izumi

大分大学工学部, Faculty of Engineering, Oita University

2 段階サンプリングに基づく臨床研究のデザインとして、2 段階ケース・コントロール研究や2段階ケース・コホート研究が、頻繁に行われる.本研究では、これらの2つのデザインの特徴と解析方法を解説し、臨床研究の計画を立てる際に必要となる標本数や統計的な検出力を求める方法についての考察を行う.

2段階ケース・コントロール研究は、ターゲットとなる母集団から2段階のサンプリン グを行うデザインに基づく.このデザインの解析では、母集団から1段階目にサンプリン グされたコホートにおける疾病の有無 Y と共変量 X のデータと, コホートから 2 段階目に サンプリングされたケースとコントロールにおける共変量 Z のデータを用いる. その際, コホートデータを, 共変量 Z が Missing At Random(MAR)で欠測した不完全データとみなし て, 重みつき尤度法(WL)や, Approximate Bayesian Bootstrap(ABB), Markov chain Monte Carlo(MCMC), Multiple Imputation by Chained Equations (MICE)のような多重代入法によっ て解析することができる(野間 他, 2012; Keogh et al., 2013). Breslow and Chatterjee (1999) は, National Wilms Tumour Studies のデータをもとにして, 2 段階ケース・コントロール研 究の事例紹介を行っている. そこで、本研究では、2 段階ケース・コントロール研究にお ける上記の解析方法および、2段階目の対象者のみを用いたロジスティック回帰分析法を 用いて検出力を算出して、研究デザインや解析方法の違いが検出力に及ぼす影響について 調べた. 加えて, 2 段階目にサンプリングされる対象者数を固定して, ケース数とコント ロール数の比による影響についても調べた. 仮想データは, National Wilms Tumour Studies のデータを参考にして作成した. コホートは母集団からのランダムサンプリングとして, 標準正規分布に従う共変量 X と, 2 値変数である共変量 Z を用いたロジスティック回帰モ デルから症例の確率を定めた. 2 段回目のサンプリングでは、共変量 X のカテゴリ変数と 共変量Zで分けられる層を用いた.共変量Zが1となる確率は,共変量Xの関数とした. 数値実験の結果と考察については、当日に報告する.

一方、2 段階ケース・コホート研究は、コホート全体からのランダムサンプリングによって得られたサブコホートと症例(ケース)全員を用いる研究デザインである。複数の疾病についてケース・コホート研究を計画する場合、それぞれの疾病のタイプごとにサブコホートを設定する必要がないため、研究コストの削減および生体試料の有効活用につながる(Lin, 2013)。2 段階ケース・コントロール研究のときと同様に、1 段階目にサンプリングされたコホートにおいて、疾病の有無 Y、観察時間 T、共変量 X のデータを得る。次に、2 段階目にサンプリングされたサブコホートとケースにおいて共変量 Z のデータを得る。その際、コホートデータを、比例ハザードモデルのもとで、共変量 Z が Missing At

Random(MAR)で欠測した不完全データとみなして,推定された重みを用いる方法(Estimated weights),Multiple Imputation by Chained Equations (MICE)のような多重代入法によって解析することができる(野間 他, 2012; Keogh et al., 2013).Keogh et al. (2013)は,食物繊維の摂取と大腸がんの関係を調べた Nested case-control study の事例紹介や,そのデータを参考にして作成した仮想データを用いて,Nested case-control study とケース・コホート研究における推定精度に関する実験結果を示した.そこで,本研究では,2 段階ケース・コホート研究における上記の解析方法および,2 段階目のサンプリング確率を重みとした通常のケース・コホート解析方法を用いてサブコホートの標本サイズを算出し,研究デザインや解析方法の違いによる標本サイズへの影響について調べた.仮想データは,Keogh et al. (2013)の数値実験データを参考にして作成した.Estimated weights と多重代入法については,共変量 X と相関のある変数である代替変数 X を用いた.数値実験の結果と考察については,当日に報告する.

参考文献

- [1] Breslow, N.E. and Chatterjee, N. (1999). Design and analysis of two-phase studies with binary outcome applied to Wilms tumour prognosis. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, **48**, 457–468.
- [2] 野間久史, 田中司郎. (2012). Multiple Imputation 法による 2 段階ケースコントロール研 究の解析. *応用統計学*, **41(1)**, 1–17.
- [3] 野間久史, 田中司郎, 田中佐智子, 和泉志津恵. (2012). Multiple Imputation法によるネステッドケースコントロール研究, ケースコホート研究の解析. *計量生物学*, **33(2)**, 101–124.
- [4] Keogh, R.H. and White, I.R. (2013). Using full-cohort data in nested case-control and case-cohort studies by multiple imputation. *Statistics in Medicine*, **32(23)**, 4021-4043.
- [5] Lin, D.Y. (2013). Survival analysis with incomplete genetic data. *Lifetime Data Analysis*, **20(1)**, 16-22.
- [6] 野間久史, 田中司郎, 田中佐智子, 和泉志津恵. (2012). Multiple Imputation 法による ネステッドケースコントロール研究, ケースコホート研究の解析. 計量生物学, 33(2), 101-124.

連絡先:和泉 志津恵, Email: shizue@oita-u.ac.jp 〒870-1192 大分県大分市大字旦野原 700 番地 大分大学工学部

LSE-type estimation for stochastic processes with small Lévy noise

Yasutaka Shimizu
Department of Applied Mathematics, Waseda University

1 Model

Let $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, \mathbb{P})$ be a stochastic basis, on which an \mathbb{R} -valued stochastic process X is defined via the stochastic integral equation

$$X_t^{\epsilon} = x + \int_0^t b(X_s^{\epsilon}, \theta_0) \, \mathrm{d}s + \epsilon \cdot Q_t, \tag{1}$$

where $x \in \mathbb{R}, \epsilon > 0$, and θ_0 is an unknown parameter that belongs to Θ_0 , which is an open bounded, convex subset of \mathbb{R}^p ; we put $\Theta := \overline{\Theta_0}$, the closure of Θ_0 ; b is a measurable function on $\mathbb{R} \times \Theta$; Q is a Lévy process such that the *characteristic exponent*: $\psi(u) := \mathbb{E}\left[\exp(iu^{\top}Q_1)\right]$, is given by

$$\psi(u) = ib^{\top}u - \frac{\sigma^2}{2}u^{\top}u + \int_{\mathbb{R}} \left(e^{iu^{\top}z} - 1 - \frac{iu^{\top}z}{1 + |z|^2}\right) \nu(\mathrm{d}z), \quad u \in \mathbb{R},$$
 (2)

where $b \in \mathbb{R}$, ν is the Lévy measure with $\nu(\{0\}) = 0$ and $\int_{|z| \le 1} |z|^2 \nu(\mathrm{d}z) < \infty$.

We suppose that the process $X^{\epsilon} = (X_t^{\epsilon})_{t \in [0,1]}$ is observed discretely in time: $\{X_{t_k^n}\}_{k=0}^n$ with $t_k^n = k/n$ (the index ϵ is abbreviated), from [0,1]-interval. We denote by $\Delta_k^n X := X_{t_{k-1}^n} - X_{t_{k-1}^n}$ and $\Delta_n := t_k^n - t_{k-1}^n = 1/n$.

Our interest is to estimate the value of the parameter θ_0 from the discrete samples under that $n \to \infty$ as well as $\epsilon \to 0$: small noise asymptotics.

2 LSE-type estimator

We consider the threshold-type least squares estimator defined as follows:

$$\widehat{\theta}_{n,\epsilon} := \arg\min_{\theta \in \Theta} \Phi_{n,\epsilon}(\theta), \tag{3}$$

where

$$\Phi_{n,\epsilon}(\theta) = \sum_{k=1}^{n} |\Delta_k^n X - b(X_{t_{k-1}^n}, \theta) \cdot \Delta_n|^2 \mathbf{1}_{\{|\Delta_k^n X| \le \delta_{n,\epsilon}\}}.$$
 (4)

Without the indicator $\mathbf{1}_{\{|\Delta_k^n X| \leq \delta_{n,\epsilon}\}}$, the estimating function is the same as in Long et al. [1], which can give unstable performance in finite sample simulations. For the stability, we use the threshold $\delta_{n,\epsilon} > 0$ in order to eliminate 'large' shocks causing bias to drift estimation.

3 Main results

We make the following assumptions:

A1 $|b(x,\theta) - b(y,\theta)| \lesssim |x-y|$ for each $x,y \in \mathbb{R}$ and $\theta \in \Theta$.

Under this assumption, the ordinary differential equation

$$dX_t^0 = b(X_t^0, \theta_0) dt, \quad X_0^0 = x,$$

has the unique solution $X^0 = (X_t^0)_{t \ge 0}$.

A2 $b \in C^{2,3}_{\uparrow}(\mathbb{R} \times \Theta; \mathbb{R}).$

A3 $\theta \neq \theta' \Leftrightarrow b(X_t^0, \theta) \neq b(X_t^0, \theta_0)$ for at least one value of $t \in [0, 1]$.

A4 $I(\theta_0) := \int_0^1 \nabla_{\theta} b(X_t^0, \theta_0) \nabla_{\theta}^{\top} b(X_t^0, \theta_0) dt \ (\in \mathbb{R}^p \otimes \mathbb{R}^p)$ is positive definite,

Q1[γ] There exists some $\gamma > 0$ such that, for any k = 1, ..., n,

$$\mathbb{P}\left\{\sup_{t\in(0,\Delta_n]}|Q_t|>\Delta_n^{\gamma}\right\}=o_p(1).$$

 $\mathbf{Q2}[q]$ For some q > 0,

$$\int_{|z|>1} |z|^q \, \nu(\mathrm{d}z) < \infty.$$

Theorem 3.1 (Consistency) Suppose A1-A3, Q1[γ], and that a sequence $\{\delta_{n,\epsilon}\}$ satisfies that

$$\delta_{n,\epsilon} \Delta_n^{-1} \to \infty, \quad \epsilon \Delta_n^{\gamma} \delta_{n,\epsilon}^{-1} = O(1).$$
 (5)

Then

$$\widehat{\theta}_{n,\epsilon} \stackrel{\mathbb{P}}{\longrightarrow} \theta_0.$$

Theorem 3.2 (Asymptotic distribution) Suppose the same assumptions as in Theorem 3.1, and further A4. Then

$$\epsilon^{-1}(\widehat{\theta}_{n,\epsilon} - \theta_0) \stackrel{\mathbb{P}}{\longrightarrow} \zeta,$$

where $\zeta := I^{-1}(\theta_0) \int_0^1 \nabla_{\theta} b(X_t^0, \theta_0) \, \mathrm{d}Q_t \in \mathbb{R}^p$.

Theorem 3.3 (Mighty convergence) Suppose the same assumptions as in Theorem 3.2, and that Q2[M] holds true for any M > 0. Then, it follows for every continuous function $f: \mathbb{R}^p \to \mathbb{R}$, of polynomial growth that

$$\mathbb{E}\left[f\left(\epsilon^{-1}(\widehat{\theta}_{n,\epsilon}-\theta_0)\right)\right]\to\mathbb{E}[f(\zeta)],$$

where ζ is given in Theorem 3.2.

References

[1] Long, H; Shimizu, Y. and Sun, W. (2013). Least squares estimators for discretely observed stochastic processes driven by small Lévy noises. *J. Multivariate Analysis*, **116**, 422–439.

ハザード関数のカーネル型推定量の漸近表現とその応用

九州大学大学院数理学研究院 前園宜彦

1 序

 X_1,\cdots,X_n を分布関数 F(x) からの無作為標本とする.データは生存時間を表すとすると,点 x_0 でのハザード関数は

$$\frac{f(x_0)}{1 - F(x_0)}$$

と定義される.ここで f(x) は F(x) の密度関数である.このハザード関数の推定量としては密度関数と分布関数を様々な方法で推定し代入したものが提案されている.本講演ではカーネル型推定量を利用した下記の推定量を考察する.K(u) をカーネルとするとき

$$\widehat{f}_n(x_0) = \frac{1}{n} \sum_{i=1}^n \frac{1}{h_n} K\left(\frac{x_0 - X_i}{h_n}\right),$$

$$\widehat{F}_n(x_0) = \frac{1}{n} \sum_{i=1}^n W\left(\frac{x_0 - X_i}{h_n}\right)$$

がそれぞれの密度及び分布関数の推定量である.ただし

$$W(t) = \int_{-\infty}^{t} K(u) du$$

であり, h_n はバンド幅で $h_n \to 0$ とする. ハザード関数のカーネル型推定量は

$$\frac{\widehat{f}_n(x_0)}{1 - \widehat{F}_n(x_0)}$$

で与えられる.この推定量は分母の収束のオーダーは $n^{1/2}$ であるが,分子のオーダーは $(nh_n)^{1/2}$ である.したがって平均二乗誤差は分子の推定量の平均二乗誤差が支配的となる.ここではカーネル関数 K(u) は 2 次のカーネル,すなわち

$$\int_{-\infty}^{\infty} K(u)du = 1, \qquad \int_{-\infty}^{\infty} K(u)u^2 du \neq 0$$

を満たすとする.また簡単のためにバンド幅は $h_n=cn^{-1/4}\;(c>0)$ とする.このとき漸近正規性が成立する.以下では分布の近似の精密化を求める.

2 カーネル型推定量の漸近表現

 \widehat{f}_n の分散を求めると

$$V\left[\widehat{f}_n(x_0)\right] = \frac{f(x_0)}{nh_n} \int K^2(z)dz + \cdots$$

となり, $\widehat{F}(x_0)$ の分散は

$$V\left[\hat{F}_{n}(x_{0})\right] = \frac{F(x_{0})(1 - F(x_{0}))}{n} + \cdots$$

であるから,標準化するときの収束のオーダーは $(nh_n)^{1/2}$ になる.このオーダーで漸近分布を持つように標準化するので,高次漸近理論を構成するには,カーネル関数と密度関数に対する適当な条件の下で,下記のような漸近表現が必要になる.

定理 適当な正則条件の下で

$$\frac{\widehat{f}_{n}(x_{0})}{1-\widehat{F}_{n}(x_{0})} - \frac{f(x_{0})}{1-F(x_{0})}$$

$$= \frac{b_{1,n}}{1-F(x_{0})} + \frac{f(x_{0})b_{2,n}}{\{1-F(x_{0})\}^{2}} + \frac{1}{n}\sum_{i=1}^{n} \left[\frac{f(x_{0})}{\{1-F(x_{0})\}^{2}} Z_{i,n} + \frac{1}{1-F(x_{0})} Y_{i,n} \right] + \frac{1}{n^{2}}\sum_{1 \leq i < j \leq n} \frac{1}{\{1-F(x_{0})\}^{2}} (Y_{i,n} Z_{j,n} + Y_{j,n} Z_{i,n}) + (nh_{n})^{-1/2} R_{n}$$

が成り立つ.ただし

$$Y_{i,n} = \frac{1}{h_n} \left[K \left(\frac{x_0 - X_i}{h_n} \right) - E \left\{ K \left(\frac{x_0 - X_i}{h_n} \right) \right\} \right],$$

$$Z_{i,n} = W \left(\frac{x_0 - X_i}{h_n} \right) - E \left\{ W \left(\frac{x_0 - X_i}{h_n} \right) \right\},$$

$$b_{1,n} = \frac{h_n^2}{2} f''(x_0) \int z^2 K(z) dz,$$

$$b_{2,n} = \frac{h_n^2}{2} f'(x_0) \int z^2 K(z) dz$$

である.ここで残差項は $E|R_n|^c=O(n^{-1/2-c/2-\delta})$ となる $c\geq 2,\ \delta>0$ が存在するものになる.

この漸近表現を利用すると , 残差項 $o(n^{-1/2})$ までのエッジワース展開を求めることができる .

参考文献

- [1] Estévez-Pérez, G, Quitela-del-Río, P and Vieu, P. (2002), Jour. Stat. Plan. Inf., Vol.104, 1-30 .
- [2] Patil, P. N. (1993), Jour. Stat. Plan. Inf., Vol.35, 15-30.

Adaptive Quantile Regression and Risk Bounds ¹

Ngai Hang Chan
Department of Statistics
Chinese University of Hong Kong
Shatin, NT, Hong Kong
nhchan@sta.cuhk.edu.hk

Abstract

An adaptive local smoothing method for nonparametric conditional quantile regression models is considered in this paper. Theoretical properties of the procedure are examined. The proposed method is fully adaptive in the sense that no prior information about the structure of the model is assumed. The fully adaptive feature not only allows varying bandwidths to accommodate jumps or instantaneous slope changes, but also allows the algorithm to be spatially adaptive. Under general conditions, precise risk bounds for homogeneous and heterogeneous cases of the underlying conditional quantile curves are established. An automatic selection algorithm for locally adaptive bandwidths is also given, which is applicable to higher dimensional cases. Simulation studies and data analysis confirm that the proposed methodology works well.

¹International Symposium on "Education, Theory and Applications in Statistical Science", October 24–26, 2014, Nigata University, Nigata. Research supported in part by grants from HKSAR-RGC-GRF.

Bayesian Inference for Time Series of Counts Models

Cathy W. S. Chen Department of Statistics, Feng Chia University, Taiwan

This paper proposes a new variant of generalized autoregressive conditional Poisson (ACP) models for a time series of counts. The proposed models can account for several characteristics, such as discreteness, over-dispersion, asymmetry, serial correlation, and structural changes. This study advances the ACP model in the following three ways: first, a threshold nonlinear ACP model is designed to capture more flexible asymmetric and nonlinear properties; second, an ACP model with generalized Poisson error distribution is proposed to accommodate the over-dispersion and under-dispersion of a time series of counts; and third, an ACP model with change points is proposed to cope with situations where the data set is experiencing changes in its pattern. An adaptive Markov chain Monte Carlo (MCMC) sampling scheme is developed for parameter estimation. The standardized Pearson residuals are employed for diagnostic checking. The proposed methods are illustrated by a simulation study and empirical analysis. The latter provides evidence for a significant improvement in the modelling of New South Wales crime data sets.

Markov chain Monte Carlo method for high-dimensional heavy-tailed distribution

Kengo Kamatani*

Osaka University kamatani@sigmath.es.osaka-u.ac.jp.

Abstract: High-dimensional asymptotics of the random walk Metropolis-Hastings (RWM) algorithm is well understood for a class of light-tailed target distributions. We develop a study for heavy-tailed target distributions, such as the Student t-distribution or the stable distribution. The performance of the RWM algorithms heavily depends on the tail property of the target distribution. The expected squared jumping distance (ESJD) is a common measure of efficiency for light-tail case but it does not work for heavy-tail case since the ESJD is unbounded. For this reason, we use the rate of weak consistency as a measure of efficiency. When the number of dimension is d, we show that the rate for the RWM algorithm is d^2 for the heavy-tail case where it is d for the light-tail case. Also, we show that the Gaussian RWM algorithm attains the optimal rate among all random RWM algorithms. Thus no heavy-tail proposal distribution can improve the rate. Finally we describe an efficient Metropolis-Hastings algorithm, which attains a better rate of consistency. This talk is based on Kamatani [2014].

References

Kengo Kamatani. Rate optimality of Random walk Metropolis algorithm in high-dimension with heavy-tailed target distribution. *Arxiv*, 2014. URL http://arxiv.org/abs/1406.5392.

ケースクロスオーバーデザインでアウトカムが再発イベントの場合の解析 慶應義塾大学医学部 衛生学公衆衛生学 竹内文乃 概要

近年、特に大気汚染を曝露として、その急性影響の評価を目指す環境疫学領域において、ケース・クロスオーバー(CCO)デザインによる解析を行う事例が増加している。CCO デザイン提案当初は、1回しか起こらないイベントを想定していたが、さまざまな疾患に適用されるにつれて、頻繁に繰り返すイベントをどのように評価するか、コントロール期間の取り方や個人内相関の考慮が課題となるようになってきた。本研究では再発イベントを想定した CCO デザインに対して複数の解析手法をレビューし、大気汚染(黄砂飛来有無)と喘息発作の有無という単純な状況でシミュレーションを実施した。

1. ケース・クロスオーバーデザインとは

ケース・クロスオーバーデザインは、疫学研究の研究デザインのひとつであり、対象とする疾患の発症者(ケース)のみを研究対象者とし、ケースが経時的に経験する曝露と非曝露を比較することで曝露の疾患発症リスクを評価する[Maclure 1991]。CCO は曝露が急性健康影響を持つ場合に利用され、その利点は、同一個人内で曝露 - 非曝露の比較ができる点だが、交絡因子が短時間で変化する場合には取り除けるとは限らない[Greenland 1996]。解析には、条件付ロジスティック回帰モデルを用いるのが一般的である[Maclure 1991]。

2. 再発を考慮した解析

i=1,...,n:対象者。 $t∈\{\tau_1,...,\tau_T\}$:研究期間。Tは喘息日誌をつけた平日の登校日。

 $j: 個人ごとのイベントを起こした回数。<math>j=1,...,m_i (\ge 1)$

t_{ii}:ケース期間。イベントをいつ起こしたか。s_{ii}:コントロール期間。

 $X_{1ii}: ケース期間(t_{ii})のiごとの共変量。<math>X_{0ii}: コントロール期間(s_{ii})のiごとの共変量。$

 O_{ij} : 対象者iのj番目のイベントで作成される 1:1 のマッチドセット。 O_{ij} ={ t_{ij} , X_{1ij} , s_{ij} , X_{0ij} } 対象者i時点tでの対数オッズは X_{it} を曝露を表し、 λ_i をベースライン共変量として以下。

$$\log \frac{p_{it}}{1 - p_{it}} = \lambda_i + \beta X_{it}$$

2-1. 複数回の疾患発症を1回に限定した解析

複数回の疾患発症を1回に限定し1人1マッチドセット O_{ij} を得る[Luo 2008] 以下の曝露の条件付尤度のスコア関数を解いてベータの条件付最尤推定量を求める。

$$S_{1}(\beta) = \sum_{i=1}^{n} \left\{ X_{1i1} - \frac{\sum_{l=0}^{1} X_{li1} \exp(\beta X_{li1})}{\sum_{l=0}^{1} \exp(\beta X_{li1})} \right\}$$

2-2. 複数回の疾患発症をプールする解析

対象者全員分のマッチドセット $\{O_{ij}; i=1,...,n, j=1,...,m_i\}$ を作り、条件付きで独立に同一の分布 $\{iid\}$ をすると考えられる。この場合、条件付尤度のスコア関数は以下のように変わる。

$$\sum_{i=1}^{n} \sum_{j=1}^{m_i} \left\{ X_{1ij} - \frac{\sum_{l=0}^{1} X_{lij} \exp(\beta X_{lij})}{\sum_{l=0}^{1} \exp(\beta X_{lij})} \right\}$$

2-3. クラスターデータとしての解析

個人をクラスターとみなし、クラスターレベルに条件付ロジスティック回帰を適用する。 ケース期間の共変量の値は $\{X_{1ij}=x_{1ij};\ j=1,...,m_i\}$ で、 S_i を E_i の中のサイズ m_i のサブセット、

 \sum_{s_i} をサブセットの合計とすると、コア関数は以下のようになる[Navidi 1998]。

$$\sum_{i=1}^{n} \left[\sum_{j=1}^{m_i} X_{1ij} - \frac{\sum_{S_i} \{ (\sum_{l \in S_i} X_l) \exp(\beta \sum_{l \in S_i} X_l) \})}{\sum_{S_i} \exp(\beta \sum_{l \in S_i} X_l)} \right]$$

2-4. 対象者内ペアリサンプリング法による解析

クラスターからランダムな復元抽出を Q 回繰り返す (Within-Cluster Resampling: WCR) ことで得られるサブサンプルを、独立なデータとして、条件付ロジスティックで解析する 手法 [Hoffman 2001] [Reiger 2002]。最終的なパラメータ推定値を以下のように得る。

$$\hat{\beta}_{WCR} = \frac{1}{Q} \sum_{q=1}^{Q} \hat{\beta}(q), \quad \hat{\Sigma}_{WCR} = \left\{ \frac{1}{Q} \sum_{q=1}^{Q} \hat{\Sigma}(q) - \frac{1}{Q} \sum_{q=1}^{Q} \left(\hat{\beta}(q) - \hat{\beta}_{WCR} \right) \left(\hat{\beta}(q) - \hat{\beta}_{WCR} \right) \right\}$$

2-5. 重みつき推定方程式による解析

上記リサンプリング法を拡張する形で以下のスコア方程式の解として $\hat{\beta}(q)$ を得る、重みつき推定方程式 (Weighted Estimating Equation; WEE) が提案された[Williamson 2003]。

$$S_{WEE}(\beta) = \sum_{i=1}^{n} \frac{1}{m_i} \sum_{j=1}^{m_i} \left\{ X_{1ij} - \frac{\sum_{l=0}^{1} X_{lij} \exp(\beta X_{lij})}{\sum_{l=0}^{1} \exp(\beta X_{lij})} \right\}$$

3. シミュレーションおよび結果・考察

200 人の小学生が $1\sim15$ 回の喘息発作が発症し(平均は 2 回もしくは 2.5 回)、曝露は黄砂飛来(20 日観測)、 λ_i はlogged beta分布(a=1, b=100)に X_{it} は確率 1/5 のベルヌーイ分布に従うとして、データはロジットモデルlogit(p_{it})= λ_i + β_i X_{it} から 1000 回発生させた。

結果、推定値のバラツキはイベントを 1 回に限定した解析で最大でクラスターとした場合が最小、次にプールした場合となった。クラスターとする場合は個人の発作回数が増えると計算に長い時間がかかった[Luo 2007]。

Hoffman EB, Sen PK, Weinberg CR. Within-cluster resampling. Biometrika. 2001;88:1121-34.

Luo X, Sorock GS. Analysis of recurrent event data under the case-crossover design with applications to elderly falls. Stat Med. 2008;27(15):2890-901.

Maclure M. The case-crossover design: a method for studying transient effects on the risk of acute events. Am J Epidemiol. 1991;133:144-53.

Navidi W. Bidirectional case-crossover design for exposures with time trends. Biometrics. 1998;54:596-605.

Reiger RH, Weinberg CR. Analysis of clustered binary outcomes using within-cluster paired resampling. Biometrics 2002;58:332-41.

Williamson JM, Datta S, Satten GA. Marginal analysis of clustered data when cluster size is informative. Biometrics. 2003;59:36-42.

全国ウツタイン統計を用いた院外心停止例の類型化

○湯澤あや1、山内慶太2、渡辺美智子2

1東芝メディカルシステムズ株式会社、2慶應義塾大学大学院健康マネジメント研究科

本研究は、総務省消防庁のウツタイン統計を活用し、心原性の院外心停止症例のうちバイスタンダーによる蘇生が実施された例について、類型化を試みた。これは、一般市民が救急蘇生に関わる際のバイスタンダーCPRにおける現状把握を主目的とする。今回、母集団を構成する個人の異質性を考慮したうえで、対象データの母集団を4つのクラスに類型化することができた。なお、今回は特にAEDの実施について注目をし、類型化後のそれぞれのクラスにおける実施状況および生存率への影響についても考察をした。さらに、2005年から2011年の経年結果では、すべての年において類似したモデルで説明することが可能であり、2011年の類型パターンに一般性があることが示唆された。なお、全国の県別傾向も調査した。

1. はじめに

わが国における心疾患による死亡は増加傾向にあり,平成23年度人口動態統計によると男女ともに死因の第2位を占めている.心疾患による死亡は突然死の形をとりやすく,その多くが院外における心停止である.我が国においては年間約7万人の院外心停止が発生しており¹⁾病院前救急医療体制の整備と院外心停止例の救命率向上は公衆衛生上のもっとも重要な医療施策のひとつであるとされている.

2. 研究方法

本報告では、総務省消防庁より借用した、全国の院外心肺停止患者データを使用している。 今回はその中から、目撃のある心原性心停止例かつ、バイスタンダーCPR が特定できたケースに限定してデータを使用し、潜在クラス分析を実施した。潜在クラス分析のクラス数特定の使用ソフトウェアは株式会社エスミ Excel アドイン潜在クラス分析 Ver1.06、また記述統計などその他の統計処理はSPSS ver.21である。潜在クラスによって説明される反応変数と潜在クラスに対する説明変数として導入した属性項目に分けて示す。

<潜在クラス分析の反応変数>

バイスタンダー種別(家族/友人/同僚/通行人/その他), 心臓マッサージの有無,AED 実施の有無,口頭指示の有無, 初期心電図波形:VF/PVT/PEA/心静止/その他

3. 結果

クラス数については解釈可能性にも配慮して判断した. 今回の分析においては、潜在クラス分析を繰り返し行った最終的にはその他の要因も考慮の上、クラス数を 4 とした場合が最も適切な類型を導くことが可能と考え、C=4 として潜在クラス分析を適用した.

表 1: 4 クラスモデルの結果(応答確率)

	クラス1	クラス2	クラス3	クラス4
クラスサイズ	0.420	0.275	0.199	0.107
パイスタンダー種別				
家族	0.881	0.101	0.600	0.116
友人	0.015	0.008	0.114	0.106
同僚	0.000	0.001	0.125	0.163
通行人	0.008	0.016	0.060	0.180
その他	0.096	0.875	0.101	0.435
心臓マッサージ				
なし	0.003	0.018	0.012	0.027
あり	0.997	0.982	0.988	0.973
人工呼吸				
なし	0.881	0.588	0.797	0.519
あり	0.119	0.412	0.203	0.481
バイスタンダーAED				
なし	1.000	0.954	0.995	0.526
あり	0.000	0.046	0.005	0.474
口頭指示				
なし	0.150	0.619	0.297	0.695
あり	0.850	0.381	0.703	0.305
波形種別				
VF	0.095	0.046	0.656	0.344
PVT	0.004	0.001	0.004	0.007
PEA	0.319	0.332	0.143	0.170
心静止	0.554	0.553	0.196	0.176
その他	0.028	0.068	0.001	0.303

4. 考察

4 つの異なるグループの存在が明らかになった.グループの特色は年齢分布に制限を受けるのではなく,年齢分布をある程度加味した上で,そこに所属する確率が高い人の生活行動パターンも反映している.また,潜在クラスを特定した後に分析した生存率のクラス比較では,クラス2でAED実施ありとなしの群比較において,最も高いオッズ比を示した.このことから,クラス2にAED実施をより実施しやすくなるような介入をすることで生存率の向上に期待ができるものである.

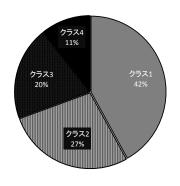


図 1: クラスサイズ

2007 年以降で搬送患者特性に類似的傾向が見られ,我が国におけるバイスタンダーCPR の傾向および,公共施設への AED 配備に一定の効果があることが確認できた.また,地域的な特長を県別に見てみると,東京については突出してクラス 4 の割合が高くなっているが,その他の多くの地域では,むしろクラス 4 の構成比は低くなっている.必ずしも公共施設への設置だけが解決策でないことを伺わせる.

X1.122									
AED使用	クラス1	クラス2	クラス3	クラス4					
なし	6.388%	6.738%	25.839%	30.837%					
あり	9.966%	13.612%	36.382%	53.683%					
オッズ比	1.622	2.181	1.641	2.6					
95%CI	6 388-4 565	0.828-5.745	0 896-3 008	1 458-4 636					

表 2: AED 使用の有無における生存率の比較

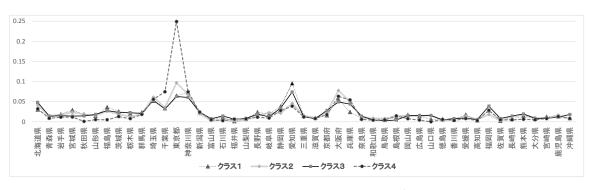


図 2: 説明変数のカテゴリ別帰属確率(都道府県別)

5. 結論

一般市民による救急蘇生の実施は、プレホスピタルケアを充実させるうえで重要である. AED の使用についてはその有用性が検証されているが、AED を市民一人ひとりがいざという時に使用できるようにするためには、配備の問題だけではなく、十分に使用できていない状況を把握する必要がある. その中で4つのグループにおける AED 使用現状を今回整理することができた.

Decompositions of symmetry using models based on f-divergence for square contingency tables

Yusuke Saigusa (Tokyo University of Science) Kouji Tahata (Tokyo University of Science) Sadao Tomizawa (Tokyo University of Science)

1. はじめに

行と列が順序のある同じ分類からなる $r \times r$ 正方分割表を考え,(i,j)セル確率を p_{ij} とする $(i=1,\ldots,r;j=1,\ldots,r)$. このとき,対称(S) モデルは次のように定義される $(Bowker\ [2])$:

$$p_{ij} = \psi_{ij}$$
 $(i = 1, \dots, r; j = 1, \dots, r),$ $triangle \psi_{ij} = \psi_{ji}.$

Sモデルが成り立たないとき、いくつかの拡張したモデルが提案されている。たとえば、準対称 (QS) モデルは次のように定義される (Caussinus [3]):

$$p_{ij} = \alpha_i \beta_j \psi_{ij}$$
 $(i = 1, \dots, r; j = 1, \dots, r),$ $trial_i = \psi_{ij} = \psi_{ji}.$

QSモデルはオッズ比 $\theta_{(i < j: s < t)} = (p_{is}p_{jt})/(p_{js}p_{it})$ を用いて次のようにも表せる:

$$\theta_{(i < j; s < t)} = \theta_{(s < t; i < j)}$$
 $(1 \le i < j \le r; 1 \le s < t \le r).$

また、周辺同等 (MH) モデルは次のように定義される (Stuart [6]):

$$p_{i.} = p_{.i}$$
 $(i = 1, ..., r),$ $ttilde{tilde} p_{i.} = \sum_{t=1}^{r} p_{it},$ $p_{.i} = \sum_{s=1}^{r} p_{si}.$

Caussinus [3] は次の定理を与えた:

定理1. Sモデルが成り立つための必要十分条件は、QSモデルとMHモデルの両方が成り立つことである.

 $\{u_s\}$ を順序のある既知のスコアとする $(s=1,\ldots,r;u_1\leq u_2\leq \cdots \leq u_r;u_1< u_r)$. このとき,順序準対称 (OQS) モデルは次のように定義される (Agresti [1], p.429):

$$p_{ij} = \alpha^{u_i} \beta^{u_j} \psi_{ij}$$
 $(i = 1, \dots, r; j = 1, \dots, r),$ $\uparrow c \not \sim c \lor \psi_{ij} = \psi_{ji}.$

fを2階微分可能な狭義凸関数, Fをfの1次導関数とする. f-ダイバージェンスに基づく準対称 (QS[f]) モデルは次のように定義される (Kateri and Papaioannou [5]):

$$p_{ij} = p_{ij}^S F^{-1}(\alpha_i + \gamma_{ij})$$
 $(i = 1, ..., r; j = 1, ..., r),$

ただし,

$$p_{ij}^S = \frac{p_{ij} + p_{ji}}{2}, \quad \gamma_{ij} = \gamma_{ji}.$$

Kateri and Papaioannou [5] は次の定理を示した:

定理 2. Sモデルが成り立つための必要十分条件は、QS[f]モデルと MHモデルの両方が成り立つことである.

f-ダイバージェンスに基づく順序準対称 (OQS[f]) モデルは次のように定義される (Kateri and Agresti [4]):

$$p_{ij} = p_{ij}^S F^{-1}(\alpha u_i + \gamma_{ij})$$
 $(i = 1, ..., r; j = 1, ..., r),$

ただし,

$$p_{ij}^S = \frac{p_{ij} + p_{ji}}{2}, \quad \gamma_{ij} = \gamma_{ji}.$$

Kateri and Agresti [4] は次の定理を示した:

定理3. Sモデルが成り立つための必要十分条件は,OQS[f]モデルと MHモデルの両方が成り立つことである.

行と列のスコアの平均一致(ME)構造を考える:

本研究では,次の定理を得た.

定理 4. Sモデルが成り立つための必要十分条件は,OQS[f]モデルと MEモデルの両方が成り立つことである.

モデルMの適合度を検定するための尤度比カイ二乗統計量を $G^2(M)$ とする.定理1に対して,Tomizawa and Tahata [7] は検定統計量に関する次の定理を与えた:

定理5. $G^2(S)$ は、 $G^2(QS)$ と $G^2(MH)$ の和に漸近的に同等である.

本研究では,次の定理を得た.

定理 6. $G^2(S)$ は、 $G^2(QS[f])$ と $G^2(MH)$ の和に漸近的に同等である.

定理7. $G^2(S)$ は、 $G^2(OQS[f])$ と $G^2(ME)$ の和に漸近的に同等である.

参考文献

- [1] Agresti, A. (2002). Categorical Data Analysis, 2nd edition. Wiley, New York.
- [2] Bowker, A. H. (1948). Journal of the American Statistical Association, 43, 572-574.
- [3] Caussinus, H. (1965). Annales de la Faculté des Sciences de l'Université de Toulouse, **29**, 77-182.
- [4] Kateri, M. and Agresti, A. (2007). Statistics and Probability Letters, 77, 598-603.
- [5] Kateri, M. and Papaioannou, T. (1997). Journal of the American Statistical Association, 92, 1124-1131.
- [6] Stuart, A. (1955). Biometrika, 42, 412-416.
- [7] Tomizawa, S. and Tahata, K. (2007). Journal de la Société Française de Statistique, 148, 3-36.

Pitman nearness 基準の下で平均と分散の双方に順序制約がある 場合の2つの正規母集団の平均の推定とその応用

目白大学・社会学部 張 慶應義塾大学 篠崎 信雄 元宗

1.はじめに

制約された母数空間における推定問題が数多く考えられるが、母平均に関する代表的な線形制約条件は次のようなものが挙げ られる。(1) 非負性 (2) 順序制約 (simple order) (3) simple tree order (4) 傘型順序制約である。例えば、順序制約は、年齢とと もに平均値が大きくなると考えられる量(児童の身長など)薬品の投与量とともに平均的に大きくなると考えられる反応量など の場合に考えられている。このような場合の統計的推測について、古くから様々の研究が進められてきているが、1988年以前の 研究については、Barlow et al. (1972) や Robertson et al. (1988) で詳しく解説されている。その後の発展、特に、点推定及 び区間推定については、Silvapulle & Sen (2005) や van Eeden (2006) のモノグラフによって解説されている。また、国内で は張(Chang)と篠崎(Shinozaki)は制約条件を考慮する最尤推定量(RMLE)による改良問題について研究を推進し、久保川 (Kubokawa), 津熊 (Tsukuma), Marchand, Perron, Strawderman らは一般ベイズ推定量の許容性およびミニマックス性などに ついて精力的に研究を進めている。

推定量の良さを評価する基準として、確率優越性あるいは平均の2乗誤差がよく使われているが、他の基準として Pitman (1937) により提案された Pitman nearness 評価基準がある。 Pitman nearness 評価基準はつぎのように定義される。 定義1. T_1 と T_2 を θ の推定量とする。下記の条件を満たすとき、 T_1 は T_2 より θ に(Pitman の意味で)近いと定義する。

$$P_r\{|T_1 - \theta| < |T_2 - \theta|\} > 1/2.$$

Pitman nearness に関する詳しい議論については Communications in Statistics - Theory and Methods A20 (11)(1992) 及び Keating, Manson & Sen (1993) がまとめたモノグラフを参照されたい。ここでは、母平均と分散の双方に順序制約条件がある 場合の2つの正規母平均の推定問題を Pitman nearness 評価基準の下で考える。

まず、母平均と分散に双方に順序制約条件がある場合の推定量及び確率優越性評価基準の下で得られた結果を紹介する。 $X_{ij}, i=1,2,j=1,\ldots,n_i$ は平均 μ_i 、分散 σ_i^2 の正規分布からの独立な観測値とする。ここで、 μ_i と σ_i^2 は共に未知で、 $ar{X}_i=\sum_{j=1}^{n_i}X_{ij}/n_i, s_i^2=\sum_{j=1}^{n_i}(X_{ij}-ar{X}_i)^2/(n_i-1)$ は μ_i と σ_i^2 の不偏推定量である。Oono & Shinozaki (2005) が母平均に 順序制約 $\mu_1 \leq \mu_2$ があるとき μ_i の打ち切り型推定量

$$\hat{\mu}_1^{OS} = \min\{\bar{X}_1, \hat{\mu}^{GD}\}, \quad \hat{\mu}_2^{OS} = \max\{\bar{X}_2, \hat{\mu}^{GD}\}$$
(1.1)

を提案し、 $\hat{\mu}_i^{OS}$ が制約条件を無視した μ_i の最尤推定量 $ar{X}_i$ を $ext{MSE}$ で改良するための必要十分条件を与えた。ここで、 $\hat{\mu}^{GD}$ は Graybill & Deal (1959) が提案した 2 つの正規分布の共通母平均の推定量

$$\hat{\mu}^{GD} = \frac{n_1 s_2^2}{n_1 s_2^2 + n_2 s_1^2} \bar{X}_1 + \frac{n_2 s_1^2}{n_1 s_2^2 + n_2 s_1^2} \bar{X}_2$$

である。しかし、分散にも順序制約条件 $\sigma_1^2 \leq \sigma_2^2$ がある場合、 $s_1^2 > s_2^2$ のとき、 $n_1 s_2^2/(n_1 s_2^2 + n_2 s_1^2) < n_1/(n_1 + n_2)$ となり不 自然である。そこで、Chang & Shinozaki (2012) は次のような推定量

$$\hat{\mu}_{1}^{CS} = \begin{cases} \hat{\mu}_{1}^{OS}, & if \ s_{1}^{2} \leq s_{2}^{2} \\ \min\{\bar{X}_{1}, \frac{n_{1}}{n_{1}+n_{2}}\bar{X}_{1} + \frac{n_{2}}{n_{1}+n_{2}}\bar{X}_{2}\}, & if \ s_{1}^{2} > s_{2}^{2}, \end{cases}$$

$$\hat{\mu}_{2}^{CS} = \begin{cases} \hat{\mu}_{2}^{OS}, & if \ s_{1}^{2} \leq s_{2}^{2} \\ \max\{\bar{X}_{2}, \frac{n_{1}}{n_{1}+n_{2}}\bar{X}_{1} + \frac{n_{2}}{n_{1}+n_{2}}\bar{X}_{2}\}, & if \ s_{1}^{2} > s_{2}^{2} \end{cases}$$

$$(1.2)$$

$$\hat{\mu}_{2}^{CS} = \begin{cases} \hat{\mu}_{2}^{OS}, & \text{if } s_{1}^{2} \leq s_{2}^{2} \\ \max\{\bar{X}_{2}, \frac{n_{1}}{n_{1}+n_{2}}\bar{X}_{1} + \frac{n_{2}}{n_{1}+n_{2}}\bar{X}_{2}\}, & \text{if } s_{1}^{2} > s_{2}^{2} \end{cases}$$

$$(1.3)$$

を考え、確率優越性の評価の基準の下で、 $\hat{\mu}_i^{CS}$ と $\hat{\mu}_i^{OS}$ との比較を行い、次のような結果を得た。

1. 大きい分散に対応する母平均 μ_2 の推定:

 $\hat{\mu}_2^{CS}$ は $\hat{\mu}_2^{OS}$ より確率的に優れている。つまり、すべての $\mu_1 \leq \mu_2,\,\sigma_1^2 \leq \sigma_2^2$ に対して、

$$P\left\{ \mid \hat{\mu}_2^{CS} - \mu_2 \mid \leq d \right\} \ge P\left\{ \mid \hat{\mu}_2^{OS} - \mu_2 \mid \leq d \right\}, \quad \forall d > 0$$

が成立する。従って、 $ext{MSE}$ を基準にしたとき、 $\hat{\mu}_2^{CS}$ は $\hat{\mu}_2^{OS}$ より優れている。

2. 小さい分散に対応する母平均 μ_1 の推定:

 $\Delta=\mu_2-\mu_1$ が十分大きいとき、 $\hat{\mu}_1^{CS}$ は $\hat{\mu}_1^{OS}$ よりも大きな $ext{MSE}$ をもつ。

 $3.~(\mu_1,\mu_2)$ の同時推定: $(\hat{\mu}_1^{CS},\hat{\mu}_2^{CS})$ は $(\hat{\mu}_1^{CS},\hat{\mu}_2^{CS})$ に対して、

$$P\bigg\{\sum_{i=1}^2 \bigg(\frac{\hat{\mu}_i^{CS} - \mu_i}{\sigma_i^2/n_i}\bigg)^2 \leq d\bigg\} \geq P\bigg\{\sum_{i=1}^2 \bigg(\frac{\hat{\mu}_i^{OS} - \mu_i}{\sigma_i^2/n_i}\bigg)^2 \leq d\bigg\}, \quad \forall \, d>0,$$

が成立する。

4. 広いクラスの推定量:

Chang, Oono & Shinozaki (2012) は、 $\hat{\mu}_i^{OS}, i=1,2$ を含むような、一般の推定量 $\gamma \bar{X}_1+(1-\gamma)\bar{X}_2$ 、を改良する広いクラスの推定量を与えた。 ここで γ は $n_1,n_2,s_1^2,s_2^2,(x_1-x_2)^2$ の関数である。

次に、2 つの母平均に順序がある場合の推定に対して、修正した Pitman nearnee 評価基準を述べる。Nayak (1990) に従い、Gupta & Singh (1992) は正の確率で合致する 2 つの推定量 $T_i, i=1,2$ を比較するときに、次のような modified Pitman nearness(MPN) 評価基準を提案した。

定義 2. T_1 と T_2 を θ の推定量とする。

$$MPN_{\theta}(T_1, T_2) = P_r\{|T_1 - \theta| < |T_2 - \theta||T_1 \neq T_2\}$$

を考え、 $MPN_{\theta}(T_1,T_2)>1/2$ ならば、 T_1 は T_2 より θ に近いと定義する。

ここで、母平均と分散に双方に順序制約条件があるとき、MPN 評価基準の下で、 $\hat{\mu}_i^{CS}$ と $\hat{\mu}_i^{OS}$ との比較をし、確率優越性ある いは MSE の評価基準の下で、Chang & Shinozaki (2012) が得た結果と合致することを明らかにする。 2.結果

1. 大きい分散に対応する母平均 μ_2 の推定:

定理1. $\hat{\mu}_2^{CS}$ は $\hat{\mu}_2^{OS}$ より $\hat{\mu}_2$ に近い。つまり、すべての $\hat{\mu}_1 \leq \hat{\mu}_2$ と $\hat{\sigma}_1^2 \leq \hat{\sigma}_2^2$ に対して、

$$MPN_{\mu_2}(\hat{\mu}_2^{CS}, \hat{\mu}_2^{OS}) > 1/2$$

が成立する。

2. 小さい分散に対応する母平均 μ_1 の推定:

定理 2. $\Delta=\mu_2-\mu_1$ が十分大きいとき、 $\hat{\mu}_1^{CS}$ は $\hat{\mu}_1^{OS}$ より μ_1 に近くならない。つまり、 Δ が十分大きいとき、ある $\sigma_1^2\leq\sigma_2^2$ に対して

$$MPN_{\mu_1}(\hat{\mu}_1^{CS}, \hat{\mu}_1^{OS}) < 1/2$$

が成立する。

 $3. (\mu_1, \mu_2)$ の同時推定:

定理3. $(\hat{\mu}_1^{CS},\hat{\mu}_2^{CS})$ は $(\hat{\mu}_1^{OS},\hat{\mu}_2^{OS})$ より次の意味で (μ_1,μ_2) に近い。

$$MPN\mu(\hat{\mu}^{CS}, \hat{\mu}^{OS}) = P_r \left\{ \sum_{i=1}^{2} (\hat{\mu}_i^{CS} - \mu_i)^2 / (\sigma_i^2/n_i) \le \sum_{i=1}^{2} (\hat{\mu}_i^{OS} - \mu_i)^2 / (\sigma_i^2/n_i), \hat{\mu}^{CS} \ne \hat{\mu}^{OS} \middle| \hat{\mu}^{CS} \ne \hat{\mu}^{OS} \right\} > 1/2$$

4. 広いクラスの推定量:

Chang & Shinozaki (2014) は、 $\hat{\mu}_i^{OS}, i=1,2$ を含むような、一般の推定量 $\gamma \bar{X}_1+(1-\gamma)\bar{X}_2$ 、を改良する広いクラスの推定量を与えた。 ここで γ は $n_1,n_2,s_1^2,s_2^2,(x_1-x_2)^2$ の関数である。

5.数値計算

 $\hat{\mu}_i^{CS}$ と $\hat{\mu}_i^{OS}$ の挙動を明らかにするため、数値計算を行い、その結果を当日に示す。

参考文献: (1)Chang Y.-T. and Shinozaki, N. (2012) "Estimation of Ordered Means of Two Normal Distributions with Ordered Variances", Journal of Mathematics and System Science Vol.2, No.1, pp. 1-7. (2)Chang Y,-T. Oono, Y. and Shinozaki, N. (2012), "Improved estimators for the common mean and ordered means of two normal distributions with ordered variances", Journal of Statistical Planning and Inference, 142, 2619-2628. (3) Chang Y.-T. and Shinozaki, N. (2014) "Estimation of two ordered normal means under modified Pitman nearness criterion", Annals of the Institute of Statistical Mathematics, (to appear). (4) Graybill, F.A. and Deal, R. B. (1959), "Combining unbiased estimators," Biometrics 15, pp. 543-550. (5)Gupta, R. D., Singh, H. (1992). "Pitman nearness comparisons of estimates of two ordered normal means". Australian Journal of Statistics, 34(3), 407–414. (6)Hwang, J. T. (1985), "Universal domination and stochastic domination," Ann. Statist., Vol.13, No.1, pp. 295-314. (7) Keating, J. P., Mason, R. L. and Sen, P. K. (1993): Pitman's measure of closeness: A comparison of statistical estimators, Philadelphia, SIAM. (8) Kubokawa, T. (1989), "Closer estimation of a common mean in the sense of Pitman", Ann. Inst. Statist. Math, Vol. 41, No.3, pp. 477-484. (9) Nayak, T. K. (1990) "Estimation of location and scale parameters using generalized Pitman nearness criterion", Journal of Statist. Plann. and Infer. No. 24, pp. 259-268. (10)Oono, Y. and Shinozaki, N. (2005), "Estimation of two order restricted normal means with unknown and possibly unequal variances," Journal of Statistical Planning and Inference, Vol. 131, Issue2, pp. 349-363. (11) Pitman, E. J. G. (1937), The closest estimates of statistical parameters, Proc. Cambridge Philos. Soc., 33, pp. 212-222. (12) Silvapulle, M. J., Sen, P. K. (2004). Constrained Statistical Inference. Wiley, New Jersey. (13) van Eeden, C., (2006). Restricted Parameter Space Estimation Problems. Lecture notes in Statistics 188, Springer.

謝辞:本研究は科研費 (基盤研究 (C)No.2633067) 助成を受けています。また、著者張元宗は目白大学の特別研究費の助成も受けています。

Frequency domain GMM estimators for stable processes and its asymptotic optimality

Fumiya Akashi *†

1 Introduction

In the last few decades, heavy-tailed data have been observed in variety of fields involving electrical engineering, hydrology, finance and physical systems. To model such data, symmetric α -stable processes are often considered. The symmetric α -stable process is a linear process which is generated from the sequence of independent and identically distributed (i.i.d.) symmetric α -stable random variables, and it is known that the stable processes does not have the finite variance. So the classical methods are not applicable, such as method of moments approach. To overcome this hurdle, we frequently make use of the frequency domain approach and self-normalization. Mikosch et al. [5] constructed Whittle estimator for the stable ARMA models, and showed the consistency and limit distribution of the estimator. Klüppelberg and Mikosch [4] studied the limit behavior of so-called self-normalized periodograms and the integrated version for the stable processes. They elucidated the asymptotic properties of the statistics. In particular, it was shown that the limit distribution of the integrated self-normalized periodograms for the stable processes is expressed as a sum of stable random variables. Subsequently, Akashi, Liu and Taniguchi [1] derived the limit distribution of the frequency domain empirical likelihood ratio statistic for the stable processes, and constructed the method of nonparametric inference for important quantities of time series models. On the other hand, generalized method of moments (GMM) is one of the most popular tools in econometrics, and the method has been applied in variety of fields. In i.i.d. case, Newey and McFadden [6] gave a unified view for the estimation problems based on the class of extremum estimators which contains the GMM estimators and the maximum likelihood estimators as special cases. For dependent data, Hansen [2] introduced the GMM approach for time series models with the finite second moment. Hansen [2] considered the moment restriction models in time domain, and Kakizawa [3] extended Hansen's approach to frequency domain. Kakizawa [3]

^{*}Graduate School of Fundamental Science and Engineering, Waseda University.

[†]This work was supported by Grant-in-Aid for JSPS Fellows Grant Number 26.3721.

proposed the frequency domain GMM estimator, and showed the consistency and asymptotic normality of the estimator. Furthermore, Kakizawa [3] gave the optimal weighting matrix of the GMM estimator, and constructed the asymptotic optimal GMM estimator without assuming that the true model is known.

In this paper, we extend the results in Kakizawa [3] to the stable processes. We focus on the estimation problem of over-identified important quantities of the stable processes, and derive the limit distribution of the frequency domain GMM estimator. Moreover, the optimality of the estimator will be discussed based on the limit distribution. Since the stable distribution does not have the finite second moments, we define two types of criteria: one of them is the pseudo covariance matrix, and the other one is the codifference, which was introduced by Samoradnitsky and Taqqu [7]. We also provide the discussion on the asymptotic optimality of the GMM estimator for stable processes via theoretical results and some simulation experiments.

References

- [1] Akashi, F., Liu, Y., and Taniguchi, M. (2014). An empirical likelihood approach for symmetric α-stable processes. *To appear in Bernoulli*.
- [2] Hansen, L. P. (1982). Large sample properties of generalized method of moments estimators. *Econometrica: Journal of the Econometric Society*, pages 1029–1054.
- [3] Kakizawa, Y. (2013). Frequency domain generalized empirical likelihood method. Journal of Time Series Analysis, 34(6):691–716.
- [4] Kluppelberg, C. and Mikosch, T. (1996). The integrated periodogram for stable processes. *The Annals of Statistics*, pages 1855–1879.
- [5] Mikosch, T., Gadrich, T., Kluppelberg, C., and Adler, R. J. (1995). Parameter estimation for arma models with infinite variance innovations. The Annals of Statistics, pages 305–326.
- [6] Newey, W. K. and McFadden, D. (1994). Large sample estimation and hypothesis testing. *Handbook of econometrics*, 4:2111–2245.
- [7] Samoradnitsky, G. and Taqqu, M. S. (1994). Stable non-Gaussian random processes: stochastic models with infinite variance, volume 1. CRC Press.