科学研究費シンポジウム　プログラム

研究集会：　統計推測理論の展開と諸モデルへの応用

開催日：　2012 年 10 月 3 日 (水)〜10 月 5 日 (金)

場所：　釧路市生涯学習センター (TEL: 0154–41–8181)

http://www.kushiro-bunka.or.jp/manabo/koutuuannai.html

研究代表者：谷口正信（早稲田大学基幹理工学部）
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」（課題番号：23244011）

開催責任者：関谷祐里（北海道教育大学教育学部釧路校）
種市信裕（鹿児島大学大学院理工学研究科）
鈴川晶夫・柿沢佳秀（北海道大学大学院経済学研究科）

10 月 3 日 (水曜日)

10:00〜12:30 (研究題目についての事前打ち合わせ)

【1–1】 座長: 鈴川晶夫 (北海道大学)

13:50〜14:20 種市信裕 (鹿児島大学), 関谷祐里 (北海道教育大学), 外山 淳 (数学利用研究所)
二項反応の一般化線型モデルにおけるパワーダイバージェンス適合度検定統計量の改良

14:30〜15:00 五十嵐 岳 (北海道大学), 柿沢佳秀 (北海道大学)
境界バイアスのない密度推定量の改良について

【1–2】 座長: 柿沢佳秀 (北海道大学)

15:20〜15:50 川崎玉恵 (東京理科大学), 瀬尾 隆 (東京理科大学)
異なる分散共分散行列をもつ２つの平均ベクトルの検定に対する検定統計量の近似分布について

16:00〜16:30 世古規子 (東京理科大学), 瀬尾 隆 (東京理科大学)
Tests for mean vector with two-step monotone missing data

16:40〜17:10 大塚芳宏 (北海道大学)
Bayesian Inference for Stochastic Volatility Model with Spatial Correlation: Application to Regional Business Cycle in Japan

10 月 4 日 (木曜日)

【2–1】 座長: 種市信裕 (鹿児島大学)

10:00〜10:30 小倉寛生 (北海道教育大学), 関谷祐里 (北海道教育大学)
２×２分割表の対称性検定におけるカイ二乗統計量の分布のエッジワース展開について

10:40〜11:10 金川秀也 (東京都市大学)
Asymptotic expansion for Hilbert space valued random variables and its application to symmetric statistics

11:20〜11:50 鈴川晶夫 (北海道大学)
２変量極値分布のノンパラメトリック推定

【2–2】 座長: 瀬尾 隆 (東京理科大学)

13:40〜14:10 生亀清貴 (東京理科大学), 田畑耕治 (東京理科大学), 富澤貞男 (東京理科大学)
多変量密度関数の対称性に関する分解

14:20〜14:50 山本紘司 (大阪大学), 富澤貞男 (東京理科大学)
正方分割表における累積確率に基づく非対称モデルと分解

15:00〜15:30 広津千尋 (明星大学)
分割表解析で有用な幾つかの多重比較法

【2–3】 座長: 加藤 剛 (上智大学)

15:50〜16:20 三浦良造 (一橋大学名誉教授)
一般化されたレーマン対立仮説を用いた市場モデルと CAPM の Jensen の α, そして順位推定

16:30〜17:00 刈屋武昭 (明治大学), 山村能郎 (明治大学), 乾 孝治 (明治大学), 王 竹 (ZW システム)
個別企業の信用価格スプレッドと倒産確率の導出

10 月 5 日 (金曜日)

【3–1】 座長: 金川秀也 (東京都市大学)

10:00〜10:30 山方 亮 (上智大学)
CAPM における回帰分析と最適ポートフォリオの構成

10:40〜11:10 加藤 剛 (上智大学)
大規模データを対象にした推定問題における高速処理

11:20〜11:50 甫喜本 司 (東京大学)
回遊性魚類の行動予測におけるマルコフスイッチング構造の効果

11:50 終わりに

二項反応の一般化線型モデルにおけるパワーダイバージェンス適合度検定統計量の改良

鹿児島大学・理工　種市信裕
北海道教育大学・釧路　関谷祐里
数学利用研究所　外山　淳

1 はじめに

我々は，先の研究（Taneichi et al. [4]）において，ロジスティック回帰モデルにおけるデビアンス（対数尤度比統計量）D の帰無仮説のもとでの分布に対する漸近展開式を導出し，その連続項に基づき D にバートレット修正を施した変換統計量 $\tilde{D}$ を構築した．さらにその上で，$\tilde{D}$ の検出力は D とほぼ同じで，極限カイ二乗分布への収束が D よりも速いことを数値的に実証した．本報告では，Taneichi et al. [4] の研究内容を，ロジスティックモデルから，より一般の一般化線型モデルへ，また検定統計量もデビアンスからパワーダイバージェンス (Cressie and Read [1]) に基づく検定統計量へと拡張をおこなった．これにより，二項反応の多様な一般化線型モデルに対する適合度検定において，標本数があまり大きくない場合であっても，極限カイ二乗分布を用いた近似検定によって，適切な検定結果を導きやすい新たな検定統計量を提案した．

2　二項反応の一般化線型モデル

一般化線型モデル (Nelder and Wedderburn [3]) を 2 項分布 $B(n,\pi)$ について考える．N 個の異なるサブグループにおける反応数に対応した確率変数 $Y_\alpha,\ (\alpha=1,\ldots,N)$ が互いに独立に二項分布 $B(n_\alpha,\pi_\alpha),\ (\alpha=1,\ldots,N)$ に従うとし，その連結関数として，単調かつ微分可能な関数 $g(\cdot)$ を用いると，二項データに対する一般化線型モデル

$$g(\pi_\alpha)=\boldsymbol{x}_\alpha'\boldsymbol{\beta},\quad(\alpha=1,\ldots,N) \tag{1}$$

が得られる．ただし，$\boldsymbol{x}_\alpha=(x_{\alpha1},\ldots,x_{\alpha p})',\ (\alpha=1,\ldots,N)$ は共変量ベクトル，$\boldsymbol{\beta}=(\beta_1,\ldots,\beta_p)'$ は未知のパラメータベクトルであり，$p<N$ とする．関数 g として特に，$g(t)=\log\{t/(1-t)\},\ g(t)=\Phi^{-1}(t),\ g(t)=\log\{-\log(1-t)\},$ を用いたときのモデル (1) はそれぞれ，ロジスティックモデル，プロビットモデル，補対数対数モデルとなる．ここで，$\Phi(\cdot)$ は標準正規分布の累積分布関数である．本報告で扱うモデルは，これら種々のモデルを含む一般の一般化線型モデル (1) を対象とする．

3　一般化線型モデルの適合度検定におけるパワーダイバージェンス統計量

一般化線型モデルが正しいという帰無仮説

$$H_0^g:\pi_\alpha=g^{-1}(\boldsymbol{x}_\alpha'\boldsymbol{\beta}),\ (\alpha=1,\ldots,N)$$

を検定するためのパワーダイバージェンス統計量は，

$$R^a=2\sum_{\alpha=1}^{N}n_\alpha\left\{I^a\left(\frac{Y_\alpha}{n_\alpha},\ \hat{\pi}_\alpha^g\right)+I^a\left(1-\frac{Y_\alpha}{n_\alpha},\ 1-\hat{\pi}_\alpha^g\right)\right\},$$

ただし，$I^a(e,f)=\{a(a+1)\}^{-1}e\{(e/f)^a-1\}\ (a\neq0,-1);\ e\log(e/f)\ (a=0);\ f\log(f/e)\ (a=-1),$ で与えられる．また，$\hat{\pi}_\alpha^g=\pi_\alpha(\hat{\boldsymbol{\beta}}^g),\ (\alpha=1,\ldots,N)$ であり，$\hat{\boldsymbol{\beta}}^g=(\hat{\beta}_1^g,\ldots,\hat{\beta}_p^g)'$ は

帰無仮説 H_0^g のもとでの $\boldsymbol{\beta}$ の最尤推定量である．ここで, 検定統計量 R^a は, 連結関数 g に応じて異なる統計量であることを注意しておく．$n = \sum_{\alpha=1}^{N} n_\alpha$ とおくとき, 条件

$$n_\alpha/n \to \mu_\alpha,\ (\alpha = 1, \ldots, N) \quad \text{as} \quad n \to \infty, \tag{2}$$

ただし，$0 < \mu_\alpha < 1, (\alpha = 1, \ldots, N)$, $\sum_{\alpha=1}^{N} \mu_\alpha = 1$, が成り立つならば，パワーダイバージェンス統計量 R^a は, a の値によらず帰無仮説 H_0^g のもとで $n \to \infty$ に伴って漸近的に自由度 $N-p$ のカイ二乗分布に従う．このことを用いて, 二項データが, (1) で与えられる一般化線型モデルに従うかどうかの適合度検定をおこなうことができる．

4　パワーダイバージェンス統計量の改良

本報告では，R^a よりもさらに極限カイ二乗分布への収束の速い統計量を構築するために, 帰無仮説 H_0^g のもとでの分布の漸近展開に基づく近似として，Yarnold [5] の考え方に従い,

$$\Pr\{R^a \le x | H_0^g\} \approx J_1^{g,a}(x) + J_2^{g,a}(x)$$

という近似を考えた．ここで，$J_1^{g,a}(x)$ は連続分布を仮定した多変量エッジワース展開の項であり，$J_2^{g,a}(x)$ は不連続性を考慮した離散項である．R^a の分布の漸近展開式を求めるために, 条件 (2) の代わりに，次の仮定１を考える．

仮定１:$n_\alpha/n = \mu_\alpha,\ (\alpha = 1, \ldots, N)$ という条件を満たしながら，$n_\alpha \to \infty,\ (\alpha = 1, \ldots, N)$ as $n \to \infty$, ただし，$0 < \mu_\alpha < 1, (\alpha = 1, \ldots, N), \sum_{\alpha=1}^{N} \mu_\alpha = 1$.

すると，g^{-1} が 4 回連続微分可能であるとき，仮定１のもとで，$J_1^{g,a}(x)$ は,

$$J_1^{g,a}(x) = \Pr\{\chi^2_{N-p} \le x\} + \frac{1}{n}\sum_{j=0}^{3} w_j^{g,a} \Pr\{\chi^2_{N-p+2j} \le x\} + O(n^{-2}) \tag{3}$$

という形式で評価される．ここで，$\sum_{j=0}^{3} w_j^{g,a} = 0$ という関係が成り立つ．さらに, $a = 0$ の場合には特に, $w_3^{g,0} = w_4^{g,0} = 0$ となる．また，χ^2_f は自由度 f のカイ二乗分布に従う確率変数を表す．(3) 式に, Bartlett 修正および改良変換の構築と漸近展開式との関係の理論 (e.g. Fujikoshi [2]) を適用した．ところで, 離散項 $J_2^{g,a}(x)$ の評価式は非常に複雑であること, および $J_2^{g,a}(x) = O(n^{-1/2})$ であることから，R^a の分布に対して，(3) 式で与えられる $J_1^{g,a}(x)$ の近似式のみを用いて，小標本におけるカイ二乗近似の改良を考えた．$a = 0$ の場合にはデビアンス, すなわち R^0 にバートレット修正を施した変換統計量

$$\tilde{R}_B^0 = \left\{1 + \frac{2\hat{w}_0^{g,0}}{n(N-p)}\right\} R^0$$

を考え，$a \neq 0$ の場合には R^a の対数形の改良変換統計量 (e.g. Fujikoshi [2]) $\tilde{R}_I^a$ $(a \neq 0)$ を構築した．さらに, $\tilde{R}_B^0$, $\tilde{R}_I^a$ のカイ二乗分布への収束の速さおよび検出力を数値的に考察した．

参考文献

[1] Cressie, N. and Read, T. R. C.: *J. R. Statist. Soc.* B, **46** (1984), 440–464.

[2] Fujikoshi, Y.: *J. Mult. Anal.* **72** (2000), 249–263.

[3] Nelder, J. A. and Wedderburn, R. W. M.: *J. R. Statist. Soc.*, A, **135** (1972), 370–384.

[4] Taneichi, N., Sekiya, Y. and Toyama, J.: *J. Mult. Anal.*, **102** (2011), 1263–1279.

[5] Yarnold, J. K.: *Ann. Math. Statist.*, **43** (1972), 1566–1580.

境界バイアスのない密度推定量の改良について

北海道大学　五十嵐　岳　　北海道大学　柿沢　佳秀

Rosenblatt (1956; AMS) を先駆けとする通常のカーネル推定量は

$$\hat{f}_h^{(K)}(x) = \frac{1}{nh}\sum_{i=1}^{n} K(\frac{x - X_i}{h})$$

である．ここで $X_1, \ldots, X_n$ は互いに独立で同一の密度 f に従う確率変数, $h > 0$ はバンド幅で, $h \to 0, nh \to \infty$ $(n \to \infty)$. カーネル推定量は, 推定する密度 f の台が $(-\infty, \infty)$ の場合, 次のような性質を持つ.

(i). 2次カーネルのとき, バイアスは$O(h^2)$, 分散は$O(n^{-1}h^{-1})$で, 最適なMSE, MISEが$O(n^{-4/5})$である．ただし, p 次カーネル K_p は

$$\int_{-\infty}^{\infty} K_p(u)du = 1,\ \int_{-\infty}^{\infty} u^j K_p(u)du = 0,\ j = 1, \ldots, p-1,\ \int_{-\infty}^{\infty} u^p K_p(u)du \neq 0$$

を満たす.

(ii). f が 4 階導関数を持てば, 4 次カーネルのとき, バイアスは $O(h^4)$ になり, 分散は $O(n^{-1}h^{-1})$ のままで, 最適な MSE, MISE が $O(n^{-8/9})$ になる．なお, Schucany and Sommers (1977; JASA) による加法型推定量

$$\hat{f}_{ADD,h,ch}^{(K)}(x) = \frac{c^2}{c^2-1}\hat{f}_h^{(K)}(x) - \frac{1}{c^2-1}\hat{f}_{ch}^{(K)}(x), \quad c \neq 1$$

は 4 次カーネル推定量の特別な場合である.

(iii). 高次カーネル推定量は負になり得るため, Terrell and Scott (1980; AS), Jones et al. (1995; Biometrika) は

$$\hat{f}_{TS,h}^{(K)}(x) = \{\hat{f}_h^{(K)}(x)\}^{4/3}\{\hat{f}_{2h}^{(K)}(x)\}^{-1/3}, \quad \hat{f}_{JLN,h}^{(K)}(x) = \frac{\hat{f}_h^{(K)}(x)}{nh}\sum_{i=1}^{n}\frac{K(\frac{x-X_i}{h})}{\hat{f}_h^{(K)}(X_i)}$$

のような非負性を保持した推定量を提案し, バイアスが $O(h^4)$ になり, 分散は $O(n^{-1}h^{-1})$ のままで, 最適な MSE, MISE が $O(n^{-8/9})$ になることを示した.

しかし, 推定する密度 f の台が有界区間や半無限区間 $[0, \infty)$ の場合, 境界付近でバイアスが $O(1)$ になるという境界問題が存在する. 境界バイアスを改良する推定量として, リノーマライゼーション法やリフレクション法, 一般化ジャックナイフ法など境界付近で異なるカーネルを用いる手法がいくつか提案されている. 境界付近でも内部と同じカーネルを用いて境界バイアスを改良する推定方法として, $[0, 1]$ の台を持つ密度に対する, ベルンシュタイン推定量 (Vitale, 1975; Statistical Inference and Related Topics Vol.2), ベータカーネル推定量 (Chen, 1999; CSDA), 正規コピュラ推定量 (Jones and Henderson, 2007; Biometrika) などが提案されている. Hirukawa (2010; CSDA) は, ベータカーネル推定量の TS 型, JLN 型修正を議論した (五十嵐は修士論文で加法型修正も追加検討した). Leblanc (2010; JNS) は, ベルンシュタイン推定量の加法型, 五十嵐, 柿沢 (2011; 統計関連学会連合大会) はベルンシュタイン推定量の TS 型, JLN 型修正を議論した. 一方, 境界と内

部で同じカーネルを用いる, $[0,\infty)$ の台を持つ密度に対する境界バイアスのない推定量として, ガンマカーネル推定量 (Chen, 2000; AISM), 逆ガウスカーネル推定量, 相反逆ガウスカーネル推定量 (Scaillet, 2004; JNS)

$$\hat{f}_b^{(G1)}(x) = \frac{1}{n}\sum_{i=1}^{n} K_{x/b+1,b}^{(G)}(X_i), \quad K_{p,q}^{(G)}(s) = \frac{s^{p-1}e^{-s/q}}{q^p\Gamma(p)},$$
$$\hat{f}_b^{(IG_s)}(x) = \frac{1}{n}\sum_{i=1}^{n} K_{x,1/b}^{(IG)}(X_i), \quad K_{m,\lambda}^{(IG)}(s) = \frac{\sqrt{\lambda}}{\sqrt{2\pi s^3}}\exp\Big\{-\frac{\lambda}{2m}\Big(\frac{s}{m}-2+\frac{m}{s}\Big)\Big\},$$
$$\hat{f}_b^{(RIG_s)}(x) = \frac{1}{n}\sum_{i=1}^{n} K_{1/(x-b),1/b}^{(RIG)}(X_i), \quad K_{m,\lambda}^{(RIG)}(s) = \frac{\sqrt{\lambda}}{\sqrt{2\pi s}}\exp\Big\{-\frac{\lambda}{2m}\Big(ms-2+\frac{1}{ms}\Big)\Big\}$$

などが提案されている. ここで $b>0$ は平滑化 (形状) パラメータで, $b\to 0,\ nb\to\infty$ $(n\to\infty)$ (相反逆ガウスカーネル推定量については $nb^2\to\infty$). いずれもバイアスは $O(b)$ であり, 分散は $O(n^{-1}b^{-1/2})$ であるため, 最適な MSE, MISE は $O(n^{-4/5})$ となる.

しかし, Scaillet (2004; JNS) の提案した逆ガウスカーネル推定量は密度 f に関わらず境界 $x=0$ で 0 を推定し, 相反逆ガウスカーネル推定量については, $x<b$ で $O(1)$ の境界バイアスを持つため, 境界付近でも正常に推定できるようパラメータを変更して推定量を

$$\hat{f}_b^{(IG)}(x) = n^{-1}\sum_{i=1}^{n} K_{x+b,(x+b)^2/b}^{(IG)}(X_i), \quad \hat{f}_b^{(RIG)}(x) = n^{-1}\sum_{i=1}^{n} K_{1/(x+b),1/b}^{(RIG)}(X_i)$$

と再定義する. この再定義された推定量は $O(b)$ のバイアスと $O(n^{-1}b^{-1/2})$ の分散を持ち, 最適な MSE, MISE は $O(n^{-4/5})$ である.

本報告では, これら $[0,\infty)$ に関する境界バイアスのない推定量の加法型修正

$$\hat{f}_{ADD,b,ab}^{(\#)}(x) = \frac{a}{a-1}\hat{f}_b^{(\#)}(x) - \frac{1}{a-1}\hat{f}_{ab}^{(\#)}(x), \quad a>1,$$

TS 型修正

$$\hat{f}_{TS,b,a}^{(\#)}(x) = \Big\{\hat{f}_b^{(\#)}(x)\Big\}^{a/(a-1)}\Big\{\hat{f}_{ab}^{(\#)}(x)\Big\}^{-1/(a-1)}, \quad a>1,$$

JLN 型修正

$$\hat{f}_{JLN,b}^{(G1)}(x) = \hat{f}_b^{(G1)}(x)n^{-1}\sum_{i=1}^{n}\frac{K_{x/b+1,b}^{(G)}(X_i)}{\hat{f}_b^{(G1)}(X_i)},$$
$$\hat{f}_{JLN,b}^{(IG)}(x) = \hat{f}_b^{(IG)}(x)n^{-1}\sum_{i=1}^{n}\frac{K_{x+b,(x+b)^2/b}^{(IG)}(X_i)}{\hat{f}_b^{(IG)}(X_i)},$$
$$\hat{f}_{JLN,b}^{(RIG)}(x) = \hat{f}_b^{(RIG)}(x)n^{-1}\sum_{i=1}^{n}\frac{K_{1/(x+b),1/b}^{(RIG)}(X_i)}{\hat{f}_b^{(RIG)}(X_i)}$$

について, バイアスのオーダーが $O(b)$ から $O(b^2)$ に改良され, 分散のオーダーは $O(n^{-1}b^{-1/2})$ のまま変わらないため, MSE のオーダーは $O(n^{-4/5})$ から $O(n^{-8/9})$ へと改良されることを報告する. さらに, 加法型と TS 型の修正では, 第 2 のパラメータ a が含まれていて, 加法型, TS 型修正は $a\to 1$ とすると MSE を小さくできる. $a\to 1$ のとき得られる極限の推定量の漸近的性質も報告する.

異なる分散共分散行列をもつ2つの平均ベクトルの検定に対する検定統計量の近似分布について

東京理科大・理・院　川崎　玉恵
東京理科大・理　　瀬尾　隆

1　はじめに

平均ベクトルの検定を考える場合，分散共分散行列がグループ間で等しいものを仮定することがあるが，この仮定が成り立たない場合は多変量 Behrens-Fisher 問題と呼ばれ，多くの議論がなされている．このとき，Welch(1938) による検定統計量を多変量に拡張した統計量（この統計量を T 統計量と表すことにする）が考えられ，その近似分布に関する議論が数多くなされている．たとえば James(1954) は，T 統計量の Cornish-Fisher 展開によるカイ 2 乗分布を用いた近似解を与え，また，Yanagihara and Yuan (2005) では F 近似による自由度調整を行った近似解，修正 Bartlett 補正における χ^2 近似解を与えている．本報告では，2 標本問題において分散共分散行列が異なる平均ベクトルの検定について，Yanagihara and Yuan(2005) における近似法をより精密にし，モンテカルロ・シミュレーションによって数値比較を行った．

2　自由度近似

n_i 個の確率ベクトル $\boldsymbol{x}_{i1},\ldots,\boldsymbol{x}_{in_i}$ がそれぞれ独立に $N_p(\boldsymbol{\mu}_i,\Sigma_i),\ i=1,2$ に従うとする．このとき以下の仮説を考えた．

$$H_0:\boldsymbol{\mu}_1=\boldsymbol{\mu}_2 \ \text{ vs. } \ H_1:\boldsymbol{\mu}_1\neq\boldsymbol{\mu}_2$$

ただし，$\Sigma_1\neq\Sigma_2$ とする．この仮説検定において，検定統計量 T は

$$T=(\overline{\boldsymbol{x}}_1-\overline{\boldsymbol{x}}_2)'\left(\frac{S_1}{n_1}+\frac{S_2}{n_2}\right)^{-1}(\overline{\boldsymbol{x}}_1-\overline{\boldsymbol{x}}_2)$$

として与えられている．ただし

$$\overline{\boldsymbol{x}}_i=\frac{1}{n_i}\sum_{j=1}^{n_i}\boldsymbol{x}_{ij},\ S_i=\frac{1}{n_i-1}\sum_{j=1}^{n_i}(\boldsymbol{x}_{ij}-\overline{\boldsymbol{x}}_i)(\boldsymbol{x}_{ij}-\overline{\boldsymbol{x}}_i)',\quad i=1,2$$

である．ここで T 統計量は $\Sigma_1=\Sigma_2$ かつ $n_1=n_2$ の場合，Hotelling の T^2 統計量に相当し，H_0 の下で $(n-p-1)T/p(n-2)$ は自由度 $p,n-p-1$ の F 分布に従う．ただし，$n=n_1+n_2$ とする．一方，$\Sigma_1\neq\Sigma_2, n_1\neq n_2$ の場合の H_0 の下での厳密な T 統計量の分布は，Nel et al.(1990) によって与えられているが，この厳密な分布は計算的に扱いにくいものとなっており，現実問題には適していない．

Yanagihara and Yuan(2005) は，T 統計量を次のように表した．

$$T=\boldsymbol{z}'W^{-1}\boldsymbol{z}=\frac{\boldsymbol{z}'\boldsymbol{z}}{U}$$

ただし

$$\boldsymbol{z}=\sqrt{\frac{n_1n_2}{n}}\overline{\Sigma}^{-1/2}(\overline{\boldsymbol{x}}_1-\overline{\boldsymbol{x}}_2),\ \ W=\overline{\Sigma}^{-1/2}\left(\frac{n_2}{n}S_1+\frac{n_1}{n}S_2\right)\overline{\Sigma}^{-1/2},$$

$$\overline{\Sigma}=\frac{n_2}{n}\Sigma_1+\frac{n_1}{n}\Sigma_2,\ \ U=\frac{\boldsymbol{z}'\boldsymbol{z}}{\boldsymbol{z}'W^{-1}\boldsymbol{z}}$$

である．ここで，U の近似分布は

$$U\approx\frac{\chi^2_\nu}{\phi}\tag{2.1}$$

のように与えることができ，よって T 統計量の近似分布は

$$\frac{\nu}{p\phi}T \approx F_{p,\nu}$$

とみなすことができる．ここで，ν と ϕ は，次のように U の 1 次，2 次モーメントを求めることによって得ることができた．

$$\rho_i = \sqrt{\frac{n_i-1}{n-2}},\ \ \Omega_i = \sqrt{\frac{n-n_i}{n}}\overline{\Sigma}^{-1/2}\Sigma_i^{1/2},\ \ V_i = \sqrt{n_i-1}(\Sigma_i^{-1/2}S_i\Sigma_i^{-1/2} - I_p),\ \ i=1,2$$

とおくと，U は次のように展開することができた．

$$\begin{aligned} U =& 1 + \frac{1}{\sqrt{N}}Q_1 + \frac{1}{N}(Q_1^2 - Q_2) + \frac{1}{N\sqrt{N}}(Q_3 - 2Q_1Q_2 + Q_1^3) \\ &+ \frac{1}{N^2}(Q_1^4 - Q_4 + 2Q_1Q_3 + Q_2^2 - 3Q_1^2Q_2) + O_p(N^{-5/2}). \end{aligned}$$

ただし，$Q_\ell = \boldsymbol{z}'\overline{V}^\ell\boldsymbol{z}/\boldsymbol{z}'\boldsymbol{z}, \ell = 1,2,\ldots,4,\ \overline{V} = \rho_1^{-1}\Omega_1V_1\Omega_1' + \rho_2^{-1}\Omega_2V_2\Omega_2'$ である．同様に U^2 を展開すると次のようになった．

$$\begin{aligned} U^2 =& 1 + \frac{2}{\sqrt{N}}Q_1 + \frac{1}{N}(3Q_1^2 - 2Q_2) + \frac{2}{N\sqrt{N}}(Q_3 - 3Q_1Q_2 + 2Q_1^3) \\ &+ \frac{1}{N^2}(5Q_1^4 - 2Q_4 + 6Q_1Q_3 + 3Q_2^2 - 12Q_1^2Q_2) + O_p(N^{-5/2}). \end{aligned}$$

したがって，Q_ℓ に関する期待値を計算することによって

$$\begin{aligned} \mathrm{E}[U] &= 1 - \frac{\theta_1}{N} + \frac{1}{N^2}(\theta_2 - \theta_3) + O(N^{-3}), \\ \mathrm{E}[U^2] &= 1 - \frac{2}{N}(\theta_1 - \theta_4) + \frac{1}{N^2}(2\theta_5 - \theta_6) + O(N^{-3}) \end{aligned}$$

を得た．ただし $N = n-2$ とし，また $\theta_k,\ k = 1,2,\ldots,6$ は $p, n, n_i, \Sigma_i, \overline{\Sigma}$ を用いて表されるもので，(2.1) 式より

$$\mathrm{E}[U] \approx \frac{\nu}{\phi},\ \ \mathrm{E}[U^2] \approx \frac{\nu(\nu+2)}{\phi^2} \tag{2.2}$$

となり，$\mathrm{E}(U)$ と $\mathrm{E}(U^2)$ の展開式と (2.2) とを連立方程式で解くことにより，次のような ν と ϕ の新たな近似解を得ることができた．

$$\begin{aligned} \nu_{KS} &= \frac{2(N^2 - N\theta_1 + \theta_2 - \theta_3)^2}{N^2(N^2 - 2N\theta_1 + 2N\theta_4 + 2\theta_5 - \theta_6) - (N^2 - N\theta_1 + \theta_2 - \theta_3)^2}, \\ \phi_{KS} &= \frac{N^2\nu_{KS}}{N^2 - N\theta_1 + \theta_2 - \theta_3}. \end{aligned}$$

ここで，ν_{KS} と ϕ_{KS} は Yanagihara and Yuan (2005) の N^{-1} の項までの展開による結果を次項まで展開し，拡張した結果となっている．また，100 万回のモンテカルロ・シミュレーションにより本報告の新しい結果と先行研究の結果の比較を行った．

参考文献

[1] James, G. S. (1954). Tests of linear hypotheses in univariate and multivariate analysis when the ratios of the population variances are unknown. *Biometrika*, **41**, 19–43.

[2] Nel, D. G., van der Merwe, C. A. and Moser, B. K. (1990). The exact distributions of the univariate and multivariate Behrens-Fisher statistics with a comparison of several solutions in the univariate case. *Comm. Statist., Theory Methods*, **19**, 279–298.

[3] Yanagihara, H. and Yuan, K. (2005). Three approximate solutions to the multivariate Behrens-Fisher problem. *Comm. Statist., Simulation Comput.*, **34**, 975–988.

[4] Welch, B. L. (1938). The significance of the difference between two means when the population variances are unequal. *Biometrika*, **29**, 350–362.

Tests for mean vector with two-step monotone missing data

東京理科大・理・院　世古　規子
東京理科大・理　　瀬尾　隆

単調欠測データの 1 標本または 2 標本問題における平均ベクトルの検定について，いくつかの論文で，ホテリング T^2 型統計量とその近似分布の議論がされている (Krishnamoorthy and Pannala(1999), Chang and Richards(2009), Yu, Krishnamoorthy and Pannala (2006) などを参照)．完全データの場合, ホテリング T^2 統計量は帰無仮説の下で正確な F 分布となるが, 欠測データの場合, 正確に F 分布にはならないため, その漸近分布である χ^2 分布を用いることになる．しかし, 標本数が少ない場合, χ^2 分布の近似精度はあまり良くない．そこで, Seko, Yamazaki and Seo (2012) では，2-step 単調欠測データの下での 1 標本問題に対するホテリング T^2 型統計量について，F 分布を用いた近似精度の良い新たな上側近似パーセント点を提案した．さらに，Seko, Kawasaki and Seo (2012) では，Seko, Yamazaki and Seo (2012) の議論を 2 標本問題へ拡張し，同様に F 分布を用いた上側近似パーセント点を提案している．

本報告では，k 標本問題における 2-step 単調欠測データの下での平均ベクトルの検定について考えた．$N_1^{(i)}$ 個の $p(=p_1+p_2)$ 変量観測ベクトル $\boldsymbol{x}_j^{(i)} = (\boldsymbol{x}_{1j}^{(i)'}, \boldsymbol{x}_{2j}^{(i)'})', j = 1, 2, \ldots, N_1^{(i)}$ が $N_p(\boldsymbol{\mu}^{(i)}, \boldsymbol{\Sigma})$ に従い, $N_2^{(i)}$ 個の p_1 変量観測ベクトル $\boldsymbol{x}_{1j}^{(i)}, j = N_1^{(i)}+1, \ldots, N^{(i)}$ が $N_{p_1}(\boldsymbol{\mu}_1^{(i)}, \boldsymbol{\Sigma}_{11})$ に従っているとする. ただし

$$\boldsymbol{\mu}^{(i)} = (\boldsymbol{\mu}_1^{(i)'}, \boldsymbol{\mu}_2^{(i)'})', \quad \boldsymbol{\Sigma} = \begin{pmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{pmatrix}, \quad i = 1, 2 \ldots, k$$

である．このようなデータセットを 2-step 単調欠測データと呼ぶ. まず最初に，k 個の平均ベクトルがすべて等しいとする仮説検定問題

$$H_0 : \boldsymbol{\mu}^{(1)} = \boldsymbol{\mu}^{(2)} = \cdots = \boldsymbol{\mu}^{(k)} \text{ vs. } H_1 : \neq H_0 \tag{1}$$

に対する尤度比検定統計量を与えた．尤度比検定統計量は漸近的に自由度 $p(k-1)$ の χ^2 分布に従う．

次に，帰無仮説 (1) が棄却されたとき，任意の 2 群の平均ベクトルが等しいとする帰無仮説

$$H_0 : \boldsymbol{\mu}^{(a)} = \boldsymbol{\mu}^{(b)} \text{ vs. } H_1 : \boldsymbol{\mu}^{(a)} \neq \boldsymbol{\mu}^{(b)} \text{ (for all } a, b,\ 1 \leq a < b \leq k) \tag{2}$$

に対するホテリング T^2 型検定統計量を与えた．固定された a, b において, 2 群が等しいとする帰無仮説に対するホテリング T^2 型検定統計量 (Seko, Kawasaki and Seo (2012) を参照) は

$$T_{ab}^2 = (\widehat{\boldsymbol{\mu}}^{(a)} - \widehat{\boldsymbol{\mu}}^{(b)})' \widehat{\boldsymbol{\Gamma}}^{-1} (\widehat{\boldsymbol{\mu}}^{(a)} - \widehat{\boldsymbol{\mu}}^{(b)})$$

と構成することができる．ここで，$\widehat{\boldsymbol{\mu}}^{(a)}, \widehat{\boldsymbol{\mu}}^{(b)}$ は，それぞれ 2-step 単調欠測データの下での $\boldsymbol{\mu}^{(a)}$ と $\boldsymbol{\mu}^{(b)}$ の最尤推定量であり，$\widehat{\boldsymbol{\Gamma}}$ は $\widehat{\boldsymbol{\mu}}^{(a)} - \widehat{\boldsymbol{\mu}}^{(b)}$ の分散共分散行列 $\boldsymbol{\Gamma}$ の推定量である．この T_{ab}^2 統計量を用いると，(2) に対する検定統計量は，以下のように与えることができる (Siotani, Hayakawa and Fujikoshi (1985) を参照).

$$T_{\max}^2 = \max_{1 \leq a < b \leq k} T_{ab}^2.$$

この $T_{\max}^2$ 統計量の上側 100α パーセントを t_α^2 とおくと，t_α^2 は以下の式を満たすものである．

$$P[T_{\max}^2 > t_\alpha^2] = \alpha. \tag{3}$$

しかし，$T_{\max}^2$ の正確な分布を導くことは難しく，完全データの場合でさえも上側パーセント点の導出には，非常に複雑な計算が必要であり，数表も十分には与えられていない．この問題の解決方法の

一つとしてボンフェローニ近似法が考えられる．k 標本すべての観測数が等しく，k 個の母集団が独立であると仮定した場合，(3) 式 は，以下のように書き換えることができる．

$$P[T^2_{\max} > t^2_\alpha] \approx \sum_{a<b} P[T^2_{ab} > t^2_{B,\alpha}] = \alpha.$$

さらに，すべての T^2_{ab} の分布は等しいので，ボンフェローニ近似法によるホテリング T^2 型検定統計量の上側 100α パーセント点は，次のように与えられる．

$$P[T^2_{12} > t^2_{B,\alpha'}] = \alpha'.$$

ただし，$\alpha' = 2\alpha/k(k-1)$ である．しかし，ボンフェローニ近似法は標本の数が大きい場合，保守的になり過ぎる傾向があり，また，正確な $t^2_{B,\alpha'}$ を求めることも難しく，シミュレーションによる導出が必要である．そこで, Seko, Kawasaki and Seo (2012) で提案した，2 標本問題における F 分布を用いた上側近似パーセント点の考え方を応用し，$t^2_{B,\alpha'}$ の近似値として F^* を提案した．

$$\begin{aligned} F^*_{\alpha'} &= T^2_{F,\alpha'} - \frac{(N^{(a)}+N^{(b)})p - (N^{(a)}_2+N^{(b)}_2)p_2}{(N^{(a)}+N^{(b)})p}\left(T^2_{F,\alpha'} - T^2_{T,\alpha'}\right) \\ &= cT^2_{F,\alpha'} + (1-c)T^2_{T,\alpha'}. \end{aligned}$$

ただし

$$\begin{aligned} &T^2_{F,\alpha'} = \frac{(n_1-k)p}{g_1}F_{\alpha';p,g_1}, \quad T^2_{T,\alpha'} = \frac{(n-k)p}{g}F_{\alpha';p,g}, \\ &n_1 = \sum_{i=1}^k N^{(i)}_1, \quad n_2 = \sum_{i=1}^k N^{(i)}_2, \quad n = n_1+n_2, \\ &c = \frac{(N^{(a)}_2+N^{(b)}_2)p_2}{(N^{(a)}+N^{(b)})p}, \quad g = n-k-p+1, \quad g_1 = n_1-k-p+1 \end{aligned}$$

である．最後に，モンテカルロ・シミュレーションにより $T^2_{\max}$ 統計量の上側パーセント点とボンフェローニ近似法による上側パーセント点を算出し，これらの値を真の値と考え，F^* 値の近似精度を評価した．

参考文献

[1] Chang, W. Y. and Richard, D. St. P. (2009). Finite-sample inference with monotone incomplete multivariate normal data I. *Journal of Multivariate Analysis*, **100**, 1883–1899.

[2] Krishnamoorthy, K. and Pannala, K. M. (1999). Confidence estimation of a normal mean vector with incomplete data. *The Canadian Journal of Statistics*, **27**, 395–407.

[3] Seko, N., Yamazaki, A. and Seo, T. (2012). Tests for mean vector with two-step monotone missing data. *to appear in SUT Journal of Mathematics*, **48**.

[4] Seko, N., Kawasaki, T. and Seo, T. (2012). Testing equality of two mean vectors with two-step monotone missing data. *to appear in American Journal of Mathematical and Management Sciences*, **31**.

[5] Siotani, M., Hayakawa, T. and Fujikoshi, Y. (1985). *Modern Multivariate Statistical Analysis : A Graduate Course and Handbook*, American Sciences Press, Ohio.

[6] Yu, J., Krishnamoorthy, K. and Pannala, K. M. (2006). Two-sample inference for normal mean vectors based on monotone missing data. *Journal of Multivariate Analysis*, **97**, 2162–2176.

Bayesian Inference for Stochastic Volatility Model with Spatial Correlation: Application to Regional Business Cycle in Japan

大塚 芳宏
北海道大学大学院経済学研究科

1 はじめに

本報告は，空間的相関を考慮した確率的ボラティリティ (Stochastic Volatility; SV) モデルを新たに提案し，日本の地域別景気循環の動態の推定を行った．景気循環の分析は，マクロ計量分野において重要な分野の一つである．なぜならば，景気循環・動向を推定することによって，政府は自国の経済に対して適切な経済政策を実行できることから国内外で盛んに実証研究が行われている．そして，景気自体は観測できない潜在変数であるために，状態空間モデルを用いて推定するほか，あるマクロ指数を景気循環の代理変数として推定する必要がある．こうした景気循環の分析は，国家単位の分析が中心であったが，近年においてデータの整備や計量モデルの改良が進み，地域単位の分析も行われてきている．地域別の景気循環を取り扱った先行研究の分析結果においては，国と地域の循環には乖離があることが指摘されており，国の景気を分析するだけでは適切な経済政策をとるには不十分であると指摘がなされている．また，各地域の経済活動は個々に独立しているのではなく，相互に依存しているという観点から地理情報を用いて空間的相互依存関係を導入した空間計量モデルによるアプローチも行われている．具体的には，Kakamu *et al*., (2010) や大塚 (2011) では日本の地域別景気循環には，スピルオーバー効果が観測され，空間計量モデルを用いた方が推定精度が高いことが報告されている．さらに，Falk and Sinabell (2009) では，欧州のデータを用いて，空間的相関がボラテリィティに内在しているとの指摘もされており，本報告では未だ検証が行われていないボラテリティにおける空間的相互関係の強さの推定を試みる．

ボラティリティを推定する計量モデルは主に GARCH 型モデルと SV モデルが挙げられる．とりわけ SV モデルはこれまで金融時系列分析において，リスク評価，ポートフォリオ選択，オプション価格の評価などに用いられてきた．しかし，Fernández-Villaverde and Rubio-Ramírez (2010) で近年のマクロ指数には分散不均一性があることが指摘されているほか，Primiceri (2005) や中島・渡部 (2012) では，可変パラメータ VAR モデルの分散構造に SV モデルを仮定し，政策分析に用いられているなど，マクロ計量分析の分野においても SV モデルは用いられてきていることから，本報告においても SV モデルを空間計量モデルへの拡張を試みる．

2 モデル

本報告で提案するモデルは以下のように定義される．

$$y_{it} = m_i + \psi_i(y_{i,t-1} - m_i) + \exp\left(\frac{1}{2}\alpha_{it}\right)\epsilon_{it},\ \epsilon_{it} \sim \mathcal{N}(0,1) \tag{1}$$

$$\alpha_{it} = \rho\sum_{j=1}^{n} w_{ij}\alpha_{jt} + h_{it}, \quad |\rho| < 1 \tag{2}$$

$$h_{i,t+1} = \mu_i + \phi_i(h_{it} - \mu_i) + \eta_{it},\ \eta_{it} \sim \mathcal{N}(0, \sigma_i^2), \tag{3}$$

(1) 式は，観測値 y_{it} を 1 次の自己回帰とボラティリティと呼ばれる非負の確率変数 $\exp(\alpha_{it}/2)$ と過去と独立な標準正規分布に従う確率変数 ϵ_{it} の積として表している. (2) 式は，ボラティリティに空間的自己回帰過程を仮定し，空間的相互相関を示す ρ と地理的情報を示すウェイト行列 w_{ij} を用いている．本報告では，ウェイト行列の構築には，1 次の隣接ダミーを用いている．これより個別要因と空間的要因にボラティリティの分解を行う．(3) は個別のボラティリティの 2 乗の対数値 h_{it} が 1 次の自己回帰モデルに従うと仮定している．ここで，ϕ_i が 1 に近いほどボラティリティに対するショックの持続性が高いことを示し，金融時系列分析においてはこうした高い持続性をボラティリティ・クラスタリングと呼んでいる．本報告では，地域別のマクロ指数においてもこうした高い持続性が見られるかどうかを検証する．ここで，ボラティリティを除くパラメータを $\mathbf{\Theta}$ と定義する．このとき，ボラティリティ $\mathbf{H}$ を条件とした尤度は以下のように書ける．

$$L(\mathbf{y}|\mathbf{\Theta},\mathbf{H}) = \prod_{t=1}^{T}\prod_{i=1}^{n} f(y_{it}|\mathbf{\Theta},\mathbf{H}_t) = \frac{1}{\sqrt{2\pi\tilde{\sigma}_{it}^2}}\exp\left(-\frac{e_{it}^2}{2\tilde{\sigma}_{it}^2}\right), \tag{4}$$

ただし，$\mathbf{y} = (\mathbf{y}_1,\ldots,\mathbf{y}_T)'$, $\mathbf{y}_t = (y_{1t},\ldots,y_{nt})$, $\mathbf{h} = (\mathbf{h}_1',\ldots,\mathbf{h}_T')'$, $\mathbf{h}_t = (h_{1t},\ldots,h_{nt})'$, $\mathbf{H}_t = diag(\exp(\alpha_{1t}),\ldots,\exp(\alpha_{nt}))$, $\mathbf{H} = (\mathbf{H}_1,\ldots,\mathbf{H}_T)$,

$$e_{it} = \begin{cases} y_{it} - m_i, & (t=1) \\ y_{it} - m_i - \psi_i(y_{i,t-1} - m_i), & (t>1) \end{cases}, \quad \tilde{\sigma}_{it}^2 = \begin{cases} \frac{\exp(\alpha_{it})}{1-\psi_i^2}, & (t=1) \\ \exp(\alpha_{it}), & (t>1) \end{cases}.$$

この尤度関数は，ボラティリティもパラメータであることから，尤度が解析的に解けないため，パラメータを最尤推定することは難しく，本報告では，ベイズ統計学の手法の一つであるマルコフ連鎖モンテカルロ (Markov chain Monte Carlo: MCMC) 法によって推定を行っていく．

そして，SV モデルの MCMC 法による推定は，$\mathbf{H}$ も条件付き事後分布からサンプリングする必要があり，$\mathbf{H}$ が標本の大きさ nT だけあることから，このサンプリングも効率的に行わないと，膨大な時間がかかり，事実上推定不可能になってしまう．そこで，本報告では，ボラティリティのサンプリングに Watanabe and Omori (2004) で提案されているマルチ・ムーブ・サンプラーを用いて行う．

3　実証分析と結果

実証分析においては，経済産業省が公表している月次の鉱工業生産指数（季節調整済）を地域の景気循環の代理変数として用いた．標本期間は 1998 年 1 月から 2012 年 5 月までとし，地域区分は経済産業省の区分に従い，北海道，東北，関東，北陸，中部，近畿，中国，四国，九州の 9 地域を使用した．また，ウェイト行列の構築は，Kakamu *et al.*(2008) で提案されている方法によって行っている．モデルの推定については，$40,000$ 回のサンプリングを行い，稼働検査期間として最初の $20,000$ 回を切り捨て，その後の $20,000$ 回のサンプルを事後分布からサンプリングされたものと見なして推定に用いた．

推定結果より，以下の結論が得られた．第一に，各地域のボラティリティの自己相関の強さを表すパラメータ ϕ の事後平均が，0.8 から 0.9 前後であったことから，鉱工業生産指数でとりわけ地域の指数においてもボラティリティ・クラスタリングが観測された．第二に，空間的相関を示すパラメータ ρ の事後平均が 95% 信用区間にゼロを含まず，0.3 と正の空間的相互作用が確認された．これより，日本のデータにおいても地域の景気変動は，周辺地域の景気状況を受けていることを示唆している結果が得られた．さらに，推定されたボラティリティの推移を検証した結果，2011 年 3 月に発生した東日本大震災において，震源地である東北地域に近い北海道，関東および中部地域のボラティリティが 2008 年に起きたリーマンショックの時よりも大きいことが明らかになった．

2 × 2 分割表の対称性検定におけるカイ二乗統計量の分布の エッジワース展開について

北海道教育大学　小倉 寛生　　北海道教育大学　関谷 祐里

2×2 分割表において，多項分布モデルを考える。すなわち，$\pi_{ij} = \Pr(A_i \cap B_j), 0 < \pi_{ij} < 1\ (i, j = 1, 2)$ であり，$\mathbf{Y} = (Y_{11}, Y_{12}, Y_{21}, Y_{22})' \sim \mathrm{Mult}_4(n; \boldsymbol{\pi})$ とする。ただし，$\boldsymbol{\pi} = (\pi_{11}, \pi_{12}, \pi_{21}, \pi_{22})'$ である。

対称性の帰無仮説 $H_0 : \pi_{12} = \pi_{21}$ のもとで，π_{ij} の最尤推定量を $\widetilde{\pi}_{ij}$ とおくと，

$$\widetilde{\pi}_{ij} = \frac{Y_{ij} + Y_{ji}}{2n} \quad (i, j = 1, 2)$$

となる。また，制約条件のない場合の π_{ij} の最尤推定量は，

$$\hat{\pi}_{ij} = \frac{Y_{ij}}{n} \quad (i, j = 1, 2)$$

となる。対称性の帰無仮説 $H_0 : \pi_{12} = \pi_{21}$ を対立仮説 $H_1 : \pi_{12} \neq \pi_{21}$ に対して検定する統計量として，カイ二乗統計量 T を用いる。T は，

$$T = n \sum_{i=1}^{2} \sum_{j=1}^{2} \frac{(\hat{\pi}_{ij} - \widetilde{\pi}_{ij})^2}{\widetilde{\pi}_{ij}}$$

として定義される。

対称性の帰無仮説 H_0 が正しいとき，n の値が大きくなると，T の分布が自由度 1 のカイ二乗分布に近づく。すなわち，

$$T \xrightarrow{L} \chi_1^2 \quad as \ \ n \to \infty$$

が成り立つことが知られている (Bowker[1])。

表 1: 2×2 分割表

A ＼ B	B_1	B_2	計
A_1	Y_{11}	Y_{12}	$Y_{1\cdot}$
A_2	Y_{21}	Y_{22}	$Y_{2\cdot}$
計	$Y_{\cdot 1}$	$Y_{\cdot 2}$	n

表 2: セル確率表

A ＼ B	B_1	B_2	計
A_1	π_{11}	π_{12}	$\pi_{1\cdot}$
A_2	π_{21}	π_{22}	$\pi_{2\cdot}$
計	$\pi_{\cdot 1}$	$\pi_{\cdot 2}$	1

以下では，極限分布に基づく近似よりもより精密な近似を求めるために，対称性の帰無仮説 H_0 のもとでの T の分布に対する漸近展開式を求めていく。ここからは，対称性の帰無仮説のもとでカイ二乗統計量の分布を考えるので，$\boldsymbol{Y} = (Y_{11}, Y_{12}, Y_{21}, Y_{22})' \sim \mathrm{Mult}_4(n; \boldsymbol{\pi}_0)$ とする。ただし，$\boldsymbol{\pi}_0 = (\pi_{11}, \pi_{12}, \pi_{12}, \pi_{22})'$ である。

確率変数 Y_{ij} から X_{ij} への変換

$$X_{ij} = \frac{Y_{ij} - n\pi_{ij}}{\sqrt{n}} \quad (i, j = 1, 2)$$

を行い，$\boldsymbol{X} = (X_{11}, X_{12}, X_{21})'$ とおく。ただし，$\pi_{21} = \pi_{12}$ とする。すると，T は $\boldsymbol{X}$ の関数として $T(\boldsymbol{X})$ として表すことができる。また，3 次元確率変数ベクトル $\boldsymbol{X}$ は，集合

$$L = \left\{ \boldsymbol{x} = (x_{11}, x_{12}, x_{21})' : \boldsymbol{x} = \frac{\boldsymbol{y} - n\boldsymbol{q}}{\sqrt{n}}, \boldsymbol{y} \in M \right\}$$

の値を取る格子点分布に従う。ここで，$\boldsymbol{q} = (\pi_{11}, \pi_{12}, \pi_{12})'$

$$M = \left\{ \boldsymbol{y} = (y_{11}, y_{12}, y_{21})' : y_{ij} \geq 0, y_{ij} \in \boldsymbol{Z}, y_{11} + y_{12} + y_{21} \leq n \right\}$$

である。ただし，$\boldsymbol{Z}$ は整数全体の集合である。

$\boldsymbol{x} = (\boldsymbol{y} - n\boldsymbol{q})/\sqrt{n}$ とすると，任意の $\boldsymbol{y} \in M$ に対して，帰無仮説 H_0 のもとでの局所エッジワース展開は，以下の形式で表現できる。

$$\Pr\{\boldsymbol{X} = \boldsymbol{x} | H_0\} = n^{-3/2}\phi(\boldsymbol{x}) \left\{ 1 + \frac{1}{\sqrt{n}} h_1(\boldsymbol{x}) + \frac{1}{n} h_2(\boldsymbol{x}) + O(n^{-3/2}) \right\}$$

$G = \{\boldsymbol{x} = (x_{11}, x_{12}, x_{21})' : T(\boldsymbol{x}) < \beta\}$ のとき，$\Pr(\boldsymbol{X} \in G)$ における漸近展開を得るためには，$G \cap L$ の範囲にあるすべての点の上での局所エッジワース展開の和が必要である。そこで，Yarnold は，格子点分布に従う確率変数が拡張された凸集合に属する確率の近似式を与えた (Yarnold[2])。その考えに基づくと，次のような近似式が考えられる。

$$\Pr(\boldsymbol{X} \in G) = J_1 + J_2 + J_3 + O(n^{-3/2})$$

ここで，J_1 項は，連続型分布の Edgeworth 展開とみなされる。一方，J_2 項は，X の不連続性を説明する項である。$J_2 = O(n^{-1/2})$, $J_3 = O(n^{-1})$ である。

本報告では，2×2 分割表における対称性の帰無仮説 H_0 のもとでの，局所エッジワース展開と J_1 の評価式を紹介した。

参考文献

[1] A. H. Bowker, A test for symmetry in contingency tables, *J. Am. Stat. Assoc.*, **43** (1948), 572-574.

[2] J. K. Yarnold, Asymptotic approximations for the probability that a sum of lattice random vectors lies in a convex set, *Ann. Math. Statist.*, **43** (1972), 1566-1580.

Asymptotic expansion for Hilbert space valued random variables and its application to symmetric statistics

Shuya Kanagawa

Department of Mathematics
Tokyo City University, Tokyo 158-8557, JAPAN

1 Introduction

Let $\{\xi_j, j \geq 1\}$ be i.i.d. real valued random variables with a distribution μ and $u(x_1, x_2, \cdots, x_m)$ be a symmetric function. A statistics with a kernel function u is called a symmetric statistics. We consider U-statistics and V-statistics which are typical examples of symmetric statistics.

Example 1.1 *(U-statistics with degree m)*

$$U_n := \frac{1}{\binom{n}{m}} \sum_{1 \leq i_1 < i_2 < \cdots < i_m \leq n} u(\xi_{i_1}, \xi_{i_2}, \cdots, \xi_{i_m})$$

Sample mean : $u(x) = x \quad (m = 1)$
Sample variance : $u(x_1, x_2) = \frac{1}{2}(x_1 - x_2)^2 \quad (m = 2)$

Example 1.2 *(V-statistics with degree m)*

$$V_n := n^{-m} \sum_{1 \leq i_1 \leq i_2 \leq \cdots \leq i_m \leq n} u(\xi_{i_1}, \xi_{i_2}, \cdots, \xi_{i_m})$$

Cramér-von Mises-Smirnov Statistics :

$$u(x_1, x_2) = \int_0^1 w(u) \left(I_{\{x_1 \leq u\}} - u\right) \left(I_{\{x_2 \leq u\}} - u\right) du \quad (m = 2)$$

and

$$V_n = \frac{1}{n-1} \int_0^1 w(x) (F_n(x) - 1)^2 dx,$$

where $w(x)$ is a weight function and $F_n(x)$ is a empirical distribution function for samples ξ_j, $1 \leq j \leq n$.

The kernel $u(x_1, x_2, \cdots, x_m)$ is called "degenerate" if for any $x_2, \cdots, x_m$

$$\int_{-\infty}^{\infty} u(y, x_2, \cdots, x_m) \mu(dy) = 0.$$

In this paper we improve the poof in [13] to show the Donsker-Varadhan large deviation principle for U-statistics and V-statistics with a degenerate kernel. The obtained result holds for not only continuous but also piecewise continuous kernel. Furthermore we consider an Edgeworth expansion of them. The main scheme used in the proofs is constructed from some limit theorems for sums of Hilbert space valued random variables and its application to a Fourier series expansion of kernel functions of these symmetric statistics.

2 Representations of U-statistics and V-statistics with degree 2

Put

$$U_n := \frac{2}{n(n-1)} \sum_{1 \le i < j \le n} u(\xi_i, \xi_j), \quad V_n := \frac{1}{n^2} \sum_{1 \le i,j \le n} u(\xi_i, \xi_j).$$

Then we have

$$nV_n = \frac{1}{n} \sum_{1 \le i,j \le n} u(\xi_i, \xi_j) = \frac{1}{n} \sum_{1 \le i,j \le n} \sum_{k=1}^{\infty} \lambda_k g_k(\xi_i) g_k(\xi_j) \tag{10}$$

$$= \frac{1}{n} \sum_{k=1}^{\infty} \lambda_k \sum_{1 \le i,j \le n} g_k(\xi_i) g_k(\xi_j) = \frac{1}{n} \sum_{k=1}^{\infty} \lambda_k \left\{ \sum_{i=1}^{n} g_k(\xi_i) \right\}^2$$

$$= h\left(\frac{1}{\sqrt{n}} \sum_{i=1}^{n} G_i \right).$$

Furthermore

$$\sqrt{n(n-1)} U_n = \frac{2}{\sqrt{n(n-1)}} \sum_{1 \le i < j \le n} u(\xi_i, \xi_j)$$

$$= \frac{2}{\sqrt{n(n-1)}} \sum_{1 \le i < j \le n} \sum_{k=1}^{\infty} \lambda_k g_k(\xi_i) g_k(\xi_j)$$

$$= \frac{1}{\sqrt{n(n-1)}} \sum_{k=1}^{\infty} \lambda_k \sum_{1 \le i < j \le n} 2 g_k(\xi_i) g_k(\xi_j)$$

$$= \frac{1}{\sqrt{n(n-1)}} \sum_{k=1}^{\infty} \lambda_k \left[\left\{ \sum_{i=1}^{n} g_k(\xi_i) \right\}^2 - \sum_{i=1}^{n} g_k^2(\xi_i) \right]$$

$$= \frac{n}{\sqrt{n(n-1)}} \sum_{k=1}^{\infty} \lambda_k \left[\left\{ \sum_{i=1}^{n} \frac{g_k(\xi_i)}{\sqrt{n}} \right\}^2 - \frac{1}{n} \sum_{i=1}^{n} \left(\frac{g_k(\xi_i)}{\sqrt{n}} \right)^2 \right]$$

$$= \frac{n}{\sqrt{n(n-1)}} \sum_{k=1}^{\infty} \lambda_k \left\{ \sum_{i=1}^{n} \frac{g_k(\xi_i)}{\sqrt{n}} \right\}^2 - \frac{1}{\sqrt{n(n-1)}} \sum_{k=1}^{\infty} \lambda_k \sum_{i=1}^{n} \frac{g_k^2(\xi_i)}{n}$$

$$= \frac{n}{\sqrt{n(n-1)}} h\left(\frac{1}{\sqrt{n}} \sum_{i=1}^{n} G_i \right) - \frac{1}{\sqrt{n(n-1)}} \sum_{k=1}^{\infty} \lambda_k \sum_{i=1}^{n} \frac{g_k^2(\xi_i)}{n}.$$

3 Edgeworth expansion of symmetric statistics

Combining Bogatyrev-Götze-Ulyanov (2006) and the representation of V-statistics defined in Section 2, we easily obtain an Edgeworth expansion of V-statistics.

Theorem 3.1 *Let $\{\xi_j, j \ge 1\}$ be sequence of i.i.d. random variables with a distribution μ. Assume all assumptions in Theorem 7.4 for $u(x,y)$. Then, under some technical conditions for G_1, we have*

$$\left| P\{n|V_n| \le r\} - P\left\{ \left| \sum_{k=1}^{\infty} \lambda_k \left(Z_k^2 - 1 \right) + E[u(\xi_1, \xi_1)] \right| \le r \right\} - A_1(r) \right| = O\left(\frac{1}{n} \right),$$

where $A_1(r)$ is the first term in an Edgeworth expansion of $P\{n|V_n| \le r\}$ and $Z_1, Z_2, \cdots$ are i.i.d. N(0,1) random variables. We also easily show the similar result for U-statistics.

２変量極値分布のノンパラメトリック推定

鈴川晶夫 (北海道大学大学院経済学研究科)

1. はじめに

２次元確率変数ベクトル (X,Y) の周辺分布はいずれも平均１の指数分布であり, 同時生存関数 $S(x,y)=\Pr(X>x,Y>y)$ が, ２変量極値分布

$$S(x,y)=\exp\left\{-(x+y)A\left(\frac{y}{x+y}\right)\right\}$$

で与えられる場合について考えた. ここに, 関数 $A:[0,1]\to[1/2,1]$ は条件 $A(0)=A(1)=1$ と $\max(1-t,t)\le A(t)\le 1$ $(\forall t\in[0,1])$ を満たす凸関数であり, Pickands (1981) の従属性関数とよばれる. 本報告において, この従属性関数のノンパラメトリック推定問題について議論した.

Pickands (1981) によって一つの推定量が提案され, その改良としていくつかの推定量が提案された（(Deheuvels 1991, Hall and Tajvidi 2000 など). これらの推定量を Pickands 型推定量とよぶ. また, Capéraà, Fougères and Genest (1997) は, Pickands 型とは異なる型の推定量 (CFG 型推定量) を提案し, 数値実験によって Pickands 型推定量と CFG 型推定量を比較した. その実験結果は, CFG 型推定量が Pickands 型より望ましいことを示している.

Segers (2008) は, Pickands 型推定量と CFG 型推定量を統一的に扱う表現を与え, ２変量の独立性の下で CFG 型推定量は Pickands 型推定量に比べて漸近的により有効であることを示した. 本報告において, Pickands 型推定量と CFG 型推定量の両方を含む推定量のクラスを定式化し, そのクラスに属する推定量の漸近的性質について議論した.

2. Pickands 従属性関数の推定量

同時生存関数 $S(x,y)$ からの無作為標本を (X_i,Y_i), $i=1,2,\ldots,n$ とする. 任意の $t\in[0,1]$ に対して, $\xi_i(t)=\min\left(\frac{X_i}{1-t},\frac{Y_i}{t}\right)$, $i=1,2,\ldots,n$ とおく. このとき, Pickands(1981) 推定量は

$$1/\hat{A}^{\mathrm{P}}(t)=n^{-1}\sum_{i=1}^{n}\xi_i(t)=n^{-1}\sum_{i=1}^{n}\min\left(\frac{X_i}{1-t},\frac{Y_i}{t}\right)$$

により定義される. これを改良した Deheuvels (1991) 推定量は

$$1/\hat{A}^{\mathrm{D}}(t)=n^{-1}\sum_{i=1}^{n}\{\xi_i(t)-(1-t)(X_i-1)-t(Y_i-1)\}$$

により定義される. また, Capéraà, Fougères, and Genest (CFG) (1997) 推定量は

$$-\log\hat{A}^{\mathrm{CFG}}(t)=n^{-1}\sum_{i=1}^{n}[\log\xi_i(t)-p(t)\log X_i-\{1-p(t)\}\log Y_i]$$

により定義される (Beirlant *et al.* 2004, Segers 2008). ただし, $p(t)$ は区間 $[0,1]$ 上の適当な重み関数である.

本報告において, $\xi_i(t)$ を Box and Cox (1964) の冪変換

$$\varphi_\lambda(x) = \lambda^{-1}(x^\lambda - 1) \quad (\lambda > 0), \qquad \varphi_0(x) = \log x$$

で変換したとき, その期待値に関する等式

$$\begin{aligned} \Gamma(1+\lambda)\varphi_\lambda\{1/A(t)\} &= E[\varphi_\lambda\{\xi_i(t)\} - a(t)\varphi_\lambda(X_i) - b(t)\varphi_\lambda(Y_i)] \\ &\quad -\lambda^{-1}\{\Gamma(1+\lambda)-1\}\{1-a(t)-b(t)\} \end{aligned}$$

が成り立つことに着目し, Pickands 従属性関数に対する新たな推定量を提案した. ただし, Γ はガンマ関数であり, $a(t)$ と $b(t)$ は $[0,1]$ 上の適当な重み関数である. その推定量 $\hat{A}_\lambda(t; a(t), b(t))$ は

$$\begin{aligned} \varphi_\lambda\{1/\hat{A}_\lambda(t; a(t), b(t))\} &= \frac{1}{n\Gamma(1+\lambda)} \sum_{i=1}^{n} [\varphi_\lambda\{\xi_i(t)\} - a(t)\varphi_\lambda(X_i) \\ &\quad -b(t)\ \varphi_\lambda(Y_i)] - c_\lambda\{1-a(t)-b(t)\} \end{aligned}$$

により定義される.

推定量 $\hat{A}_\lambda(t; a(t), b(t))$ の漸近的性質（一致性, 漸近正規性など）を調べた. その漸近分散を最小化するという意味においての最良な重み関数の選択についても議論した. また, Marshall and Olkin (1967) の 2 変量指数分布モデルのもとで漸近分散を評価する公式を与え, その公式に基づいて推定量の漸近比較を行った.

参考文献

[1] Beirlant, J., Goegebeur, Y., Segers, J. and Teugels, J. (2004). *Statistics of Extremes, Theory and Applications.* Wiley, Chichester.

[2] Box, G. E. P. and Cox, D. R. (1964). *Journal of the Royal Statistical Society, Series B*, **26**, 211-252.

[3] Capéraà, P., Fougères, A. L. and Genest, C. (1997). *Biometrika*, **84**, 567-577.

[4] Deheuvels, P. (1991). *Statistics & Probability Letters*, **12**, 429-439.

[5] Hall, P. and Tajvidi, N. (2000). Bernoulli, **6**, 835-844.

[6] Marshall, A. W. and Olkin, I. (1967). *Journal of the American Statistical Association*, **62**, 30-44.

[7] Pickands, J. (1981). *Proceedings of the 43rd Session of the ISI*, Buenos Aires, **49**, 859-878.

[8] Segers, J. (2008). *Topics in Extreme Values* (M. Ahsanullah and S. N. U. A. Kirmani, Eds.), 185-207. Nova Science Publishers, New York.

多変量密度関数の対称性に関する分解

東京理科大学大学院理工学研究科　生亀清貴[1]
東京理科大学理工学部　田畑耕治
東京理科大学理工学部　富澤貞男

行と列が同じ分類からなる正方分割表において, Caussinus (1965) は定理「対称モデルが成り立つための必要十分条件は, 周辺同等モデルと準対称モデルの両方が成り立つことである」を与えた. Bhapkar and Darroch (1990) は Caussinus (1965) の定理を多元分割表に拡張した. 一方, 連続型の 2 変量確率密度関数に対して同様の定理が Tomizawa, Seo and Minaguchi (1996) によって与えられている.

本講演では連続型の多変量確率密度関数に対して準対称性と周辺対称性を定義し, 確率密度関数に関して対称性を準対称性と周辺対称性に分解した.

ここでは 3 変量の場合について記述する. X_1, X_2, X_3 を連続型確率変数とし, その結合密度を $f(x_1, x_2, x_3)$ とする. また (1,2,3) の任意の並べ替えを (π_1, π_2, π_3) とする. $f(x_1, x_2, x_3)$ の対称性は次のように定義される：すべての $(x_1, x_2, x_3) \in R^3$ に対して,

$$f(x_{\pi_1}, x_{\pi_2}, x_{\pi_3}) = f(x_1, x_2, x_3)$$

が成り立つ. 詳細については Tong (1990) を参照されたい. また $f(x_1, x_2, x_3)$ の 1 次周辺対称性を次のように定義した：すべての $t \in R$ に対して,

$$f_{X_1}(t) = f_{X_2}(t) = f_{X_3}(t)$$

が成り立つ. さらに 2 次周辺対称性を次のように定義した：すべての $(s, t) \in R^2$ に対して,

$$f_{X_1X_2}(s, t) = f_{X_1X_2}(t, s) = f_{X_1X_3}(s, t) = f_{X_2X_3}(s, t)$$

が成り立つ. すなわち 2 次周辺対称性はそれぞれの周辺確率密度関数 (X_1, X_2), (X_1, X_3), (X_2, X_3) が対称かつ同等であることを表している. 密度関数が対称性を満たせば 2 次周辺対称性を, また 2 次周辺対称性を満たせば 1 次周辺対称性をそれぞれ満たす.

ここで密度関数 $f(x_1, x_2, x_3)$ の台を K^3 とする. ただし

$$K^3 = \{(x_1, x_2, x_3); f(x_1, x_2, x_3) > 0, a < x_i < b, i = 1, 2, 3, -\infty \leq a < b \leq \infty\}.$$

[1]〒 278-8510 千葉県野田市山崎 2641
e-mail: kiyotaka_iki@ybb.ne.jp

一般に確率密度関数は

$$f(x_1,x_2,x_3)=\mu\alpha_1(x_1)\alpha_2(x_2)\alpha_3(x_3)\beta_{12}(x_1,x_2)\beta_{13}(x_1,x_3)\beta_{23}(x_2,x_3)\gamma(x_1,x_2,x_3),\ (1)$$

のように表すことが可能である. ただし $(x_1,x_2,x_3)\in K^3$, また任意の $c\in(a,b)$ について

$$\alpha_1(c)=1,\quad \beta_{12}(c,x_2)=\beta_{12}(x_1,c)=1,\quad \gamma(c,x_2,x_3)=\gamma(x_1,c,x_3)=\gamma(x_1,x_2,c)=1,$$

$\alpha_2,\alpha_3,\beta_{13},\beta_{23}$ についても同様. このとき, 確率密度関数 $f(x_1,x_2,x_3)$ が対称的であるとは, 式 (1) に次の制約を課すことと同値である：

$$\begin{cases}\alpha_1(x_1)=\alpha_2(x_1)=\alpha_3(x_1),\\ \beta_{12}(x_1,x_2)=\beta_{12}(x_2,x_1)=\beta_{13}(x_1,x_2)=\beta_{23}(x_1,x_2),\\ \gamma(x_{\pi_1},x_{\pi_2},x_{\pi_3})=\gamma(x_1,x_2,x_3).\end{cases}$$

また 1 次準対称性を次のように定義した：式 (1) に対して,

$$\begin{cases}\beta_{12}(x_1,x_2)=\beta_{12}(x_2,x_1)=\beta_{13}(x_1,x_2)=\beta_{23}(x_1,x_2),\\ \gamma(x_{\pi_1},x_{\pi_2},x_{\pi_3})=\gamma(x_1,x_2,x_3).\end{cases}$$

さらに 2 次準対称性を次のように定義した：式 (1) に対して,

$$\gamma(x_{\pi_1},x_{\pi_2},x_{\pi_3})=\gamma(x_1,x_2,x_3).$$

このとき, 次の定理を得た.

定理 1： $k=1,2$ に対して, 3 変量確率密度関数 $f(x_1,x_2,x_3)$ が対称的であるための必要十分条件は, $f(x_1,x_2,x_3)$ が k 次周辺対称性と k 次準対称性の両方を満たすことである.

参考文献

Bhapkar, V. P., and Darroch, J. N. (1990). Marginal symmetry and quasi symmetry of general order. *Journal of Multivariate Analysis*, **34**, 173-184.

Caussinus, H. (1965). Contribution à l'analyse statistique des tableaux de corrélation. *Annales de la Faculté des Sciences de l'Université de Toulouse*, **29**, 77-182.

Tomizawa, S., Seo, T., and Minaguchi, J. (1996). Decomposition of bivariate symmetric density function. *Calcutta Statistical Association Bulletin*, **46**, 129-133. *Journal de la Société Française de Statistique*, **148**, 3-36.

Tong, Y. L. (1990). *The Multivariate Normal Distribution.* New York: Springer-Verlag.

正方分割表における累積確率に基づく非対称モデルと分解

山本　紘司　(大阪大学 医学部)
富澤　貞男　(東京理科大学 理工学部)

行と列が順序のある同じ分類からなる $r \times r$ 正方分割表において，(i, j) セル確率を p_{ij} とする $(i = 1, \ldots, r; j = 1, \ldots, r)$．正方分割表解析においては分類間の関連性が強いため独立性が成り立たない場合が多く，代わって対称性に関するモデルを用いることがある．セル確率の対称性を示す対称モデルは次のように定義される (Bowker, 1948):

$$p_{ij} = p_{ji} \quad (i \neq j).$$

また，対称モデルは累積確率 $\{G_{ij}\}$ を用いて次のようにも表される:

$$G_{ij} = G_{ji} \quad (i \neq j).$$

ここに，G_{ij} は

$$G_{ij} = \sum_{s=1}^{i} \sum_{t=j}^{r} p_{st} \quad (i < j), \quad G_{ij} = \sum_{s=i}^{r} \sum_{t=1}^{j} p_{st} \quad (i > j)$$

で定義される．また，m パラメータ周辺同等 (MH(m)) モデルは次のように定義される (Tahata and Tomizawa, 2008): 固定した m（$m = 1, \ldots, r-1$）に対して

$$\frac{G_{i,i+1}}{G_{i+1,i}} = \Delta_i^{(m)} \quad (i = 1, \ldots, r-1) \quad \text{ただし} \quad \Delta_i^{(m)} = \prod_{k=0}^{m-1} \psi_k^{i^k}.$$

特に MH(1) モデルは拡張周辺同等 (EMH) モデル (Tomizawa, 1993) である．さらに，累積 2 比パラメータ対称 (C2RPS) モデルおよび累積拡張準対称 (CEQS) モデルはそれぞれ次のように定義される (Tomizawa et al., 2007):

$$\frac{G_{ij}}{G_{ji}} = \Gamma \Theta^{j-i} \quad (i < j);$$

$$\frac{G_{ij}}{G_{ji}} = \gamma \frac{\gamma_j}{\gamma_i} \quad (i < j),$$

ただし $\gamma_1 = 1$ とする．ここに C2RPS モデルは CEQS モデルの特別な場合である．特に C2RPS モデルにおいて $\Gamma = 1$，CEQS モデルにおいて $\gamma = 1$ とおいたモデルはそれぞれ累積線形対角パラメータ対称 (CLDPS) モデルおよび累積準対称 (CQS) モデルである (Miyamoto et al., 2004)．

Yamamoto et al. (2011) は次の分解定理を与えた:

定理 1: C2RPS モデルが成り立つための必要十分条件は，CEQS モデルと EMH モデルの両方が成り立つことである．

系 1: CLDPS モデルが成り立つための必要十分条件は，CQS モデルと EMH モデルの両方が成り立つことである．

本講演では，C2RPS (CLDPS) モデルを一般化したモデルを提案し，さらに定理 1 を一般化する．

C2RPS モデルを一般化したモデルとして次の C2RPS(m) モデルを提案する: 固定した m ($m = 1, \ldots, r-1$) に対して，

$$\frac{G_{ij}}{G_{ji}} = \Gamma\Omega_{ij}^{(m)} \quad (i < j) \quad \text{ただし} \quad \Omega_{ij}^{(m)} = \prod_{t=1}^{m} \Theta_t^{j^t - i^t}.$$

特に C2RPS(1) モデルは C2RPS モデルである．また，C2RPS(m) モデル ($m = 1, \ldots, r-1$) において，$\Gamma = 1$ とおいたモデルを CLDPS(m) モデルと呼ぶことにする．特に CLDPS(1) モデルは CLDPS モデルである．さらに C2RPS($r-1$) モデルおよび CLDPS($r-1$) モデルはそれぞれ CEQS モデルおよび CQS モデルである．

このとき次の分解定理を得る:

定理 2: 固定した m ($m = 1, \ldots, r-1$) に対して，C2RPS(m) モデルが成り立つための必要十分条件は，CEQS モデルと MH(m) モデルの両方が成り立つことである．

系 2: 固定した m ($m = 1, \ldots, r-1$) に対して，CLDPS(m) モデルが成り立つための必要十分条件は，CQS モデルと MH(m) モデルの両方が成り立つことである．

参考文献

Bowker, A. H. (1948). *Journal of the American Statistical Association*, **43**, 572-574.

Miyamoto, N., Ohtsuka, W. and Tomizawa, S. (2004). *Biometrical Journal*, **46**, 664-674.

Tahata, K. and Tomizawa, S. (2008). *Advances in Data Analysis and Classification*, **2**, 295-311.

Tomizawa, S. (1993). *Biometrics*, **49**, 883-887.

Tomizawa, S., Miyamoto, N., Yamamoto, K. and Sugiyama, A. (2007). *Statistica Neerlandica*, **61**, 273-283.

Yamamoto, K., Ando, S. and Tomizawa, S. (2011). *Journal of Statistics: Advances in Theory and Applications*, **5**, 1-13.

分割表において有用な幾つかの多重比較法の応用について

明星大学　広津千尋

1. 序論

1元配置において多様な多重比較法が研究され，応用されているのに比して，2元表における交互作用の多重比較法は極めて立ち遅れているように思われる．一つには，1元配置の場合のように自由度1まで分解した交互作用成分の多重比較は効率が悪く，解釈も明解でないためであろう．分割表の行を単位とした Scheffe 型の多重比較法は例えば広津(1977)や Hirotsu(1983)で提案されているが，Gilula(1986, JASA)や，Greenacre(1988, J. Classification)の correspondence analysis も本質的に同じ統計量を提案している．統計量としては，2節で述べる行間の二乗距離がまず定義され，それは直に幾つかの行をプールして出来る2群間の二乗距離に拡張された．Hirotsu(1983)ではその有意性を保守的に評価するための参照分布として，Wishart 行列の最大根の分布を導いており，Greenacre(1988)でもそれが応用されている．Gilula(1986)は参照分布として通常の適合度 χ^2 を用いており，より保守的である．Hirotsu(1991, 2009)ではさらに3群以上の場合に群間の一般化二乗距離を導入することにより，参照分布として Wishart 行列の最大根の分布を用いることの保守性を解消している．

列の水準に自然な順序がある場合には，上記の統計量は累積 χ^2 を基礎とする統計量に自然に拡張される．その場合，参照分布は斜交 Wishart 行列の最大根の分布となり扱いが困難になるが、この場合に有用な χ^2 近似を提案することが出来る．行の水準に自然な順序がある場合は行の全ての順列を考えることは意味をなさず，Scheffe 型多重比較法は適用出来ない．そこで，変化点型対比に限って適用することを提案する．列の多重比較法は、行のそれと対称に考えればよいが，行，列の両方に自然な順序がある場合は，両方に変化点型対比を適用する興味ある統計量を提案することが出来る．

2. 群間の一般化二乗距離

$a\times b$ 分割表の (i,j) セルの頻度を y_{ij}，行和を R_i，列和を C_j，総和を N で表す．$\boldsymbol{r}=N^{-1/2}(\sqrt{R_1},\cdots,\sqrt{R_a})'$，　$\boldsymbol{c}=N^{-1/2}(\sqrt{C_1},\cdots,\sqrt{C_b})'$ として，R' および C' をそれぞれ $\boldsymbol{r}$ および $\boldsymbol{c}$ に直交する $(a-1)\times a$ および $(b-1)\times b$ の直交行列として定義する(Hirotsu, 1983)．さらに $z_{ij}=y_{ij}\Big/\sqrt{R_iC_j/N}$ と変換し，z_{ij} を添字に関して辞書式に並べたベクトルを $\boldsymbol{z}$ とする．ここで，一般性を失うことなく行の m 群を，$G_k=\{q_1+\cdots+q_{k-1}+1,\cdots,q_1+\cdots+q_k\}$，$k=1,\cdots,m$ とする．このとき，群間の一般化二乗距離を

$$\chi^2(G_1;\cdots;G_m)=\max\left\|(\boldsymbol{\gamma}'\otimes C')\boldsymbol{z}\right\|^2 \qquad (1)$$

で定義する．ただし，max は $\boldsymbol{\gamma}=(\gamma_1,\cdots,\gamma_a)'$ に関する次の条件下での最大化を意味する，

$$\boldsymbol{\gamma}'\boldsymbol{r}=0,\quad \|\boldsymbol{\gamma}\|=1,\quad \gamma_i\equiv\lambda_k(R_i/T_k)^{1/2},\quad i\in G_k,\quad T_k=G_k\text{の行和計},\quad k=1,\cdots,m.$$

ここで，Y_{kj} を行をプールした $m\times b$ 2元表の要素とすると，(1)式は若干の計算により，$\boldsymbol{w}_k=(T_k/N)^{-1/2}C'(C_1^{-1/2}Y_{k1},\cdots,C_b^{-1/2}Y_{kb})'$ として，$\sum\boldsymbol{w}_k\boldsymbol{w}_k'$ の最大固有値と一致する．これは特別な場合として、2行間や2群間の二乗距離を含む一般化となっている．ここで，参照分布は Wishart 行列 $W(I_{\min(a-1,b-1)},\max(a-1,b-1))$ の最大根の分布であるが，これは(1)式において各群がそれぞれ1行から成る場合に相当する．すなわち，1 一般化二乗距離では Wishart 行列の最大根による評価の保守性が解消されていることが分かる．

3. 列に自然な順序があり，累積 χ^2 統計量を基礎とする場合

前節の統計量(1)で，単に C' の各行を変化点型基準化対比に書き換えた $C^{*'}$ (Hirotsu, 1983 参照)を用いればよい．$a\geq b$ の場合に参照分布は，漸近正規性の仮定の下に斜交 Wishart 行列

$W(C^{*'}C^{*}, a-1)$の最大根の分布として与えられる. 例えば, C_jが揃っている時には, $C^{*'}C^{*}$の最大根の第２根に対する比が３となることから, $\rho_{(1)}\chi^2_{a-1}$による近似が最大根の分布の大変良い近似を与える(Hirotsu, 1991). ただし, $\rho_{(1)}$は$C^{*'}C^{*}$の最大根である. C_jが不揃いの場合の近似精度については当日報告する.

4. 自然な順序を考慮した列の多重比較

3節の設定で、列に変化点型多重比較法を適用する. この場合の基本統計量は列に b - 1 通りの分点を入れ, それぞれその両側の列をプールして出来る(b – 1)通りの a x 2 分割表の適合度 χ 二乗の最大が適切である. この帰無仮説の下での漸近分布は, 相関のある χ 二乗統計量の最大であり, 分布論は広津, 栗木(1990)に与えられている.

5. 行、列の両方に自然な順序がある場合の max max χ 法

応用場面としては、測定値が順序応答である場合の用量反応解析がある. 行、列にそれぞれ(a – 1), (b – 1)通りの分点を考え, 分点の前後をプールして出来る(a – 1) x (b – 1)通りの 2 x 2 分割表の適合度 χ 二乗の最大を基本統計量とする。これには, 分割表に潜在するマルコフ性を利用した正確な確率計算法が提案出来るので, 当日紹介する.

6. 結語

本論では 2 次元分割表のみについて述べたが, 3 元表への拡張も試みられている(例えば, Hirotsu et al. , 2001). さらに, 順序に応じた傾向性仮説の拡張として, 最近, 凸性仮説およびスロープ変化点仮説を対象とした 2 重累積和統計量に基づく方法について若干成果が得られている.

参考文献

広津千尋 (1977). 分割表における多重比較と水準の群分け. 日本品質管理学会誌 7, 2, 27-33.

Hirotsu, C. (1983). Defining the pattern of association in two-way contingency tables. Biometrika 70, 579-590, 1983.

広津千尋, 栗木哲 (1990). 累積カイ二乗の最大成分に基づく多重比較－交互作用の場合－. 応用統計学会誌 19, 115-132, 1990.

Hirotsu, C. (1991). An approach to comparing treatments based on repeated measures. Biometrika 78, 583-594, 1991.

Hirotsu, C., Aoki, S., Inada, T. and Kitao, Y. (2001). An exact test for the association between the disease and alleles at highly polymorphic loci with particular interest in the haplotype analysis. Biometrics 57, 769-778.

Hirotsu, C. (2009). Clustering rows and/or columns of a two-way contingency table and a related distribution theory. Computational Statistics and Data Analysis 53, 4508-4515.

一般化されたレーマン対立仮説モデルを用いた市場モデルと CAPM と Jensen のアルファ、そして順位推定

三浦良造　一橋大学名誉教授

一期間の金利を r とする。ポートフォリオの投資収益率から r を差し引いたものを被説明変数とし、(時価総額ウエイトの) 株価指数から r を差し引いたものを説明変数とする単純線形回帰モデルをファイナンス理論では、市場モデルと呼ぶ。$Y=\alpha+\beta x+\varepsilon$.

ここでは、市場モデルの偶然誤差項を一般化されたレーマン対立仮説モデルを用いて表わす。観測が時間経過と共に行われるが、偶然誤差項を独立で同一分布に従う場合と同一分布に従うが独立ではなく弱い依存性を持つ場合を考える。$G(x:\theta)=h(F(\cdot):\theta)$

まず回帰係数の推定を順位統計量を用いて行う。すでに知られているように、回帰係数の推定を順位統計量から導く場合は、偶然誤差項 ε がどのような確率分布に従っていてもよい：ゼロの周りで対称でなくてもよいし、期待値がゼロでなくてもよい。良い、と云うのは、推定量が定義できるし、偶然誤差項を独立で同一分布に従う場合は、推定量の漸近正規性も示すことができる。これは Jureckova(1971)で示されている。本報告では、偶然誤差項が一般化されたレーマン対立仮説モデルに従う場合にも、同様のこと、つまり回帰係数 β の推定に加えて、切片 α の推定、さらに　一般化されたレーマン対立仮説モデルの変換パラメター θ の推定量の漸近正規性が成立することを説明する。上記論文の結果と Tsukahara&Miura(1993)の結果を組み合わせることでこれが可能である。

次に、観測における偶然誤差項 $\varepsilon_1, \varepsilon_2, \cdots, \varepsilon_n$ が、弱い依存性を持つ場合に、上記 iid の場合と同様の結果、つまり、β, α, θ の推定量の漸近正規性が成立することを論じる。厳密な数学的証明は、まだ不完全なところが残るが Shao&Yu(1996)と Louhichi(2000)の結果を援用して、数学的証明の大枠を説明する。

ファイナンス理論における CAPM では、上記の市場モデルの切片はゼロである。しかし、その CAPM 理論の実証研究では、それがゼロであるという帰無仮説がしばしば棄却される。そのようにして検出されるゼロでない切片値を、ファイナンス理論では、Jensen のアルファと呼んでいる。ここで扱う単純線形回帰モデルの偶然誤差項 ε が非対称分布に従うことに起因して、偶然誤差項 ε の期待値はゼロではない。それを $m(\theta)$ と書くことにすると、$(\alpha+m(\theta))$ が Jensen のアルファに該当すると考えられる。統計モデル上で表現されるこの量 $m(\theta)$ は関数 $(\cdot)$ に θ の推定値を代入することにより得られる。$m(\theta)$ は θ の値に応じて、ゼロまたは正負の値をとる。偶然誤差が従う、本来の確率分布 F が変換の関数 h(•:

θ)により、ゆがみ(skew)を与えられ、それにより正あるいは負である m(θ)がもたらされると解釈できると考える。ファイナンス理論では、市場モデルの偶然誤差項がどのような確率分布に従うかを明示しているわけではないが、情報効率的な市場では、これがゼロの周りで対称な確率分布に従うと想定する、あるいは少なくとも期待値がゼロであると想定していると思われる。

このようなパラメター推定は、順位統計量を用いることにより可能である。ファイナンス理論の実証分析に於いては、最小二乗法が使われることがほとんどではないかと思われるが、順位統計量を用いる推定によりこのような議論展開が可能であることを指摘しておくことは、ファイナンス理論の実証研究だけでなく、実務における“ベータの推定”とポートフォリオの銘柄選択、期待リターンの計測に於いても重要であると考える。さらに、βの推定を利用するヘッジポートフォリオの議論にもこのモデルが有効に活きると考える。

参考文献

[1] Jureckova, J. (1971). “Nonparametric estimate of regression coefficients.” Annals of Mathematical Statistics. Vol. 42. 1328-1338.

[2] Miura R. (1985). “Hodges–Lehmann type estimators and Generalized Lehmann’s Alternatives.” A special lecture (in Japanes) at Annual Meeting of Japan Mathematical Associations. April 1985.

[3]. Miura R. , D. Yokouchi and Y. Aoki (2009). “A Note on Statistical Models for Individual Hedge Fund Returns.” Math.Meth. Oper. Res. 69. 553-577.

[4].Tsukahara, H. & Miura, R.(1993). “One sample estimation for generalized Lehmann’s alternative models.” Statistica Sinica. Vol.3. 83-101.

[5] Louhichi S.(2000) “Weak convergence for empirical processes of associated sequences. “Ann.Inst.Henri Poincare. Probabilites et Statistiques. 36 (2000), 5, 547-567.

[5] Shao Q.M. and Yu H. (1996) “Weak Convergences for Weighted Empirical Processes of Dependent Sequences.” Annals of Probability. 24. 2098-2127

[6] 三浦良造(1985)。“順位統計量に基づくベータの推定” 経営研究(大阪市立大学)。

個別企業の信用価格スプレッドと倒産確率の導出

刈屋武昭・山村能郎・乾孝治
（明治大学グローバル・ビジネス研究科）
王竹（ZWシステム）

問題

- 信用リスクの基礎情報とモデル
- 国債価格とイールドの問題
- 価格ｽﾌﾟﾚｯﾄﾞvsイールドスプレッド
- 格付けと業種：カテゴリカル同質集団？
- 信用リスク価格ｽﾌﾟﾚｯﾄﾞの構造
- 倒産確率の期間構造

信用リスクの基礎情報問題

Forward-Looking vs Backward-Looking

Backward-Looking：過去の金利関係データ,過去データに基づくモデル

－金利：過去時系列ﾃﾞｰﾀからの構造分析

－信用：倒産・非倒産の過去ﾃﾞｰﾀによる統計分析、推移確率、ﾛｼﾞｯﾄ･ﾌﾟﾛﾋﾞｯﾄ、判別分析

Forward-Looking：現在時点の市場データ:投資家の,経済や金融のリスク・リターンに関する将来ﾊﾟｰｽﾍﾟｸﾃｨﾌﾞの情報を内包ー現行経済・金融や企業の将来動向への関心　ｸﾛｽｾｸｼｮﾝ･ﾃﾞｰﾀの位置づけー投資家は過去の時系列ﾃﾞｰﾀのもとに将来を予想

－金利：現在時点の国債価格、スワップレート

－信用：現在時点の社債、CDS等

属性依存型国債価格モデル

現在 t=0 で将来CF発生時点

$$s_{g1} < s_{g2} < \cdots < s_{gM(g)} \quad (g = 1, \cdots, G)$$

$C_g(s)$; CF関数　＝0 unless $s=s_{gj}$　　$s_{aM(a)} = \max_g s_{gM(g)}$

$D_g(s)$; 属性依存型確率的割引関数　$0 < s \le s_{aM(a)}$

$$P_g = \sum_{j=1}^{M(g)} C_g(s_{gj}) D_g(s_{gj}) \qquad (g = 1, \cdots, G)$$

P_g　realization　$\Longleftrightarrow$　$\{D_g(s) : 0 \le s \le s_{gM(g)}\}$

$$D_g(s) = \bar{D}_g(s) + \Delta_g(s)$$

$$P_g = \sum_{m=1}^{M(g)} C_g(s_{gm}) \bar{D}_g(s_{gm}) + \eta_g \qquad \eta_g = \tilde{C}_g \tilde{\Delta}_g$$

$$\underset{\sim}{y} = X\underset{\sim}{\beta} + \underset{\sim}{\eta} \qquad Cov(\underset{\sim}{\eta}) = (Cov(P_g, P_h)) \equiv \sigma^2 \Phi(\rho, \xi)$$

属性依存型平均割引関数への多項式近似

(w_1, w_2, w_3)　　$z_{g1=1}$,　z_{g2} : maturity,　z_{g3} : coupon

M0 : (1,0,0); basic model with no attributes
M1 : (1,1,0): M0 + maturity effect,
M2 : (1,0,1); M0 + coupon effect
M3 :(1,1,1);　M0 +maturity effect＋coupon effect

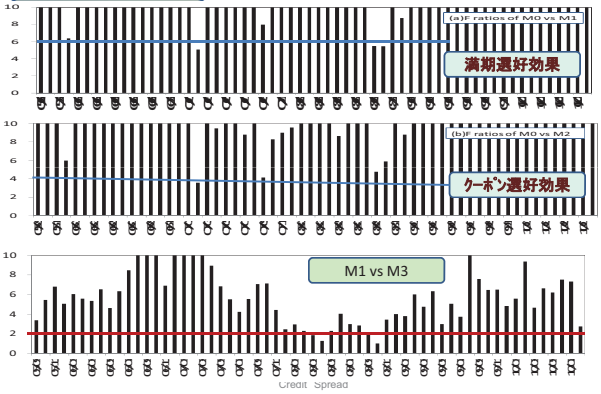
$$\bar{D}_g(s) = 1 + (\delta_{11} w_1 z_{g1} + \delta_{12} w_2 z_{g2} + \delta_{13} w_3 z_{g3}) s + \cdots + (\delta_{p1} w_1 z_{g1} + \delta_{p2} w_2 z_{g2} + \delta_{p3} w_3 z_{g3}) s^p$$

$$\bar{D}_g(s_{gj}) = E[D_g(s_{gj})] = E[\exp(-\int_0^{s_{gj}} f_{gs} ds)]$$

$$R_s = -\frac{1}{s} \log \bar{D}(s)$$　Term structure of interest rates

No属性仮説 F比検定 ψ

F ratio =[{$QSR(0) - QSR(1)$}/#]/[$QSR(1)/df$],
[{$QSR(0) - QSR(1)$}/#] > 2[$QSR(1)/df$].



社債信用価格ｽﾌﾟﾚｯﾄﾞ

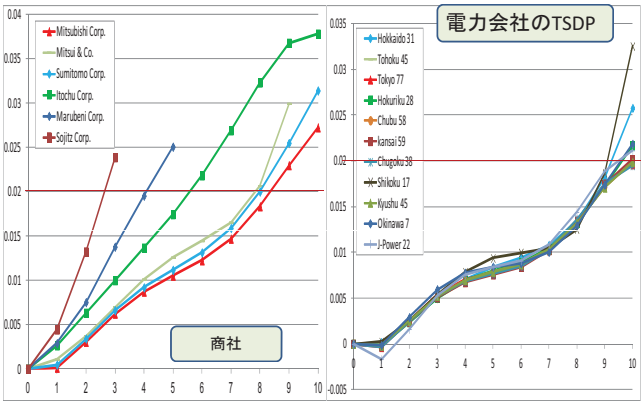
- 個別社債の信用価格ｽﾌﾟﾚｯﾄﾞ
  ＝社債価格－社債の無リスク理論価格
  ＝社債価格－同属性(ｸｰﾎﾟﾝ、満期)国債同等価格

$$y_k^{(i)} = V_k - \hat{P}_k^{(i)}, \quad \hat{P}_k^{(i)} = \sum_{j=1}^{M(k)} C_k(s_{kj}) \overline{D}_k^{(i)}(s_{kj})$$

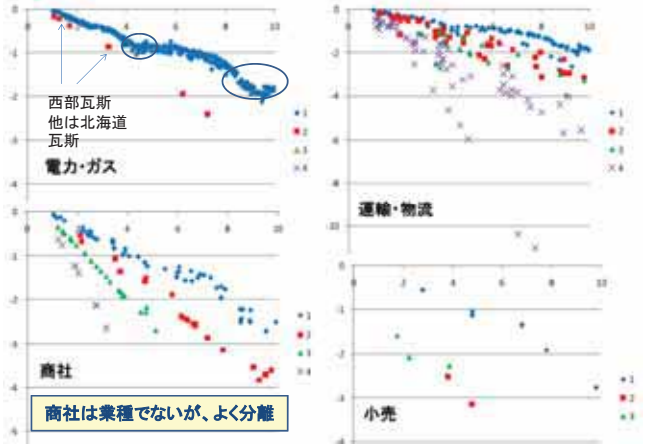
$(i = 0,3)$ MOｽﾌﾟﾚｯﾄﾞ vs M3ｽﾌﾟﾚｯﾄﾞ

- 基準化信用価格スプレッド(不確実性の基準化)

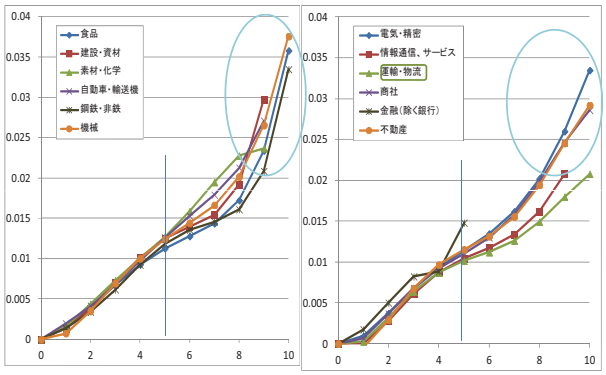
$$\varsigma_k^{(3)} = y_k^{(3)} / s_{kM(k)}$$



M3モデルの業種別・格付け別価格ｽﾌﾟﾚｯﾄﾞ



CG1×業種ｶﾃｺﾞﾘｰのディフォルト期間構造



CB 価格モデル

$$V_k = \sum_{j=1}^{M(k)} \overline{C}_k(s_{kj}) D_k(s_{kj})$$

$$D_k(s) = \overline{D}_k(s) + \Delta_k(s)$$

$$p_k(s : i(k)) \equiv \sum_{j=1}^{J} w_k(j) p(s : i(k), j)$$

格付i と(純粋)業種jのDefault確率関数

$$w_k(j) \geq 0, \quad \sum_{j=1}^{J} w_k(j) = 1$$

投資家の将来CFへの期待

$$\overline{C}_k(s_{mj}) = C_k(s_{mj})[1 - p_k(s_{mj} : i(k))] + 100\gamma(i(k))[p_k(s_{mj} : i(k)) - p_k(s_{mj-1} : i(k))]\chi_k(s_{mj})$$

$$p(s : i, j) = \alpha_1^{ij} s + \alpha_2^{ij} s^2 + \cdots + \alpha_q^{ij} s^q$$

格付i と(純粋)業種jのDefault確率関数

要約

- 信用ﾘｽｸ価格ｽﾌﾟﾚｯﾄﾞ測度を、国債市場の投資家行動の検証に基づいて定義
- ｲｰﾙﾄﾞｱﾌﾟﾛｰﾁに対して、価格ｽﾌﾟﾚｯﾄﾞｱﾌﾟﾛｰﾁの情報的価値の優位性を議論
- 基準化信用ﾘｽｸ価格ｽﾌﾟﾚｯﾄﾞ測度を定義し、信用ﾘｽｸ同質的ｸﾞﾙｰﾌﾟの基礎とした
- ｸﾗｽﾀｰ分析による実証的ｸﾞﾙｰﾌﾟに基づく信用ﾘｽｸ分析の有効性を示した
- 格付・業種ｶﾃｺﾞﾘｰと価格ｽﾌﾟﾚｯﾄﾞの関係を検証
- ﾃﾞｨﾌｫﾙﾄ確率の期間構造をｸﾗｽﾀｰｸﾞﾙｰﾌﾟに対して導出した。Kariya(2012) ﾓﾃﾞﾙ

CAPM における回帰分析と最適ポートフォリオの構成

上智大学大学院　理工学研究科
理工学専攻　数学領域　加藤剛研究室
山方　亮

1. CAPM

CAPM とは Capital Asset Pricing Model の略で、資本資産評価モデルと呼ばれるものである。このモデルは、リターンの値を理論的に求めるためにしばしば利用される。

今回は、CAPM におけるベータと呼ばれるリスク指標の考え方に着目し、リスクの程度を指定したときに、ポートフォリオを構成する銘柄の最適配分を与える重みを決定する方法を研究した。ここで、ベータとは個別の銘柄や各種ポートフォリオなどのリスク評価に用いられる指標で、線形単回帰モデルの言葉に置き換えれば、説明変数の回帰係数に該当する。

2. リスク感応度のモデル

2.1 リターンの定義

X_t を時刻 t における金融商品（株式、債権など）の価格とするとき、リターン R を次の式で定義する。

$$R = \frac{X_t}{X_{t-1}} - 1 \qquad (1)$$

今回の解析では、X_t として株価の日次データにおける 1 日の終値を採用した。

2.2 個別銘柄のリスク感応度

n 企業の株式を対象にするとき、各企業の株式のリターンを(1)にもとづいて定義し、

R_i: 第 i 銘柄のリターン (i=1,2, …, n)

と記す。また、Q_m を市場ポートフォリオのリターンとする。今回の解析では、Q_m として TOPIX を用いた。

Q_m と各 R_i について、ε_i を誤差項として

$$R_i = \beta_{0i} + \beta_{1i} Q_m + \varepsilon_i \qquad (2)$$

という線形モデルを考える。これは、R_i の TOPIX に対するリスク感応度のモデルと解釈できる。回帰係数 β_{1i} が第 i 銘柄の TOPIX に対するリスク感応度になる。

以下の議論では、リスクフリーレートを 0 として、切片項 β_{0i} を消去しておく。変数を改めて R_i、Q_m と記すことにすれば、(2)のモデルは

$$R_i = \beta_i Q_m + \varepsilon_i \qquad (3)$$

と書き換えられる。R_i と Q_m の実際のデータにもとづいて最小 2 乗法によって β_i の推定を行い、不偏推定量 $\widehat{\beta_i}$ を求める。これにより、R_i の Q_m に対するリスク感応度を表すモデル式

$$R_i = \widehat{\beta_i} Q_m \qquad (4)$$

が決まる。

2.3 ポートフォリオのリスク感応度

リスク感応度の概念は、個別銘柄だけでなく、ポートフォリオに対しても用いることができる。

Q_p：任意のポートフォリオのリターン

$\{w_i\}$ ポートフォリオを構成する各銘柄の重み

とおくと、Q_p は

$$Q_p = \sum_{i=1}^{n} w_i R_i \quad (w_i \geq 0,\ \sum_{i=1}^{n} w_i = 1) \qquad (5)$$

と定義される。そして、Q_p の Q_m に対するリスク感応度は、(4)と同様に

$$Q_p = \beta Q_m \qquad (6)$$

の形で表せる。

3. 最適ポートフォリオの構成

各 R_i のもとになる n 銘柄は固定する。(6)の式で事前に β を指定すると、それは、Q_p がどの程度 Q_m に依存してよいかを表すリスク感応度と解釈できる。

そこで、

β_D: 自分が設定したいリスク感応度

とおくと、(4)、(5)、(6)から

$$\sum_{i=1}^{n} w_i R_i = \sum_{i=1}^{n} (w_i \widehat{\beta_i}) Q_m = \beta_D Q_m \qquad (7)$$

という関係式が得られる。(7)にもとづいて各銘柄の重み $\{w_i\}$ を決定すれば、市場ポートフォリオ（今回は TOPIX）に対して事前に設定したリスク感応度 β_D に適応した最適なポートフォリオを構成することができる。

4. 重みの推定

最適なポートフォリオの重みを求めるための定式化を行う。s をデータ数とし、以下のように記号を定める。

$$\mathbf{R}_n = \begin{pmatrix} R_{11} & \cdots & R_{1n} \\ \vdots & \ddots & \vdots \\ R_{s1} & \cdots & R_{sn} \end{pmatrix} \qquad (8)$$

$$\mathbf{w} = (w_1, \cdots, w_n)' \qquad (9)$$

$$\boldsymbol{\beta} = (\widehat{\beta_1}, \cdots, \widehat{\beta_n})' \qquad (10)$$

$$\mathbf{Q}_m = (Q_{m1}, \cdots, Q_{ms})' \qquad (11)$$

ここで、R_{ki} は第 i 銘柄の第 k 日におけるリターン、Q_{mk} は市場ポートフォリオの第 k 日におけるリターンであ

る。(8)から(11)を利用すると、(7)の式を s 日分考えることにより、

$$\mathbf{R}_n\mathbf{w} = \mathbf{Q}_m\boldsymbol{\beta}'\mathbf{w} = \mathbf{Q}_m\beta_D \quad (12)$$

という表現を得る。したがって、

$$\boldsymbol{\beta}'\mathbf{w} = \beta_D \quad (13)$$

$$\mathbf{R}_n\mathbf{w} = \mathbf{Q}_m\boldsymbol{\beta}'\mathbf{w} \quad (14)$$

をみたす $\mathbf{w}$ が理論的な解となる。

ところが、リスク感応度の式(4)は理論的なモデル式であって、実際のR_{ki}とQ_{mk}のデータを代入したときは、e_{ki}を誤差項として、

$$R_{ki} = \hat{\beta}_i Q_{mk} + e_{ki} \quad (15)$$

という形になる。そこで、

$$\mathbf{e} = \begin{pmatrix} e_{11} & \cdots & e_{1n} \\ \vdots & \ddots & \vdots \\ e_{s1} & \cdots & e_{sn} \end{pmatrix} \quad (16)$$

とおいて、(15)の関係を(14)へ組み込むと、

$$\mathbf{R}_n\mathbf{w} = \mathbf{Q}_m\boldsymbol{\beta}'\mathbf{w} + \mathbf{ew} \quad (17)$$

を得る。$\mathbf{ew}$ はデータにもとづく重み付き誤差ベクトルなので、2 乗誤差

$$\|\mathbf{ew}\|^2 = \mathbf{w}'(\mathbf{R}_n - \mathbf{Q}_m\boldsymbol{\beta}')'(\mathbf{R}_n - \mathbf{Q}_m\boldsymbol{\beta}')\mathbf{w} \quad (18)$$

を最小にする $\mathbf{w}$ が望ましい。

以上のことから、$\mathbf{w}$ に対する制約条件を考慮に入れると、s 日分のデータが得られたとき、問題は次の形に帰着される。

Minimize

$$\mathbf{w}'(\mathbf{R}_n - \mathbf{Q}_m\boldsymbol{\beta}')'(\mathbf{R}_n - \mathbf{Q}_m\boldsymbol{\beta}')\mathbf{w} \quad (19)$$

Constraints

$$w_1, w_2, \ldots, w_n \geq 0 \quad (20)$$

$$\sum_{i=1}^{n} w_i = 1 \quad (21)$$

$$\boldsymbol{\beta}'\mathbf{w} = \sum_{i=1}^{n} w_i\hat{\beta}_i = \beta_D \quad (22)$$

実際には、(20)、(21)、(22)の制約条件のもとで(19)を最小にする$\mathbf{w}^* = \{w_i^*\}$を、数理計画法のアルゴリズムによって求める。

5. 銘柄の選択

5.1 一般的な方法

ポートフォリオは、異なる分野や業種、異なる要因により株価が変化する銘柄で構成する。同じ分野、業種、変動要因の銘柄を入れてしまうと、それらが同じような動きをするので、リスク分散の観点から意味がない。

また n の個数については、最低でも 3、最高で 10 くらいが望ましい。

5.2 データ解析のための銘柄選択

今回は n=10 として計算を行った。$\mathbf{w}$ の成分に対する条件 (20) により$w_i = 0$となるものが出ることも想定して、最大の 10 として銘柄数の確保を図った。

ポートフォリオを求めるにあたっては、幅広い銘柄から 5 パターンのポートフォリオを作成した。最終的にそれらのポートフォリオの銘柄とそれらに対する最適な重みを決定した。

実際に解析に用いたデータは、各銘柄の日毎の終値を用いて、期間としては、2012 年 1 月 1 日から 2012 年 6 月 30 日までを用いた。

6. TOPIX との比較

結果の検討として、TOPIX との連動性等の検証を行った。また、比較として TOPIX 連動型上場投資信託を用いた。最初に各ポートフォリオ、TOPIX、連動商品について単位根検定を行いランダム運動でないことを確認した。

次にそれぞれに関して自己相関関数を求める。最後に、各ポートフォリオと TOPIX との相互相関関数を求めた。

大規模データを対象にした推定問題における高速処理

上智大学 理工学部 情報理工学科　加藤 剛
（共同研究：テンソル・コンサルティング株式会社　藤本浩司）

1　背景

1.1　異なり数の推定問題

インターネット取引のデータをもとにした異なり数推定問題を例にして，大規模データの高速処理が要求されている状況を報告する．例として取り上げる異なり数推定問題を数学的な表現で整理すると，次のようになる．

定義　有限なデータセットのデータ項目（変量）に対し，そのデータ項目が取る相異なる値の数を，データ項目の異なり数という．

問題　有限な母集団から無作為抽出された標本が得られたとする．母集団の大きさと標本データの基本的な統計量は既知であると仮定して母集団の異なり数を推定することを，異なり数の推定問題という．

【例】標本サイズが 20 のデータ項目 $\{\mathrm{A},\mathrm{A},\mathrm{A},\mathrm{A},\mathrm{A},\mathrm{A},\mathrm{A},\mathrm{A},\mathrm{B},\mathrm{B},\mathrm{B},\mathrm{B},\mathrm{B},\mathrm{C},\mathrm{C},\mathrm{C},\mathrm{D},\mathrm{D},\mathrm{E},\mathrm{F}\}$ において，異なる値は A，B，C，D，E，F．したがって，異なり数は 6.

1.2　対数圧縮二項確率行列

参考文献 [2] において，母集団のサイズが 10^7 に達するような大規模なものであるときに異なり数を適切かつ効率的に推定する方法として，対数圧縮二項確率行列という概念を利用した方法が提案されている．その要点を簡潔に述べると，次のようになる．

母集団において度数 N をもつ要素から k 個が抽出される確率が二項確率であることをもとにして，二項確率行列 P を考える．

$$P=(P_{ij})_{1\le i,j\le N},\quad P_{ij}=\begin{cases}\binom{i}{j}r^j(1-r)^{i-j} & :\ i\ge j\\ 0 & :\ i<j\end{cases},\quad 0<r<1$$

理論的には，P そのものを使っても異なり数の推定は可能である．しかし，商業的利用を踏まえたときに N が 10^7 を超えるような大規模なものになることや，逆行列の計算を伴うことなどの理由から，その理論を計算機上に実装して推定作業を実現することは難しい．そこで，2 の指数で行を区切り，2 の指数を抽出率で割った数字で列を区切って P をいくつかの部分行列に分割し，各部分行列内の確率値の重み付き平均を成分とする新たな行列 LP を作る．LP のことを対数圧縮二項確率行列とよぶ．すなわち，行列 P から行列 LP への変換で情報の圧縮をする．そして，LP にもとづいて異なり数の推定を行う．

参考文献 [2] では，1000 万件の購買履歴データにもとづく異なり数の推定や，312 万世帯の標本から全国の世帯における姓の異なり数の推定を行った事例を挙げて，対数圧縮二項確率行列を利用した方法の有用性を論じている．

2　商業利用に際しての問題と高速近似計算法の必要性

2.1　商業利用に際しての問題点

対数圧縮二項確率行列を利用する異なり数の推定方法は有用であることが [2] によって示された．しかし，この方法の適用範囲を大きく拡張することを考えたとき，新たな問題が生じる．

例えば参考文献 [1] で採り上げている問題を考えると，母集団のサイズは億の単位に達する．一般に，インターネット・モールにおける総アクセス数を母集団として，無作為抽出標本から IP アドレスの異なり数を推定する問題では，母集団のサイズが 10 億（$=10^9$）にまで達することはごく普通である．し

たがって，二項確率行列 P のサイズが 10^9 次かそれ以上という大規模なものになるため，P を分割して作る部分行列の成分について重み付き平均を計算をするときに逐一足し算を行うと計算量が膨れあがり，対数圧縮二項確率行列 LP の生成に非常に長い時間を要してしまう．インターネット・モールで，特にセキュリティ監視の観点からアクセス履歴のある IP アドレスの異なり数を利用したいときなどは，標本の入手から異なり数推定結果算出までがほぼリアルタイムで行えなければ意味がない．ここに，大規模な二項確率行列 P から対数圧縮二項確率行列 LP を生成する際に適用できる高速近似計算法の開発の必要性が生まれる．

2.2 高速近似計算法の開発

高速近似計算法の開発にあたって要求される条件は，少なくとも次の 3 つある．

1. P のサイズが 10^9 次程度の大規模なものであっても，ほぼリアルタイムとみなせる時間内で LP を生成できること
2. 新聞広告で通販されているような普及型コンピュータに，高くても 1 万円ほどで購入できるメモリを増設した程度の性能で，1 の計算時間を達成できること
3. 商業目的において十分耐えうる精度の近似値を与えること

諸般の事情により式の提示は割愛するが，非常に多くの項の重み付き和を積分近似する方法を利用して，上の 3 条件をみたす高速近似計算法を厳密な数学の理論として導き出すことに成功した．その理論をソフトウェア上に実装して計算を行ったところ，次のような，きわめて有用な結果が得られた．

- 環境：Intel Core i5-2500K 3.30GHz の CPU，16GB のメモリ，Windows 7 (64bit)，Mathematica 8.0.4.0
- 実験：$r = 0.5$ とした $2^{30} \approx 1073741824$ 次 (> 10 億次) の二項確率行列 P から LP を生成
- 計算時間：約 7.3 秒（同じ P に対して逐一足し算をして LP を求めると，日の単位の時間を要する）
- 近似精度：小数第 3 位まで真値と一致

3 大規模データ解析における高速計算の要求

インターネット取引を例に考える．スマートフォンやタブレット型コンピュータの普及，および，通信速度の大幅な向上に伴い，インターネットを介した商取引は，急速に増加しつつある．インターネット取引におけるデータ解析において要求されることは，大きく分けて 2 つある．1 つは，10 億かそれ以上の大規模データに対して，処理がほぼリアルタイムでこなせること．もう 1 つは，処理の対象となる問題に対して，十分な精度を保証できることである．けれども，現場からの要求は，後者に比べて前者が圧倒的に多い．どんなに精密かつ正確な処理であっても，ほぼリアルタイムで処理できる速さを伴わなければ評価されない．

伝統的な統計学では，母集団からの限られた標本データをもとにして，母集団に対する統計的推測（点推定，区間推定，検定など）を行ってきた．ところが，例えば東京証券取引所が始めた指数高速配信サービスでは，TOPIX 等の株価指数が 1/1000 秒間隔で配信される．したがって，データの大きさはむしろ手に余るほどであり，しかも，データはコンマ何秒で更新される．このような大規模データには，伝統的な統計学の枠組みでは十分に対処することができない．

研究の実社会への貢献が求められている現在，統計学に携わる者（特にこれからの若い人）は，すべての人ではないにせよ,「高速性」という要求を念頭において研究を進める必要があると考える．2.2 節で紹介した結果は，たまたまではあるが，現場の要求に適切に応えることができた例となっている．

参考文献

[1] 石橋圭介 他．異なり数上位 N ホストの推定および異常検出への応用．電子情報通信学会技術研究報告 ネットワークシステム 106(14)，pp.53-56，2006.

[2] 藤本浩司，加藤 剛．大規模データにおける対数圧縮 2 項確率行列を用いた母集団の異なり数の推定．日本計算機統学会第 26 回大会論文集，2012.

[3] 加藤 剛，藤本浩司．大規模データにおける異なり数推定のための対数圧縮 2 項確率行列に対する高速近似計算．日本計算機統学会第 26 回大会論文集，2012.

回遊性魚類の行動予測における隠れマルコフスイッチング構造の効果

東京大学　甫喜本 司, 水産総合研究センター・国際水産資源研究所　清藤 秀理, 魚崎 浩司,
慶應義塾大学　清水 邦夫

1. はじめに

まぐろ類等の高度回遊性魚類は，海中を生息域としているために自然下における行動を直接観察することは難しく，陸，空に生息する生物と比較して行動の理解は遅れているのが現状である．10年程前から，水深，腹腔内温度，海水温，照度を計測できるアーカイバルタグやポップアップタグの開発が進み，取得された計測値に基づく行動研究が発展している．魚類の行動を解析する方法については1950年頃から生態学を中心として議論されており, 初期の分析においてはCorrelated Random Walk (CRW) モデル (例えば, Skellan (1951)) とよばれる確率過程モデルがよく検討されてきた．一方, 近年のタグに基づく計測技術の普及に伴って, 計測により得られた魚類の行動データの時系列構造を推測するための統計モデルの検討が行われるようになった．最も検討されているクラスは状態空間型のモデル (例えば, Josen et al. (2007)) であるが, 異なる対象魚の行動を分類するためにスイッチング構造をもつ隠れマルコフモデルの有効性についても近年報告されている (Patterson et al. (2009))．しかし, 海洋物理要因の変化がこうした行動に与える影響については, 現段階で明らかになっていない．そこで本研究では, 海洋物理要因を外生要因として考慮に入れることにより対象魚の行動を分析するためのスイッチング型モデルを開発をしながら, この点について検討を行う．

2. マルコフスイッチングモデルに基づく回遊行動の記述

本研究では, 2002年4月18日から2003年2月25日までの期間に実施したビンナガ (*Thunnus alalnga*) の標識放流調査で得られた日次の位置データを使用して検討を行った．アーカイバルタグを取り付けて計測したビンナガの位置データに基づいてその予測を精度よく行うモデルの検討を行うこと, 及び海洋環境要因の一つとして海中の植物プランクトン濃度の計測値を外生要因として考慮に入れることにより, 位置予測をより効果的に行うためのモデルを考えることが具体的な目標である．

位置データより, 前日からの移動距離 (step length) の変化 $\{V_t\}$ と仰角の変化 (turning angle)$\{\phi_t\}$ を定義し, その時間的な相関の構造について予備調査を行ったところ, 月毎の自己相関の特徴がいくつかに分類されること, 及び植物プランクトン濃度の変化との相互相関が負の相関を与える傾向があることが観察された．この点を手がかりとして, 植物プランクトン濃度を外生変数としながら確率構造の変化を念頭においたマルコフスイッチングモデルを検討する．$\{V_t\}$ と $\{\phi_t\}$ のそれぞれに対して以下の形のスイッチングモデルを考える．

$$Z_t = \mu_{S_t} + \sigma_{S_t} X_t, \quad \sigma_{S_t} > 0$$

S_t は時点 t における状態 ($S_t \in \{1, \ldots, N\}$) であり, μ_{S_t} と σ_{S_t} は時点 t における状態と共にランダムに変化する平均と標準偏差のパラメータをそれぞれ表している．$\{X_t\}$ は $\{Z_t\}$ に影響を与える潜在的なプロセス (観測不能) を示し, 外生変量である植物プランクトン濃度 $\{C_t\}$ の影響の有無について以下の2つの仮説をおく．

$$X_t = \rho X_{t-1} + \delta_t, \quad |\rho_t| < 1 \tag{1}$$

$$X_t = \rho_{t-1} X_{t-1} + \delta_t, \quad |\rho_t| < 1, \quad \rho_t = \rho / C_t \tag{2}$$

ρ は未知係数, δ_t は攪乱項で, $\{V_t\}$ の場合は正規分布を, $\{\phi_t\}$ の場合には円周上の von Mises 分布に従う確率変数を仮定する. 仮説 (1) は $\{C_t\}$ の変化が $\{V_t\}$ や $\{\phi_t\}$ へ影響を与えないことを, 仮説 (2) は有意義な影響を与えることを示している.

3. 結果

$\{V_t\}$ と $\{\phi_t\}$ に関する観測値のヒストグラムを調べたところ, 双方共に 3 つの峰があることが観察されたため, 状態数 N を 3 として上記のモデルをあてはめて分析を行った. 分析結果は以下のように要約される.

(i) $\{V_t\}$ と $\{\phi_t\}$ の双方において, 東進時 (観測開始から 10 月半ばまで) における変化の過程は発生確率が最も高い状態が頻繁に変化するのに対し, 西進時 (10 月半ばから翌年 2 月まで) においては状態の変化が発生しない. また, 両者共に東進時における $\{X_t\}$ の自己回帰係数 ρ は相対的に低く, 西進時において高くなる. この結果より, ビンナガは遭遇する海域における植物プランクトン濃度の状態に影響を受け, 索餌を行うための日和見的行動と移動を主体とした遊泳行動とを繰り返しているのではないかという仮説を与えることができる.

(ii) 構造変化を考慮に入れたスイッチングモデルの有用性を統計的に評価するため, $\{V_t\}$ と $\{\phi_t\}$ のそれぞれについて, ARIMA モデル, 時変係数型 AR モデル, 及び植物プランクトン濃度を考慮した階差型の多変量自己回帰モデル, の線形非定常な時系列モデルを用いて予測を行い, その精度を上記のモデルと比較した. その結果, スイッチングモデルが平均 2 乗予測誤差 (MSE) の観点より線形非定常なモデルのクラスに比べて効果があると評価された. また, この結果はビンナガの経度方向と緯度方向の位置予測においてもそのまま反映され, 位置予測にロバストな効果を与えることが示された.

参考文献

[1] Box, G.E.P., Jenkins, G.M. (1976). Time Series Analysis, Forecasting and Control (revised edition), Holden-Day, San Francisco.

[2] Brockwell, P.J., Davis, R.A. (1996). Introduction to Time Series and Forecasting, Springer-Verlag, New York.

[3] Buckle, R.A., Haugh, D., Thomson, P. (2002). Growth and volatility regime switching models for New Zealand GDP data, Working paper, New Zealand Treasury.

[4] Ekstrom, P.A. (2004). An advance in geolocation by light, Mem. Natl Inst. Polar Res. Spec. Issue, 58, 210-226.

[5] Jammalamadaka, S.R. and SenGupta, A. (2001). Topics in Circular Statistics, World Scientific.

[6] Josen, I.D., Myers, R.A., James, M.C. (2007). Identifying leatherback turtle foraging behavior from satellite telemetry using a switching state-space model, Marine Ecology-Progress Series, 337, 255-264.

[7] Patterson, T.A., Basson, M., Bravington, M.V., Gunn, J.S. (2009). Classifying movement behaviour in relation to environmental conditions using hidden Markov models, Journal of Animal Ecology, 78, 1113-1123.

[8] Skellam, J.G. (1951). Random Dispersal in theoretical populations, Biometrika, Vol.38, No. 1/2, 196-218.

[9] Smith, P., Goodman, D. (1986). Determining fish movement from an “archival” tag: precision of geographical positions made from a time series of swimming temperature and depth, NOAA Tech. Memo, NOAA-TM-NMFS-SWFC-60.

[10] Turchin, P. (1991). Translating foraging movements in heterogeneous environments into the spatial distribution of foragers, Ecology, 72, 1253-1266.

[11] Zucchini, W., MacDonald, I.L. (2009). Hidden Markov Models for Time Series, An Introduction Using R, Chapman & Hall/CRC, Boca Raton.