(12)「統計的モデリングと推測に関する新展開」に関する研究報告

定常ガウス型時系列の情報量規準

東京大学・工　岩崎秀夫　　東京大学・工　駒木文保

1　はじめに

定常 AR モデル等の次数選択には, 赤池情報量規準 (Akaike[1]) が用いられてきた. 一方, 竹内情報量規準 (竹内 [6]) を時間領域で構成することは, 時系列に相関があるために困難である.

栗木, 甘利 [5] は, 竹内情報量規準の補正項を全確率分布族に埋め込まれたモデルの m−平均曲率ベクトルに関する量としてとらえている. ガウス型時系列のスペクトル密度の空間に関する微分幾何学的構造は確率分布族の幾何学の自然な拡張として Amari[2] で導入している. Komaki[3] では, 形式的な実座標系を利用することにより, AR(p) モデルの計量や接続等を陽に求め, 微分幾何学的考察を通してスペクトルを推定する問題について考察している. また, 駒木 [4] では, ARMA(p,q) モデルの計量や接続を陽に求めている.

本報告では, 定常ガウス型時系列のスペクトル表現を利用して, 竹内情報量規準を構成し, AR モデル $\mathcal{S}=\{S(\lambda;\theta)\}$ をすべてのスペクトル密度の空間 $\mathcal{A}$ の部分多様体と見なしたときの, モデルの m−平均曲率ベクトルを利用して, AR モデルの竹内情報量規準の補正項を導出する.

2　i.i.d. における竹内情報量規準の幾何学的解釈

竹内情報量規準の補正項は, 幾何学的に解釈できることが, 栗木, 甘利 [5] でしめされている.

統計モデル $\mathcal{P}:=\{p(x|\theta)|\theta\in\Theta\}$ を考える. $\mathcal{P}$ のもとでの最尤推定量を $\hat{u}$ とする. さらに, $F(x|\theta)$ を $p(x|\theta)$ の分布関数とする. このとき, 竹内情報量規準の補正項をテンソル表記でかくと, $I_{ab}(\hat{G})J^{ab}(\hat{G})$ となる. ただし, $\hat{G}$ は, 真の分布 G を推定した分布であり, I_{ab}, J_{ab} は$\partial_a:=\partial/\partial\theta^a$ として, 以下の式で定義される. J^{ab} は, J_{ab} の逆行列である.

$$I_{ab}(\hat{G})=\int\partial_a\log p(x|\hat{\theta})\partial_b\log p(x|\hat{\theta})d\hat{G}(x),\qquad J_{ab}(\hat{G})=-\int\partial_a\partial_b\log p(x|\hat{\theta})d\hat{G}(x).$$

以上の式を, $\hat{G}(x)-F(x|\hat{\theta})$ の 2 次以降を無視して計算することにより,

$$I_{ab}J^{ab}=m+\int h(x;\theta)\frac{1}{p(x|\hat{\theta})}d\{\hat{G}(x)-F(x|\hat{\theta})\}$$

となる. m はモデルのパラメータ数, $h(x;\theta)$ はモデルの m−平均曲率ベクトルである.

3　形式的実座標を利用した AR モデルの計量と接続の表現

Komaki[3] で与えられた, AR モデルにおける微分幾何学的量を用いて, スペクトル密度の空間における竹内情報量規準の表現をみる.

観測値 $x_1,\ldots,x_N$ が定常 AR(p) 過程 $L_a(B)x_t=\epsilon_t\quad L_a(z):=1+a_1z+\ldots+a_pz^p$ から得られた仮定する. B は後向きシフトオペレータ, $\{\epsilon_t\}$ は平均 0, 分散σ^2 のガウス白色雑音である. 定常性より, 方程式 $\tilde{L}_a(z)=z^p+a_1z^{p-1}+\cdots+a_p=0$ の根の絶対値は 1 より小さい. AR 過程のスペクトル密度は $z:=\exp(i\lambda)$ とおいて

$$S(\lambda;a_1,\ldots,a_p,\sigma^2)=\frac{\sigma^2}{2\pi}\frac{1}{L_a(z)L_a(z^{-1})}=\frac{\sigma^2}{2\pi}\frac{1}{\tilde{L}_a(z)\tilde{L}_a(z^{-1})}$$

で与えられる.

方程式 $\tilde{L}_a(z)=0$ の根を $z_1, z_2, \ldots, z_q, z_{q+1}, z_{q+2}, \ldots, z_{2q}, z_{2q+1}, \ldots, z_{2q+r}(2q+r=p)$ とおく. $(z_1, \ldots, z_p)$ を座標として採用し, 形式的複素微分 $\partial/\partial z := 1/2(\partial/\partial x - i\partial/\partial y)$ を用いる. 以下 $\theta^0 := \sigma^2, \theta^i := z_i(i=1,\ldots,p)$ とおき, 添字 $I, J, K, \cdots$ は $0, 1, \ldots, p$ の範囲を, 添字 $i, j, k, \cdots$ は $1, 2, \ldots, p$ の範囲を動くものとする.

ここで, 簡単に計量や接続の定義 (Amari[2]) をまとめておく. モデル $\mathcal{S}=\{S(\lambda;\theta)\}$ の点 θ における接ベクトル空間はベクトル $\partial_I S(\lambda;\theta)$ $(I=0,\ldots,p)$ で張られる線形空間として定義される. ここで座標 θ とスペクトル密度 $S(\lambda;\theta)$ を同一視している. また, $\partial_I := \partial/\partial\theta^I$ である. 点 $S(\lambda;\theta)$ での接ベクトル $\partial_I S(\lambda;\theta)$ と$\partial_J S(\lambda;\theta)$ との内積, すなわち計量テンソルは,

$$g_{IJ}(\theta) := \langle \partial_I S(\lambda,\theta), \partial_J S(\lambda,\theta)\rangle := \frac{1}{4\pi}\int_{-\pi}^{\pi} \partial_I S(\lambda;\theta)\partial_J S(\lambda;\theta)\frac{1}{\{S(\lambda;\theta)\}^2}d\lambda \quad (I,J=0,\ldots,p),$$

で与えられる. $m-$接続係数は

$$\overset{\mathrm{m}}{\Gamma}_{IJK}(\theta) := \frac{1}{4\pi}\int \partial_I\partial_J S(\lambda;\theta)\cdot\partial_K S(\lambda;\theta)\frac{1}{\{S(\lambda;\theta)\}^2}d\lambda \quad (I,J,K=0,\ldots,p),$$

と定義される (Amari, 1987).

以上を用いると, AR モデルの $m-$平均曲率ベクトルは,

$$h(\lambda;\theta) = 2S(\lambda;\theta)\left\{\frac{\tilde{L}_a(z^{-1})M_a(z^{-1})}{\{L_a(z^{-1})\}^2} + \frac{\tilde{L}_a(z)M_a(z)}{\{L_a(z)\}^2}\right\},$$

となる (Komaki[3]). ただし, $M_a(z)$ は次の多項式で定義される.

$$M_a(z) := \begin{cases} \dfrac{z^2}{z^2-1}\{\tilde{L}_a(z) - L_a(z)\}, & p\text{ が偶数}, \\ \dfrac{z}{z^2-1}\{z\tilde{L}_a(z) - L_a(z)\}, & p\text{ が奇数}. \end{cases}$$

4 定常ガウス型時系列に対する竹内の情報量規準

AR モデルに対するスペクトル密度の空間における竹内情報量規準の補正項は, パラメータ θ を最尤推定量 $\hat{\theta}$ で, 真のスペクトル密度をピリオドグラムで推定するとき, p をパラメータ数として,

$$p + \langle h(\lambda;\hat{\theta}), I(\lambda) - S(\lambda;\hat{\theta})\rangle = p + \int_{-\pi}^{\pi} 2\frac{I(\lambda) - S(\lambda;\hat{\theta})}{S(\lambda;\hat{\theta})}\left\{\frac{\tilde{L}_a(e^{-i\lambda})M_a(e^{-i\lambda})}{\{L_a(e^{-i\lambda})\}^2} + \frac{\tilde{L}_a(e^{i\lambda})M_a(e^{i\lambda})}{\{L_a(e^{i\lambda})\}^2}\right\}d\lambda$$

で表される. $\lambda_t := (2\pi t)/n$ とおくと, ピリオドグラム $I(\lambda_t)$ は, 観測値 x_s のフーリエ変換, $z_t := \sum_{s=1}^{n} e^{-i2\pi ts/n}x_s/\sqrt{n}$ $(1 \le t \le n)$, を用いて, $I(\lambda_t) := |z_t|^2/2\pi$ により与えられる. n は標本数.

参考文献

[1] Akaike, H. (1973). *in Proceedings of 2nd International Symposium on Information Theory*, Akademiai Kiado, Budapest, 267-281.

[2] Amari, S. (1987). *Math. Systems Theory*, **20**, 53-82.

[3] Komaki, F. (1999a). *Journal of Time Series Analysis*, **20**, 31-50.

[4] 駒木文保. (1999b). private note.

[5] 栗木哲, 甘利俊一. (1982). 東京大学工学部計数工学科卒業論文.

[6] 竹内啓. (1976). 数理科学, **153**, 12-18.

A model based declustering method for estimating the background seismic activity

Jiancang Zhuang, Department of Statistical Sciences, Graduate University for Advanced Studies
Yosihiko Ogata, Insitute of Statistical Mathematics
David Vere-Jones, School of Mathematical and Computing Sciences, Victoria University of Wellington

1 Methodology

The epidemic type aftershock sequence model (ETAS model) is developed by Ogata in his series of papers (see Ogata, 1988, 1989, 1992) which can be represented completely by the conditional intensity function which is defined by

$$\begin{aligned} & \Pr\left[N\left\{[t,t+dt)\times[x,x+dx)\times[y,y+dy)\right\}>0|\ \mathcal{H}(t)\right] \\ = \ & \lambda(t,x,y)dtdxdy+o(dtdxdy), \end{aligned} \tag{1}$$

where $\mathcal{H}(t)$ denotes the occurrence history up to time t. The conditional intensity function for the space-time model can be written as

$$\lambda(t,x,y)=\mu(x,y)+\sum_{k:t_k<t}\kappa(M_k)g(t-t_k)f(x-x_k,y-y_k|M_k). \tag{2}$$

In the above equation, μ is the background rate which is independent of time, $g(t)$, $f(x,y|M_k)$, $j(M|M_k)$ are respectively the probability density functions of the occurrence times, the locations, the magnitudes of the offspring from an ancestor of magnitude M_k, and $\kappa(M_k)$ is the expected number of offspring that an ancestor of size M_k can produce.

Integrate over the history on both sides of Eq.(2) yields the equation for the mean rate for event with coordinate (x,y,M), i.e.,

$$\begin{aligned} m_1(x,y) \ = \ & \mu(x,y)+\iint_{R^2}\int_0^{+\infty}\kappa(M^*)f(x-x^*,y-y^*|M^*)j(M^*)\times \\ & m_1(x^*,y^*)dM^*dx^*dy^*. \end{aligned} \tag{3}$$

To make a better interpretation of each term in (3), we consider the thinning operation on this process. We suppose that the points of this process are deleted to produce a thinned process, with a probability ϵ_j for the j-th event.

Here we set ϵ_j to be the probability of the j-th event being an offspring in the process. If we delete the j-th event in the process with probability ϵ_j for all $j=1,2,...,N$, with N being the total number of events, then it is easy to show that we get a subprocess of a non-homogeous Poisson process with the spatial intenisity $\mu(x,y)$. We call this subprocess as the background subprocess or the background subprocess, and the compensation of this subprocess as the cluster subprocess or the offspring process. The conditional intensity of the background subprocess and the offspring subprocess are respectively $\mu(x,y)$ and

$$\sum_{k:t_k<t}\kappa(M_k)g(t-t_k)f(x-x_k,y-y_k|M_k), \tag{4}$$

where the subscription k runs over all the events in both subprocesses.

As it is shown in last section, the total rate, the background rate and the clustering rate are respectively the occurrence rates of the whole process, the background subprocess and the cluster subprocess respectively. And the estimations of the total rate, the background rate and the clustering rate are also the estimations the occurrence rate of these process.

Estimating the mean rate $m_1(x,y)$ can be carried out by using several methods, such as kernel estimates (see Vere-Jones, 1992), 2-dimensional spline method (see Ogata and Katsura, 1988). Here we adopt the kernel estimate, i.e.,

$$\hat{m}_1(x,y)=\sum_j k_{d_j}(x-x_j,y-y_j)=\iint_S k_{d_j}(x-x^*,y-y^*)N(dx^*dy^*) \tag{5}$$

to estimate the mean rate, where $k_d(x, y)$ denotes the kernel function and d_j represents the varying bandwidth. For each j, $j = 1, 2, ..., N$, d_j for the j-th event can be calculated in the following way: given an integer n_p* between 10 and 100, find the smallest disk centred at the epicentre of the j-th event of those include at least n_p neighbouring events, and let this minimum radius be d_j.

It is not difficlut to see that the occurrence rates of the cluster subprocess and the background subprocess can be estimated by the statistic

$$\hat{C}(x, y) = \sum_j \epsilon_j k_{d_j}(x - x_j, y - y_j). \tag{6}$$

and

$$\hat{\mu}(x, y) = \hat{m}_1(x, y) - \hat{C}(x, y) = \sum_j (1 - \epsilon_j) k_{d_j}(x - x_j, y - y_j). \tag{7}$$

2 Algorithms

We use an iterative algorithm to estimate the background rate and the branching structure at the same time.

1. Fix n_p, say 20, calculate the bandwidth d_j for each event (t_j, x_j, y_j, M_j), $j = 1, 2, .., N$;
2. Set $l = 0$, and $u^{(0)}(x, y) = const.$
3. Fit the space-time ETAS model to the earthquake data with the conditional intensity function

$$\lambda(t, x, y) = \gamma u^{(l)}(x, y) + \sum_{k: t_k < t} \kappa(M_k) g(t - t_k) f(x, y | M_k), \tag{8}$$

   where γ is another parameter to be estimated.
4. Calculate ϵ_j in Equation (7, for each j, $j = 1, 2, ..., N$.
5. Calculate $\mu(x, y)$, and record as $u^{(l+1)}(x, y)$.
6. If $\max |u^{(l+1)}(x, y) - u^{(l)}(x, y)| > \epsilon$, where ϵ is a small positive number, then set $l = l + 1$, and go to step 3. Otherwise take $u^{(l+1)}(x, y)$ as the background rate, and stop.

With the background rate and the branching structure (determined by the parameters in the space-time ETAS model) evaluated by the above procedure, we can continue to carry out a thinning algorithm to decluster a catalogue based on probabilities.

3 Applications

We have applied this method to the earthquake data from the central New Zealand and region from the western and central Japan region.

References

Kagan Y. (1991), "Likelihood analysis of earthquake catalogues," *Journal of Geophysical Research*, 106, Ser. B7, 135–148.

Musmeci F., and Vere-Jones D. (1992), "A space-time clustering model for historical earthquakes," *Annals of the Institute of Statistical Mathematics*, 44, 1 – 11.

Ogata Y. (1998), "Space-time point-process models for earthquake occurrences," *Annals of the Institute of Statistical Mathematics*, 50, 379 – 402.

Rathbun S. L. (1993), "Modeling marked spatio-temporal point patterns," *Bulletine of the International Statistical Institute*, 55, Book 2, 379–396.

Vere-Jones D. (1992), "Statistical methods for the description and display of earthquake catalogues". In: *Statistics in the environmental and earth sciences*, edited by. A. Walden, P. Gottorp. Edward Arnold, a division of Hodder & Stoughton, London.

無情報事前分布・縮小型事前分布を利用したベイズ予測

東京大学大学院工学系研究科計数工学専攻　駒木　文保

1. はじめに

統計モデル

$$\mathcal{P} = \{p(x, y|\theta) | \theta := (\theta^a) \in \Theta, a = 1, \ldots, m\}.$$

が実際の問題の十分よい近似になっているとき，観測値 x をもとにして y を予測する問題を考える．x, y が統計モデル $\mathcal{P}$ に属する分布 $p(x, y|\theta)$ から得られるとき，y を x に基づいて構成した分布 $\hat{p}(y; x)$ で予測する．この分布 $\hat{p}(y; x)$ を予測分布と呼ぶ．

以下では簡単のため，θ が与えられたとき，x と y が独立になる場合 ($p(x, y|\theta) = p(x|\theta)p(y|\theta)$ が成立する場合) について述べる．

Kullback-Leibler ダイバージェンス

$$D\{p(y|\theta); \hat{p}(y; x)\} = \int p(y|\theta) \log \frac{p(y|\theta)}{\hat{p}(y; x)} \mathrm{d}y \quad (1)$$

を損失関数として採用すると，パラメータ空間上の事前分布 $f(\theta)$ が与えられる場合には，ベイズリスクは

$$\int f(\theta) \int D(p(y|\theta); \hat{p}(y; x)) p(x|\theta) \mathrm{d}x \mathrm{d}\theta \quad (2)$$

である．ベイズリスク (2) で予測分布の良さを評価するとき，

$$p(y|x) = \int p(y|\theta) \frac{f(\theta) p(x|\theta)}{\int f(\theta') p(x|\theta') \mathrm{d}\theta'} \mathrm{d}\theta$$

が最良の予測分布であることが知られている．ベイズリスク (2) に基づく議論では $f(\theta)$ の選択が大きな問題になるためリスクの θ の関数としての挙動の考察の必要性について指摘がなされている (Akaike 1978)．

2. ベイズ予測分布の漸近理論

次の定理はベイズ予測分布の漸近展開を与える．計量テンソル，接続係数等の記号の定義は Amari (1985) 参照．

定理 1. (Komaki, 1996)

事前分布 $f(\theta)$ を利用して構成したベイズ予測分布

$$\hat{p}(x; x_{N-}) = \int p(x|\theta) f(\theta|x_{N-}) \mathrm{d}\theta$$

は，

$$\hat{p}(x; x_{N-}) = p(x|\hat{\theta}_{\mathrm{mle}}) + \frac{1}{2N} \Delta(x; \hat{\theta}_{\mathrm{mle}}) + \frac{1}{N} \tilde{\Delta}_f(x; \hat{\theta}_{\mathrm{mle}}) + o(N^{-1}),$$

のように展開される．ここで，$\hat{\theta}_{\mathrm{mle}}$ は θ の最尤推定量であり，

$$\tilde{\Delta}_f(x; \hat{\theta}_{\mathrm{mle}}) := \left\{ \partial_a \log f(\hat{\theta}_{\mathrm{mle}}) - \overset{e}{\Gamma}_{ab}{}^{b}(\hat{\theta}_{\mathrm{mle}}) \right\} g^{ac}(\hat{\theta}_{\mathrm{mle}}) \partial_c p(x|\hat{\theta}_{\mathrm{mle}}).$$

とおいた．

3. 変換群構造をもつモデルの右不変事前分布

正規分布 $N(\mu, \sigma^2)$ にしたがう独立な確率変数 $x_1, x_2, \cdots, x_N$ が観測されたとき，x_{N+1} を予測する問題を考える．ベイズ予測分布を構成するための事前分布をリスク関数で評価するとき，

$$f(\mu, \sigma) \mathrm{d}\mu \mathrm{d}\sigma \propto \frac{1}{\sigma} \mathrm{d}\mu \mathrm{d}\sigma \quad (3)$$

があるクラスのなかで最も優れていることが（多変数正規分布の場合について）Geisser (1993) により指摘されている．ここでは，この議論がより一般の変換群構造をもつモデルの場合に拡張できることを示す．簡単のため，$G = X = \Theta$ の場合について述べる．

補題 2. (Komaki, 2000c) データ x に基づいて y を予測する予測分布 $q(y; x)$ を構成することを考える．$q(g^{-1}y; x) = q(y; gx)$ が成立する予測分布のクラスで考えると，右不変分布を事前分布とする予測分布が最良になる．

4. 縮小型予測分布

予測分布を構成する場合にも縮小の手法を利用することで予測の精度が改良できることを示す．

Stein は事前分布

$$f(\mu) = ||\mu||^{-m+2} \quad (4)$$

を利用したベイズ推定量を議論し，James-Stein 推定量を改良するものとして知られる positive part James-Stein 推定量と挙動が近いことを示した (Stein, 1973).

定理 3. (Komaki, 2000b) 分散共分散行列既知の多変数正規分布モデル $N(\mu, I)$ を考える．このとき，事前分布

$$f(\mu) = ||\mu||^{-m+2}$$

を利用して構成される予測分布は Jeffreys 事前分布を利用して構成される予測分布を (厳密に) 優越する．

5. 縮小型事前分布の漸近理論

James-Stein 推定量 $\hat{\mu}_{\mathrm{JS}}$ は

$$\hat{\mu}_{\mathrm{JS}} = \hat{\mu}_{\mathrm{mle}} + \frac{\partial}{\partial x_i} \log ||x||^{-m+2}$$

の形に表現できる．ここで，$||x||^{-m+2}$ のかわりに他の（超）調和関数 $\phi(x)$ を利用して

$$\hat{\mu}^i = \hat{\mu}^i_{\mathrm{mle}} + g^{ij}\partial_j \log \phi(x)$$

としても最尤推定が改良できることが知られている (例えば Rukhin (1995)).

一般のモデルに対して厳密な理論を展開するのはパラメータ推定の場合でも困難であり，Brown (1979) はリスク関数を展開して得られる楕円型微分作用素を含む微分不等式を利用して推定量を構成することを提案している．Eguchi and Yanagimoto (1998) は，損失関数としてここでの設定とは逆向きの Kullback-Leibller ダイバージェンスを採用して，一般の統計モデルで，(超) 調和関数が存在すればリスクを 2 次のオーダーまで評価したときに，最尤推定量を漸近的に改良できることを示した．

これらのことから，予測分布を構成する場合でも，縮小型の事前分布を利用した予測が良いのではないかと想像される．実際，確率分布 $p(x|\theta)$ にしたがう独立な確率変数 $x_1, x_2, \cdots, x_N$ が観測されたとき，x_{N+1} を予測するためのベイズ予測分布 $p(x_{N+1}|x_1, \cdots, x_N)$ を構成することを考えるとき，以下の定理が成立する．

定理 4. (Komaki, 2000d) 統計モデルが 3 次元以上の負曲率の単連結完備リーマン多様体，または厳密に負曲率の 2 次元以上の単連結完備リーマン多様体になるときは，Jeffreys 型事前分布をもとに構成される予測分布を漸近的に優越する縮小型ベイズ予測分布が構成できる．

リーマン多様体 (M, g) ($\dim M \geq 3$) 上に事前分布 $h(\theta)$ が与えられているとき，($h(\theta)$ は Jeffreys 分布に対する密度) (M, g) に対する共形変換

$$\tilde{g}_{ab} = h^{\frac{2}{m-2}} g_{ab}$$

により得られるリーマン多様体 $(M, \tilde{g})$ を考える．

次の定理は，様々な事前分布を利用して構成された予測分布に対して縮小型の改良を加えたベイズ予測分布を構成する方法を与える．

定理 5. (Komaki, 2000d) $(M, \tilde{g})$ が 3 次元以上の負曲率の単連結完備リーマン多様体になるときは，$h(\theta)$ を事前分布として構成されるベイズ予測分布を漸近的に優越する縮小型ベイズ予測分布が構成できる．

参考文献

Akaike, H. (1978). A new look at the Bayes procedure, *Biometrika* vol. 65, pp. 53-59.
Amari, S. (1985). *Differential-Geometrical Methods in Statistics.* New York: Springer-Verlag.
Brown, L. (1979). A heuristic method for determining admissibility of estimators - with applications. *Annals of Statistics*, vol. 7, pp. 960-994.
Eguchi, S. and Yanagimoto, T. (1998). Asymptotical improvement of maximum likelihood estimators using relative entropy risk, Research Memorandum No. 665, Institute of Statistical Mathematics, Japan.
Geisser, S. (1993). *Predictive Inference: An Introduction*, London: Chapman & Hall.
Komaki, F. (1996). On asymptotic properties of predictive distributions, *Biometrika*, vol. 83, pp. 299-313, 1996.
Komaki, F. (1999). An estimating method for parametric spectral densities of Gaussian time series, *Journal of Time Series Analysis*, vol. 20, pp. 31-50.
駒木 文保 (2000a). 予測分布の理論について. 電子情報通信学会論文誌A, vol. J83-A, pp. 612-619.
Komaki, F. (2000b). A shrinkage predictive distribution for multivariate Normal observables, tentatively accepted for publication in *Biometrika.*
Komaki, F. (2000c). Some remarks on non-informative priors and prediction, *Proceeding of the 4th International Triennial Calcutta Symposium.*
Komaki, F. (2000d). On shrinkage predictive distributions, in preparation.
Rukhin, A. L. (1995). Admissibility: survey of a concept of progress. *International Statistical Review*, vol. 63, pp. 95–115.
Stein, C. (1973). Estimation of the mean of a multivariate normal distribution, *Proceeding of the Prague Symposium on Asymptotic Statistics*, pp. 345–381.

不等式制約下のモデリング

明星大学理工学部　　広津 千尋

1. はじめに

不等式制約の下の統計的推測でもっとも古くから議論されているのは，独立な正規分布の平均に関する単調制約の仮説検定問題であろう．これについては Bartholomew が 1950 年代後半から 60 年代前半にかけ，JRSS や Biometrika に一連の論文を発表し，さらに 1972 年にいわゆる 4B の本 (Barlow 他, 1972) が出版されて一段落したように見える．実際 4B の本では，この問題に関して単調回帰，maxmin 線形対比検定そして Williams 等の最大対比検定を取り上げ，大きな流れはほぼこれに尽きるであろうと述べられている．ところが，その後の研究の発展を見ると，この記述は必ずしも当を得ていないように思われる．それは単調制約問題の拡張に伴い，新たなアプローチの提案が見られるからである．実際，単調回帰のアプローチは直観的にも受け入れ易く性能も良いが，例えば 2 元配置交互作用問題に拡張するには困難が伴う．本論では，まず次節で単調制約も含め，様々な不等式制約の統計モデルを導入し，その後，それら拡張モデルに対するアプローチを単調回帰以外のものを中心に述べることとする．

2. 不等式制約で規定される統計的問題

単調制約は説明変数が温度，期間，年令のように自然な順序を持ち，それに応じて反応が単調増大性を満たすものの，必ずしも線形性が保証されないときに，線形回帰モデルに替って用いられる．最もよく用いられるのは用量反応解析であって，正規分布モデル $N(\mu_i, \sigma^2)$，2 項分布モデル $B(n_i, p_i)$，それぞれの場合に単調制約

$$H_m : \mu_1 \le \cdots \le \mu_K, \text{ 又は } p_1 \le \cdots \le p_K \tag{2.1}$$

が仮定される．パラメータ空間の制約により，推測の効率を上げることが目的である．

ノンパラメトリックな用量反応解析では凸性仮説や S 字性仮説も有用である．凸性仮説を積極的に利用した μ_i の同時信頼区構成法については Hirotsu & Sirvastava(2000) を参照されたい．用量水準が等間隔の場合の凸性仮説は既に Hirotsu(1988) において，3 次元コホートモデルにおけるコホート効果の変化点検出問題で用いられている．一方，経済モデルでは入力に対し出力が頭打ちを示すことが多く，凹性仮説が良く用いられる (Matzkin, 1994)．なお，単調，凸性，S 字性仮説は k 次元空間において平均ベクトル $\boldsymbol{\mu}$ をある凸錐内に制限する．興味あることに，それら凸錐のコーナーベクトルはそれぞれ階段型変化，スロープ変化そして変曲点モデルを定義する．そしてそれは新たな検定統計量を自然に導入するのに用いられる．

次に，2 元配置分散分析モデルでは，交互作用に関する傾向仮説

$$\mu_{i+1j} - \mu_{ij} \le \mu_{i+1j+1} - \mu_{ij+1}, \quad i = 1, \cdots, a-1; j = 1, \cdots, b-1 \tag{2.2}$$

が導入されている (Hirotsu, 1978)．モデル (2.2) は行の第 $i+1$ 水準と第 i 水準の差が列の水準が進むに従って増大することを示している．もし，行の水準には順序が無く，列の水準のみに順序があると両側仮説 $\mu_{ij} - \mu_{i'j} \le \mu_{ij+1} - \mu_{i'j+1}$, $i \ne i'; j = 1, \cdots, b-1$ に興味が持たれる．用量設定のための臨床試験でよく得られる 2 次元順序分割表では，用量水準 i が増すごとに効果 j が強まるという仮説

$$\frac{p_{i+11}}{p_{i1}} \le \frac{p_{i+12}}{p_{i2}} \le \cdots \le \frac{p_{i+1b}}{p_{ib}}, \ i = 1, \cdots, a-1 \tag{2.3}$$

が設定される．(2.3) 式は，加法モデル (2.2) に対する乗法モデル版である．

3. 検定統計量

単調仮説 H_c の検定で 4B の本以降に提案されたものに累積 χ^2 統計量 $\chi^{*2}=\sum_1^{K-1}t_i^2$ と $\max t$ 統計量 $\max t=\max_{i=1,\cdots,K-1}t_i$ がある．ただし，t_i は第 i 水準までの平均と，$i+1$ 水準以降の平均とを比較する t 統計量である．累積 χ^2 は元来は田口氏によって直観的に導入され，その後竹内 (1979)，および Hirotsu(1982) によって数理的構造と最適性が明らかにされたものである．とくに σ 既知 ($\hat{\sigma}$ の自由度 ∞) のときに，単調性仮説検定の完全類は各 t_i, $i=1,\cdots,K-1$, について単調増大，かつ凸な受容域を持つ検定の全体として与えられ，χ^{*2} と $\max t$ はその条件を満たしていることが分かる．ただし，χ^{*2} は両側仮説に対して適切な検定統計量である．これら検定の検出力は，様々な機会に検討され，それ以前の方法に比べて遜色のないことが示されている．さらに特筆すべきは，単調回帰に比べて，より複雑なモデルへの拡張が容易なことである．それには，χ^{*2} と $\max t$ の成分 t_i が，H_m のコーナーベクトルに対応する階段型変化点モデルの第 i 成分に対する尤度比検定統計量になっていることに注目する．すなわち，スロープ変化モデルおよび変曲点モデルに対して同じように尤度比検定を構成することにより，χ^{*2} と $\max t$ を拡張することができる．それらは検定の完全類からも支持される．同じ考え方は交互作用仮説 (2.2)，(2.3) にも適用され，とくに順序分割表に関する累積 χ^2，$\max\chi^2$ 統計量等を導く (Hirotsu, 1983)．$\max t$ と $\max\chi^2$ は多重比較法としての特長を有している．従って，例えば用量反応解析で反応が段差的に変化する水準を検出するのに適切な統計量となる．例えば行が用量水準，列が順序カテゴリ応答の場合には，列に関して $\max\chi^2$ 型の統計量を構成し，列に関して線形順位和，累積 χ^2 あるいは再び $\max\chi^2$ 型統計量を構成するなどのバリエーションも考えられる (Hirotsu, 1993, 1998b)．計量値の場合に，列に沿った累積 χ^2 型統計量に基づく行の多重比較により，行のクラスタリングに成功した興味ある例が Hirotsu(1991) に与えられている．

3 元表に関する不等式制約モデルの例はあまり多くないが，広津 (1992)，Hirosu(1998a) そして広津他 (2000) などに例が見られる．

＜引用文献＞

Barlow, R. E. et al. : Wiley, New York (1972).

Hirotsu, C. : *Biometrika* 65, 561-570 (1978).

Hirotsu, C. : *Biometrika* 69, 567-577 (1982).

Hirotsu, C. : *Biometrika* 70, 579-590 (1983).

Hirotsu, C. : *Annals Inst. Statist. Math.* A 40, 451-465 (1988).

Hirotsu, C. : *Biometrika* 78, 583-594 (1991).

広津千尋：共立出版，東京 (1992).

Hirotsu, C. : *Int. Statist. Rev.* 61, 183-201 (1993).

Hirotsu, C. : *Proc. Int. Sympos. Quality Improvement through Statistical Methods*, Birkhauser, Boston, 275-287 (1998a).

Hirotsu, C. : *Encyclopedia of Biostatistics*, Wiley, New York, 2107-2115 (1998b).

広津千尋, 青木敏，稲田俊也，北尾淑恵：日本統計学会第 68 回講演予稿集，407-408 (2000).

Hirotsu, C. and Srivastava, M. S. : *Statistics and Probability Letters* 49, 25-37 (2000).

Matzkin, R. L. : *Handbook of Econometrics* 4, North-Holland, Amsterdam, 2523-2558 (1994).

竹内啓：経済学論集 45, 3, 2-10 (1979).

Sequential Ordinal Probit Model With Autoregressive Random Effects

Yasuhiro Omori
Department of Economics, Tokyo Metropolitan University

1 Introduction

Economic duration data are subject to some external macroeconomic factors represented by common (possibly unobserved) time-dependent variables. In multiple economic times series analysis, for example, a dynamic factor model are often considered to explain such a comovement among time series. This possibly unobserved common dynamic component can be modeled as a time-dependent random effect and needs to be considered in discrete duration models for economic data.

2 Model

We consider a discrete duration time Y_{it} for the i-th subject at calendar time t $(i = 1, \ldots, n, t = 1, \ldots, T)$. The duration has a discrete distribution at duration time $j \in \{1, 2, \ldots, J+1\}$ given the parameter θ. The discrete duration time points correspond to certain intervals $A_1, A_2, \ldots, A_{J+1}$ where $A_j = (a_{j-1}, a_j]$, $j = 1, 2, \ldots, J+1$ with $a_0 = 0 < a_1 < \ldots < a_{J+1} = \infty$. We denote the failure at duration time j by $Y_{it} = j$ for the i-th subject at calendar time t and

$$\Pr(Y_{it} = j | Y_{it} \geq j, \alpha, \beta, \gamma_t^*) = F(\theta_{it}), \quad \theta_{it} = \alpha_j - x_{it}'\beta - z_{it}'\gamma_t,$$

where F is a normal cumulative distribution function, and $\alpha = (\alpha_1, \ldots, \alpha_J)'$ represents a baseline hazards function (or cutpoints), and x_{it}, z_{it} are vectors of (calendar) time-varying covariates, and β is a regression parameter vector. The coefficient γ_t is a vector of (calendar) time-variyng parameters, and $\gamma_t^* = \{\gamma_1, \ldots, \gamma_t\}$.

This setup of the model can be derived by incorporating latent variables. Suppose that the duration of the i-h subject starts at calendar time $t - j + 1$ and $Y_{it} \geq j$. Then define latent variables W_{it} and w_{it} such that

$$W_{it} \quad = \quad x_{it}'\beta + z_{it}'\gamma_t + e_{it}, \quad w_{it} = W_{it} - \alpha_j,$$

where $e_{it} \sim$ i.i.d. F. We observe $Y_{it} = j$ when $W_{i,t-j+1} > \alpha_1, \ldots, W_{i,t-1} > \alpha_{j-1}$ and $W_{it} \leq \alpha_j$ (equivalently, $w_{i,t-j+1} > 0, \ldots, w_{i,t-1} > 0$ and $w_{it} \leq 0$) .

To impose smoothness conditions on the baseline hazard function by some prior process to obtain stable estimators, we assume that α_j follows a simple random walk as in Gamerman (1992).

$$\alpha_{j+1} = \alpha_j + u_j, \quad u_j \sim \text{i.i.d. } N(0, \sigma_u^2), \quad j = 1, \ldots, J-1,$$

and $\alpha_1 \sim N(0, 10^3)$.

The γ_t 's introduce autoregressive random effect components to account for unobserved common dynamics subject to the macroeconomic condition. We assume that the γ_t 's follow a stationary process. For example, we consider the following processes for the univariate γ_t ,

$$\gamma_{t+1} \quad = \quad \phi\gamma_t + \epsilon_t, \quad \epsilon_t \sim \text{i.i.d. } N(0, \sigma_\epsilon^2), \quad \gamma_1 \sim N\left(0, \sigma_\epsilon^2/(1-\phi^2)\right), \quad |\phi| < 1$$

for $t = 1, \ldots, T-1$ where the ϕ 's are unknown parameters.

3 Markov Chain Monte Carlo Implementation

Generation of the w_{it} *'s.* Generate $w|y,\alpha,\beta,\gamma$ by sampling $w_{it}|(y_{it}=j,\beta,\gamma)$ for each i and t. If $y_{it}=1$, we generate w_{it} from $TN_{(-\infty,0)}(x'_{it}\beta+z'_{it}\gamma_t)$. If $y_{it}=j$ $(2\le j\le J)$, generate $w_{i,t-j+k}$ from $TN_{(0,\infty)}(x'_{i,t-j+k}\beta+z'_{i,t-j+k}\gamma_{t-j+k})$ for $k=1,\dots,j-1$, and w_{it} from $TN_{(-\infty,0)}(x'_{it}\beta+z'_{it}\gamma_t)$ independently. If $y_{it}=J+1$, generate $w_{i,t-J-1+k}$ from $TN_{(0,\infty)}(x'_{i,t-J-1+k}\beta+z'_{i,t-J-1+k}\gamma_{t-J-1+k})$ for $k=1,\dots,J$.
Generation of the β *'.* We assume a multivariate normal prior for β, $\beta\sim N(b_0,B_0)$, which is conjugate. The posterior distribution is also multivariate normal with a mean $\hat\beta$ and a covariance matrix B_1 such that

$$\hat\beta=B_1\left\{B_0^{-1}b_0+\sum_{t=1}^{T}\sum_{i\in R_t}(W_{it}-z'_{it}\gamma_t)x_{it}\right\},\quad B_1=\left(B_0^{-1}+\sum_{t=1}^{T}\sum_{i\in R_t}x_{it}x'_{it}\right)^{-1},$$

where R_t denotes a set of labels attached to the subjects at risk at calendar time t.
Generation of the γ_t *'s.* First we define a matrix Q_t which extracts observations at risk at calendar time t. For example, if $Y_{it}\ge j$, the i-th row of Q_t is $(0,\dots,0,1,0\dots,0)$ with the j-th element equal to 1. Let $\hat y_t=Q_tW_t$, $\hat x_t=Q_tx_t$, $\hat z_t=Q_tz_t$, $\hat e_t=Q_te_t$ where $W_t=(W_{1t},\dots,W_{nt})',x_t=(x_{1t},\dots,x_{nt})',z_t=(z_{1t},\dots,z_{nt})'$ and $e_t=(e_{1t},\dots,e_{nt})'$. Then we consider a linear Gaussian system

$$\begin{aligned}\hat y_t &= \hat x'_t\beta+\hat z'_t\gamma_t+\hat e_t,\quad \hat e_t\sim N(0,I_{n_t}),\\ \gamma_{t+1} &= \Phi\gamma_t+\epsilon_t,\quad \epsilon_t\sim N(0,\Sigma),\end{aligned}$$

for $t=1,\dots,T$ where $\gamma_1\sim N(0,\Sigma_0)$ and n_t is a number of subjects at risk at calendar time t. By using the simulation smoother given by de Jong and Shephard (1995), we sample disturbances term and obtain samples of the γ_t 's.
Generation of the α_j *'s.* Generate α by using the simulation smoother to the linear Gaussian system with

$$\begin{aligned}\hat y_j &= \alpha_j+v_j,\quad v_j\sim N(0,m_j^{-1}),\\ \alpha_{j+1} &= \alpha_j+u_j,\quad u_j\sim N(0,\sigma_u^2),\end{aligned}$$

for $j=1,\dots,J$ where $\alpha_1\sim N(0,10^3)$, $\hat y_j=\frac{-1}{m_j}\sum_{t=1}^{T}\sum_{i\in R_{jt}}(w_{it}-x'_{it}\beta-z'_{it}\gamma_t)$, $R_{jt}=\{i:Y_{it}\ge j\}$ and m_j is the number of such observations. Generations of $\phi,\sigma_\epsilon^2,\sigma_u^2$ are straightforward.

References

Albert, J. and Chib, S. (1997), "Sequential Ordinal Modeling with Applications to Survival Data," Technical report, Olin School of Business, Washington University, St. Loius.

de Jong, P. and Shephard, N. (1995), "The Simulation Smoother for Time Series Models," *Biometrika*, 82, 339-350.

Gamerman, D. (1992), "A Dynamic Approach to Statistical Analysis of Point Processes," *Biometrika*, 79, 39-50.

有意水準が小さい時の検定の検出力の比較について

東京大学・経済 竹村彰通

1 導入

仮説検定問題において検定統計量の帰無分布が近似的にも知られていない場合は多い. そのような統計量の対立仮説のもとでの検出力を求めることは当然より難しい問題であり, 通常はシミュレーションによって検出力の検討がなされる. ところで最近になって「最大値型」の検定統計量についてはいわゆる tube 法により帰無分布の裾確率の近似式が求められるようになって来た. Sun, Worsley, 竹村・栗木, 二宮らの研究によれば tube 法による帰無分布の裾確率の近似は良好である場合が多く, 実際の検定の場面で必要とされるような上側 5% あたりからは十分な近似を与えるようである.

Tube 法により理論的に扱える帰無分布の範囲が拡大されたが, この延長として対立仮説のもとでの分布にも tube 法を応用し検出力の挙動について理論的な結果を得ることがここでの目的である. 固定した対立仮説のもとでは, 有意水準 α を 0 に近づけるとともに検出力 $\beta=\beta(\alpha)$ も 0 に近づく. 正の α に対して検出力の値 $\beta(\alpha)$ を厳密に評価する事は難しくても, 以下の例で見るように $\alpha\to 0$ と極限操作をおこなった時の $\beta(\alpha)$ の挙動は明示的に記述できる場合がある. これによって α が十分小さい場合について検定の検出力の明示的な比較が可能となる.

2 記法と定義

いま帰無仮説 $H_0:\theta=\theta_0$ に関する検定統計量を T とし, 棄却域が $T>c$ の形で与えられるとする. ここでは簡単のため T の分布は連続分布であるとする. 帰無仮説が必ずしも単純仮説にならない場合もあるが, 本稿では簡単のため T は相似検定であり, 帰無仮説の各点で T の分布が共通であるとする. パラメータの値 θ のもとでの T の上側確率を

$$\bar{F}_\theta(x)=P_\theta(T>x)$$

と表すと, 有意水準が α の時の棄却点は $\bar{F}_{\theta_0}^{-1}(\alpha)$ である. また対立仮説 $\theta\neq\theta_0$ のもとでの検出力は

$$\beta(\alpha)=\beta(\alpha,\theta)=\bar{F}_\theta(\bar{F}_{\theta_0}^{-1}(\alpha))$$

である. 以下対立仮説 θ は固定しておいて検出力を α の関数と考える. いま明示的に評価できる関数 $g(\alpha)$ があり

$$\beta(\alpha)\sim g(\alpha),\qquad (\alpha\to 0) \tag{1}$$

となったとする．本稿では記号 "$\sim$" を exact asymptotics の意味，つまり両辺の比が 1 に収束するという意味で用いることとする．(1) 式の $g(\alpha)$ をここでは "order of small size power of T at θ" とよぶ事にする．そして Order of SSP あるいは OSSP と書くこととする．

3 いろいな検定問題に関する OSSP

指数分布

$$f(x,\theta) = (1/\theta)e^{-x/\theta}, \qquad x, \theta > 0$$

で帰無仮説を $H: \theta = \theta_0 = 1$ とする．対立仮説として $\theta > 1$ を固定する．MP 検定について $\beta_{MP}(\alpha,\theta)$ は

$$\beta_{MP}(\alpha,\theta) = e^{(\log\alpha)/\theta} = \alpha^{1/\theta},$$

と容易に求められる．次に両側に $\alpha/2$ の確率を残す形の両側検定の OSSP は

$$\beta(\alpha) \sim \alpha^{1/\theta}2^{-1/\theta} = g(\alpha)$$

と評価される．

1 変量正規分布の平均の場合を考える．$X \sim N(\theta,1)$ として帰無仮説を $H: \theta = 0$ とおく．対立仮説として $\theta > 0$ を固定する．片側検定の OSSP は

$$\beta_{MP}(\alpha) \sim \alpha \exp(-\frac{1}{2}\theta^2)\exp(\sqrt{-2\log\alpha}\theta)$$

と評価される．次に両側検定の OSSP を評価すると

$$\beta(\alpha) \sim \frac{\alpha}{2}\exp(-\frac{1}{2}\theta^2)\exp(\sqrt{-2\log\alpha}\theta)$$

となる．

さらに多変量解析の問題としては MANOVA での最大根検定, ordered alternative に関する尤度比検定について OOSP を明示的に評価する事ができる．

OOSP の考え方は Large deviation に基づく exact slope と似た点もあるが, sample size と有意水準の極限操作の順序が異なっており, 実際の結果も違う形で与えられることが多い．

参考文献

[1] Takemura, A. and Kuriki, S. (1997). Weights of $\bar{\chi}^2$ distribution for smooth or piecewise smooth cone alternatives. *Ann. Statist.*, **25**, 2368–2387.

[2] Kuriki, S. and Takemura, A. (1998). Tail probabilities of the maxima of multilinear forms and their applications. *Discussion Paper CIRJE-F-4*, Faculty of Economics, Univ. of Tokyo.

[3] A. Takemura and S. Kuriki (1999). Tail probability via tube formula and Euler characteristic method when critical radius is zero, Discussion Paper CIRJE-F-59, Faculty of Economics, Univ. of Tokyo.

2項混合分布のコンポーネント数の検定と tube 法

統計数理研究所　栗木哲

1　混合分布のコンポーネント数の推測

tube 法の一つの典型的な応用例として，有限混合分布のコンポーネント数の検定問題がある．ここではその中でも簡単な場合である2つの2項分布の混合分布をとりあげる．1節では Chernoff and Lander (1995, *JSPI*) の結果を再導出 (別証明) し，2節ではその結果として現われる極限分布の上側裾確率を tube 法で評価する．3節では 1, 2節の拡張を行う．

2つの2項分布の混合分布 $\alpha \mathrm{Bin}(k,p)+(1-\alpha)\mathrm{Bin}(k,1/2)$ に従う n 個の i.i.d. 観測値が得られているとする．母集団が $\mathrm{Bin}(k,1/2)$ であるという帰無仮説，すなわち H_0 : $\alpha=0$ or $p=1/2$ の検定問題を尤度比検定により考えることにする．

観測値は多項分布 $\mathrm{Mult}(n;q_0,\ldots,q_k)$, $q_i=\binom{k}{i}\{\alpha p^i(1-p)^{k-i}+(1-\alpha)(1/2^k)\}$ の実現値とみなすことができる．多項分布の母数空間は単体であり，帰無仮説 H_0 はその中の一点 $q^o=(q_0^o,\ldots,q_k^o)'$, $q_i^o=(1/2)^k\binom{k}{i}$ からなる．

真値の近傍で対立仮説を錐で近似することを考える．真値を原点とすると対立仮説は錐

$$S=\{\alpha f(\phi)=\alpha(f_0(\phi),\ldots,f_k(\phi))' \mid |\phi|\le 1,\ \alpha\ge 0\},\quad f_i(\phi)=\binom{k}{i}\left\{(1+\phi)^i(1-\phi)^{k-i}-1\right\} \tag{1}$$

で近似できる．真値で評価した多項分布の漸近分散行列を $V(q^o)=\mathrm{diag}(q_i^o)-q^oq^{o\prime}$ とおく．その一般逆行列の一つを $V(q^o)^-=\mathrm{diag}(1/q_i^o)$ ととる．Chernoff (1954, *AMS*) の議論より以下が従う．

補題 1 $Z=(Z_0,\ldots,Z_k)'$ を $N_{k+1}(0,V(q^o))$ に従う確率ベクトルとする．Z の錐 S への直交射影を Z_S と書く．このとき尤度比検定統計量は Z_S の長さの2乗 $\|Z\|^2-\inf_{x\in S}\|Z-x\|^2=\|Z_S\|^2$ に分布収束する．ただしここで内積，ノルムは $\langle x,y\rangle=x'V(q^o)^-y$, $\|x\|=\langle x,x\rangle^{\frac{1}{2}}$ で定義する．

簡単な計算より次のことが分かる．

$$\|Z_S\|=\max\left\{0,\ \sup_{|\phi|\le 1}Z(\phi)\right\},\quad Z(\phi)=\frac{\langle f(\phi),Z\rangle}{\|f(\phi)\|}=\frac{2^{k/2}\sum_{i=0}^k\{(1+\phi)^i(1-\phi)^{k-i}-1\}Z_i}{\sqrt{\sum_{i=0}^k\binom{k}{i}\{(1+\phi)^i(1-\phi)^{k-i}-1\}^2}}.$$

$Z(\phi)$ は正規変量の線形結合であるので正規分布に従う．$Z(\phi)$ の平均と分散は 0, 1 であることが容易に確認できる．また相関構造は

$$\mathrm{Cov}(Z(\phi),Z(\tilde\phi))=\frac{(1+\phi\tilde\phi)^k-1}{\sqrt{(1+\phi^2)^k-1}\sqrt{(1+\tilde\phi^2)^k-1}}$$

である．従って変数変換により $Y_1,\ldots,Y_k\sim N(0,1)$ i.i.d. を用いて

$$Z(\phi)=\sum_{i=1}^k g_i(\phi)Y_i,\qquad g_i(\phi)=\frac{1}{\sqrt{(1+\phi^2)^k-1}}\binom{k}{i}^{\frac{1}{2}}\phi^i$$

と表現することができる．

2 tube 法による裾確率の評価

$Y=(Y_1,\ldots,Y_k)'\sim N_k(0,I_k)$, $g(\phi)=(g_1(\phi),\ldots,g_k(\phi))'$ とおく．$\phi\approx 0$ のとき $g_i(\phi)=\mathrm{sgn}(\phi)$ $(i=1)$, $=0$ $(i\geq 2)$ であるので $g(\phi)$ は $\phi=0$ で不連続である．また $M=\{g(\phi)\in R^k\mid 0<|\phi|\leq 1\}$ とおく．このとき (Euclid ノルムで) $\|g(\phi)\|=1$ なので $M\subset S^{k-1}$ (単位球面) である．$\sup_{|\phi|\leq 1}Z(\phi)=\sup_{|\phi|\leq 1}g(\phi)'Y=\sup_{u\in M}u'Y$ に注意する．以下では $x\to\infty$ のときの極限分布の上側裾確率 $P(\|Z_S\|\geq x)=P(\sup_{u\in M}u'Y\geq x)$ を tube 法で評価する．

M は単位球面上の 2 本の滑らかな曲線 (連結成分 2 の，境界を持った 1 次元多様体) である．M の半径 θ の tube は $M_\theta=\{u\in S^{k-1}\mid \min_{v\in M}\cos^{-1}(u'v)\leq\theta\}$ で定義される．M の体積 (長さ) を $\mathrm{Vol}(M)$ と書く．このとき tube M_θ の体積は

$$\mathrm{Vol}(M_\theta)=\Omega_k\left\{\frac{\mathrm{Vol}(M)}{2\pi}P\left(\mathrm{Beta}\left(1,\frac{k-2}{2}\right)\geq\cos^2\theta\right)+2\times\frac{1}{2}P\left(\mathrm{Beta}\left(\frac{1}{2},\frac{k-1}{2}\right)\geq\cos^2\theta\right)\right\}$$

と表わされる．ここで $\Omega_k=2\pi^{k/2}/\Gamma(k/2)$ は単位球面 S^{k-1} の体積である．この tube 公式に対応して，裾確率の漸近公式

$$P\left(\max_{u\in M}u'Y\geq x\right)\sim\frac{\mathrm{Vol}(M)}{2\pi}P(\chi^2(2)\geq x^2)+2\times\frac{1}{2}P(\chi^2(1)\geq x^2),\quad x\to\infty,$$

が得られる．M の長さは，体積要素

$$\|g'(\phi)\|\,d\phi=\frac{\sqrt{k\left\{(1+\phi^2)^k-1-k\phi^2\right\}(1+\phi^2)^{k-2}}}{(1+\phi^2)^k-1}\,d\phi$$

を用いて $\mathrm{Vol}(M)=\{\int_0^1+\int_{-1}^0\}\|g'(\phi)\|\,d\phi=2\int_0^1\|g'(\phi)\|\,d\phi$ と評価することができる．

3 問題の拡張

1, 2 節で扱った問題の拡張として，2 つのコンポーネントの 2 項比率 $p,\tilde{p}$ がともに未知である 2 項混合分布モデル $\alpha\mathrm{Bin}(k,p)+(1-\alpha)\mathrm{Bin}(k,\tilde{p})$ においてコンポーネント数が 1 であることの検定，すなわち $\tilde{H}_0:\alpha=0$ or $\alpha=1$ or $p=\tilde{p}$ の検定を考える．ただし簡単のために帰無仮説の下での 2 項比率の真値は 1/2 とする．

1, 2 節で扱った問題とは異なり，帰無仮説は単純仮説ではない．Chernoff (1954, *AMS*) の議論に帰着させるために，帰無仮説および対立仮説の接錐を構成する．

帰無仮説のなす空間は $\{q\mid q_i=\binom{k}{i}\tilde{p}^i(1-\tilde{p})^{k-i},0\leq\tilde{p}\leq 1\}$ である．これを真値のまわりで錐近似すると，1 次元線形空間

$$\tilde{S}_0=\left\{\alpha(x_0,\ldots,x_k)'\mid x_i=\binom{k}{i}(2i-k),\ \alpha\in R^1\right\}$$

を得る．また対立仮説の接錐は Lindsay (1995, IMS モノグラフ), Dacunha-Castelle and Gassiat (1999, *AS*) の議論にあるように $\tilde{S}_0$ と 1 節 (1) 式で現れた接錐 S との和 $\tilde{S}_1=\tilde{S}_0+S=\{x+y\mid x\in\tilde{S}_0,\ y\in S\}$ となる．$\tilde{S}_1$ は $\tilde{S}_1=\tilde{S}_0\oplus\tilde{S}_1'$,

$$\tilde{S}_1'=\left\{\alpha(\tilde{f}_0(\phi),\ldots,\tilde{f}_k(\phi))'\mid|\phi|\leq 1,\ \alpha\geq 0\right\},\quad \tilde{f}_i(\phi)=\binom{k}{i}\left\{(1+\phi)^i(1-\phi)^{k-i}-1-(2i-k)\phi\right\}$$

と直交直和分解できる．帰無仮説の下での尤度比検定統計量は $\max\{0,\sup_{|\phi|\leq 1}\tilde{Z}(\phi)\}^2$, ただし

$$\tilde{Z}(\phi)=\sum_{i=2}^{k}\tilde{g}_i(\phi)Y_i,\qquad \tilde{g}_i(\phi)=\frac{1}{\sqrt{(1+\phi^2)^k-1-k\phi^2}}\binom{k}{i}^{\frac{1}{2}}\phi^i$$

に分布収束する．$g_i(\phi)$ とは異なり $\tilde{g}_i(\phi)$ は $\phi=0$ においても連続である．裾確率計算には $\tilde{M}=\{\tilde{g}(\phi)\in R^{k-1}\mid|\phi|\leq 1\}$ の tube の体積を考えればよい．$\tilde{M}$ は S^{k-2} 上の連結した曲線となる．

Asymptotic Properties of Conditional Maximum Likelihood Estimator
In a Certain Exponential Model

東京工業大学・情報理工　藤澤洋徳

1. Introduction. Consider an exponential model given by

$$p(x;\theta) = \exp\{\theta \cdot t(x) - \psi(\theta)\}\, b(x),$$

where the natural parameter θ and the sufficient statistic t are vectors, '$\cdot$' means inner product, and ψ is the potential function. Vectors are taken to be row vectors. In the following, this model is expressed as $x \sim \mathrm{EM}(t;\theta)$, where EM is an abbreviation for exponential model. The expectation parameter $\tau = \mathrm{E}(t)$ is dual to the natural parameter. Let $x_1,\ldots,x_n$ be the random samples and let $\bar{t}$ be the sample mean of $t(x_1),\ldots,t(x_n)$. The maximum likelihood estimator is given by $\hat{\tau} = \bar{t}$. Let $\theta = (\theta_1,\theta_2)$, $\tau = (\tau_1,\tau_2)$, and $\bar{t} = (\bar{t}_1,\bar{t}_2)$ be similar partitions. The orthogonal parameter $w = (\theta_1,\tau_2)$ is another parameterization and sometimes convenient. This paper regards θ_1 as the objective parameter and τ_2 as the nuisance parameter. The conditional maximum likelihood estimator of θ_1 is given by

$$\check{\theta}_1 = \arg\max p(\bar{t}_1 \,|\, \bar{t}_2;\theta_1),$$

where $p(\bar{t}_1 \,|\, \bar{t}_2;\theta_1)$ is the conditional probability density function of $\bar{t}_1$ given $\bar{t}_2$. This paper investigates asymptotic properties of the conditional maximum likelihood estimator in the exponential model with a τ-parallel foliation and compares the conditional maximum likelihood estimator with the maximum likelihood estimator. It is proved that the exponential model possesses a τ-parallel foliation if and only if the conditional maximum likelihood estimator is asymptotically efficient and its asymptotic variance depends on the objective parameter θ_1 only. It is shown that the bias of the conditional maximum likelihood estimator is more robust than the bias of the maximum likelihood estimator. We see that two estimators are very close, especially in the sense of bias-corrected version. The mean Pythagorean relation implies that the conditional maximum likelihood estimator is better than the maximum likelihood estimator in the sense of Kullback-Leibler risk under an appropriate condition.

2. Reproductivity and τ-parallel foliation. The exponential model is said to be reproductive (in t_2) if $\bar{t}_2 \sim \mathrm{EM}\left((H(\bar{t}_2),\bar{t}_2);n\theta\right)$, where H is an appropriate function. This is equivalent to $s = \bar{t}_1 - H(\bar{t}_2) \sim \mathrm{EM}\left(s;n\theta_1\right)$ and equivalent to the property that s and $\bar{t}_2$ are mutually independent. We say that the exponential model possesses a τ-parallel foliation if

$$\theta_2 = -\theta_1 h(\tau_2) + k(\tau_2),$$

where $h(\tau_2) = \partial H/\partial\tau_2'$, $k(\tau_2) = \partial K/\partial\tau_2$, and $H(\tau_2)$ and $K(\tau_2)$ are appropriate functions. This relation is equivalent to $\tau_1 = m(\theta_1) + H(\tau_2)$, where $m(\theta_1) = \partial M/\partial\theta_1$ and $M(\theta_1)$ is an appropriate function. In subsequent sections, it is assumed that the exponential model possesses the τ-parallel foliation. It is known that if the exponential model is reproductive, it possesses a τ-parallel foliation with common H. Consider the converse. If k is constant and t_2 is one-to-one correspondent to x, the converse is true. In other cases, the converse is a conjecture. It is shown in Section 4 that a part of the conjecture is asymptotically true.

3. Tensors and connections. This section provides some lemmas. They play important roles to reveal asymptotic properties. In the following, abbreviation notations are employed. Let $\theta = (\theta^i)$, $\tau = (\tau_i)$, and $w = (w^i)$. The partial differential operator is expressed as $\partial_i = \partial/\partial w^i$. Indexes $a, b, c, \ldots$ are used in former components and indexes $\kappa, \lambda, \mu, \ldots$ are used in latter components; e.g. $\theta_1 = (\theta^a)$ and $\theta_2 = (\theta^\kappa)$. Remember that $M(\theta_1)$ and $H(\tau_2)$ depend on $\theta_1 = w_1$ and $\tau_2 = w_2$ only, respectively. Partial differentials are expressed as $M_{(a)} = \partial_a M$, $M_{(ab)} = \partial_a \partial_b M$, $H_{a(\kappa)} = \partial_\kappa H_a$, $H_{a(\kappa\lambda)} = \partial_\kappa \partial_\lambda H_a$, and so on, where $H = (H_a)$. The Einstein summation convention is adopted; e.g. $\theta^i t_i$ means $\sum \theta^i t_i$. The Fisher information metrics and the skewness tensors are given by $g_{ab} = M_{(ab)}$, $g_{a\kappa} = 0$, $S_{abc} = M_{(abc)}$, $S_{ab\kappa} = 0$, $S_{a\kappa\lambda} = H_{a(\kappa\lambda)}$. The α-connections are the following: $\Gamma^{(\alpha)}_{abc} = ((1-\alpha)/2)M_{(abc)}$, $\Gamma^{(\alpha)}_{\kappa\lambda a} = ((1-\alpha)/2)H_{a(\kappa\lambda)}$, $\Gamma^{(\alpha)}_{\kappa a\lambda} = \Gamma^{(\alpha)}_{a\kappa\lambda} = -((1+\alpha)/2)H_{a(\kappa\lambda)}$, $\Gamma^{(\alpha)}_{ab\kappa} = \Gamma^{(\alpha)}_{a\kappa b} = \Gamma^{(\alpha)}_{\kappa ab} = 0$.

4. Relation between s and $\bar{t}_2$. Let $\hat{w} = (\hat{\theta}_1, \hat{\tau}_2)$ be the maximum likelihood estimator and let $\tilde{w} = \sqrt{n}(\hat{w} - w)$. It is well-known that the asymptotic distribution of $\tilde{w}$ is normal with mean zero and covariance g^{ij}. We see that $g_{a\kappa} = 0$ and $g_{ab} = M_{(ab)}(\theta_1)$. This means that $\hat{\theta}_1$ and $\hat{\tau}_2$ are asymptotically independent and the asymptotic distribution of $\hat{\theta}_1$ depends on θ_1 only. We know that the likelihood equation is $\bar{t} = \tau(\hat{w})$ and this can be rewritten as follows: $(\bar{t}_1, \bar{t}_2) = (\tau_1(\hat{w}), \tau_2(\hat{w})) = (m(\hat{\theta}_1) + H(\hat{\tau}_2), \hat{\tau}_2)$. Since $s = \bar{t}_1 - H(\bar{t}_2)$, this is equivalent to $s = m(\hat{\theta}_1)$, $\bar{t}_2 = \hat{\tau}_2$. It holds that $\partial m/\partial \theta_1' = (M_{(ab)}) = (g_{ab})$ is positive-definite. Therefore, it is seen that s and $\bar{t}_2$ are asymptotically independent and the asymptotic distribution of s depends on θ_1 only. This property is similar to the reproductivity. Furthermore, higher-order asymptotic relations between $\hat{\theta}_1$ and $\hat{\tau}_2$ imply that the same property as in the asymptotic case holds up to the higher-order.

5. Asymptotic properties. This section shows some theorems.

THEOREM 5.1 *Consider the exponential model. This possesses a τ-parallel foliation if and only if the conditional maximum likelihood estimator is asymptotically efficient and its asymptotic covariance depends on θ_1 only.*

THEOREM 5.2 *It holds that*

$$\begin{aligned} E(\hat{\theta}^a - \theta^a) &= -\frac{1}{2n}(M_{(bce)}g^{bc}g^{ea} + H_{e(\kappa\lambda)}g^{\kappa\lambda}g^{ea}) + O\left(n^{-2}\right), \\ E(\check{\theta}^a - \theta^a) &= -\frac{1}{2n}M_{(bce)}g^{bc}g^{ea} + O\left(n^{-2}\right), \end{aligned}$$

Therefore, the bias of $\check{\theta}^a$ is more robust than that of $\hat{\theta}^a$ because the first-order term of the latter does not depend on $H_{e(\kappa\lambda)}g^{\kappa\lambda}g^{ea}$.

THEOREM 5.3 *The difference between $\check{\theta}^a$ and $\hat{\theta}^a$ can be expressed as*

$$\check{\theta}^a - \hat{\theta}^a = \frac{1}{2n}Q^a(\theta_1) + O_p\left(n^{-3/2}\right), \qquad \check{\theta}^a - \hat{\theta}^a = \frac{1}{2n}Q^a(\hat{\theta}_1) + O_p\left(n^{-2}\right),$$

where $Q^a(\theta_1) = H_{e(\kappa\lambda)}g^{\kappa\lambda}g^{ea}$.

THEOREM 5.4 *Let $\check{\theta}^a_*$ and $\hat{\theta}^a_*$ be the bias-corrected versions of $\check{\theta}^a$ and $\hat{\theta}^a$, respectively. Then, we have*

$$\check{\theta}^a_* - \hat{\theta}^a_* = O_p\left(n^{-2}\right).$$

空間および時空間モデルのベイズ的モデリング：モデル選択と残差解析

統計数理研究所　尾形良彦

地域によって地震活動パタンが違うことを考慮して、時空間 ETAS モデル (Ogata, 1998) の特性パラメータ7つのうち主要な5つを位置(経度、緯度)の関数として**ベイズ的時空間ETASモデル**に拡張した。この関数を日本領域とその周辺部の地震発生位置を頂点とする2次元デロネ分割上の一次スプライン関数で表現し、平滑化のためにこの5つの関数の傾き(微分)が全体的に小さくなるような事前正規分布を導入し、ABIC 法によって事後分布から最適推定パラメタ (posterior mode)を求めるアルゴリズムを実用化した。

気象庁震源データを用いて3/4世紀の地震活動(マグニチュード 5 以上)の計測と以下の様な点過程残差解析を行なった。ABIC 法で求めた最適なベイズ的時空間ETASモデルの条件付強度関数と、これによって予測される各地の地震活動度と実際の地震発生数を比べる**比モデル発生率関数**(時空間上の関数)をかけ合わせた条件付強度関数で尤度を定義し、大規模ベイズモデルによる平滑化問題を考え推定した。比モデル発生率関数は地震の震源(発生時刻を含む)を頂点とするような3次元デロネ分割上の一次スプライン関数で表現し、平滑化の為この傾き(微分)が全体的に小さくなるような事前分布を導入し ABIC 法によって事後分布から最適推定パラメタ (posterior mode) を求めた。

この様にして推定した相対的地震発生率関数を3次元画像解析ソフト AVSによって可視化したところ、いくつかの巨大地震の震源域周辺で前兆現象であるとみられている相対的静穏化現象や地震空白域が見られた。

デロネ分割上のスプライン関数は、第一に集中性の強い点配置上や次元の高い空間上の統計的モデルのパラメタ化に適していることが実証されただけでなく、広く時空間モデル開発の実施に必要な一歩を記したものと考える。

文献

Ogata, Y. (1998). Space-time point-process models for earthquake occurrences, *Ann. Inst. Statist. Math.*, 50, pp. 379-402.

Ogata, Y. (2000). Spatially heterogeneous space-time modeling for seismic activity and space-time residual analysis for point processes. presented at the Annual Meeting of the Statistical Society of Japan held at Hokkaido University, 26-28 July 2000, and also at an invited talk at the Symposium on Analysis of Spatio-Temporal Data, held at Australian National University, 11-14 September 2000.

観測次元が大規模な状態空間モデルの推定

川崎 能典・統計数理研究所 予測制御研究系

1. モデルとアルゴリズム

状態ベクトル $(p\times 1)$ を x_t，その一期先予測の分散共分散行列 $(p\times p)$ を $\Sigma_{t|t-1}$ とする．更に観測ベクトル $(N\times 1)$ を $z_t(t=1,\dots,T)$ と書く．いま推定したいモデルに

$$x_{t+1} = F_t x_t + G_t w_t,\ w_t \sim \mathrm{N}(0, Q_t) \quad (1)$$

$$z_t = H_t x_t + v_t,\ v_t \sim \mathrm{N}(0, R_t) \quad (2)$$

という状態空間表現が与えられているとしよう．ここで w_t と v_t は互いに独立な正規分布にしがたい，その共分散行列をそれぞれ Q_t，R_t と記す．このとき一期先予測の共分散行列は尤度関数は

$$\Lambda_{t|t-1} = H_t \Sigma_{t|t-1} H_t' + R_t \quad (3)$$

と表される．モデルの最尤推定には $\Lambda_{t|t-1}^{-1}$ の評価と $|\Lambda_{t|t-1}|$ の評価が必要であるが，Kalman filter の利用を前提にすれば，観測次元が高い場合には巨大な行列の反転と三角分解が必要となる．このとき (3) に行列の逆転公式 (Anderson and Moore (1979), § 6.3) を適用すると，

$$\Lambda_{t|t-1}^{-1} = R_t^{-1} - R_t^{-1} H_t \left[\Sigma_{t|t-1}^{-1} + H_t' R_t^{-1} H_t\right]^{-1} H_t' R_t^{-1} \quad (4)$$

を得る．これより，仮に観測ベクトルの次元が大きくても，R_t^{-1} が簡単に与えられる限りは (4) によって $\Lambda_{t|t-1}^{-1}$ を効率よく計算できるので，二次形式の計算は容易であると言える (情報量フィルタ)．しかし，モデルの最尤推定には行列式の評価が伴うので，単に情報量フィルタを適用しただけでは問題は解決しない。以下で、この要請を満たす 2 つのモデルを考える．まず

$$R_t = (1-\rho)\omega^2 I_N + \rho\omega^2 \iota_N \iota_N' \quad (5)$$

というパターン行列の利用が考えられる．ここで ι_N は要素がすべて 1 の $N\times 1$ ベクトルである．このとき逆行列は明示的に与えられ，本質的に状態次元の正方行列の逆転で二次形式は計算される．行列式は明示的に書き表せないが，特殊形を利用して対角成分と準対角成分の計算のみで済む．(Kawasaki et al. (2000)) 一方，Kawasaki et al. (1998) は，回帰の design matrix H_t を利用し，次の形で R_t を与えた．

$$R_t = H_t D_2 H_t' + \sigma^2 I \quad (6)$$

ただしここで，$D_2 = \mathrm{diag}(\lambda_1^2,\dots,\lambda_p^2)$ である．$\Phi_{t|t-1} = \Sigma_{t|t-1} + D_2$ として，一期先予測の分散共分散行列を書き下して逆転公式を適用すると，

$$\Lambda_{t|t-1}^{-1} = \sigma^{-2} I - \sigma^{-4} H_t \left[\Phi_{t|t-1}^{-1} + \sigma^{-2} H_t' H_t\right]^{-1} H_t' \quad (7)$$

を得る．一方，行列式については，

$$\left|\sigma^2 I_N + \Delta_{t|t-1}\Delta_{t|t-1}'\right| = \sigma^{2(N-p)} \left|\sigma^2 I_p + \Delta_{t|t-1}'\Delta_{t|t-1}\right| \quad (8)$$

となることがわかる．

2. 応用: 大規模マルチファクターモデル

マルチファクターモデルとは，株式収益率 (往々にしてインデックスに対する相対超過収益) を，財務データやテクニカル指標といった「ファクター」で説明し，それを外挿して収益を予測するモデルで，主に自己売買を念頭に置いた分析法と言える．実務的なプロシジャの定義は様々であるが，統計モデルとしては時変係数を持つベクトル値回帰モデル定式化が可能であるので，(1) (2) の枠内で表現可能である．このとき H_t は，いわゆるファクターのデータが並んだ説明変数行列となり (添字は t であるが実際には $t-1$ 期末に得られるデータから構成される)，R_t の取り方の違いが，ここで考慮するモデル間の相違になる．

- $R_t = \sigma^2 I_N$ (OLS，時間不変)
- $R_t = (1-\rho)\omega^2 I_N + \rho\omega^2 \iota_N \iota_N'$ (パターン行列，時間不変)
- $R_t = H_t D_2 H_t' + \sigma^2 I_N$，$D_2 = \mathrm{diag}(\lambda_1^2, \ldots, \lambda_p^2)$，(Temporal Effect，時変)

1985 年 1 月から 1997 年 12 月までの東証一部上場銘柄の株価データ (月次) を用いた分析の結果を要約する．OLS 型のモデルは平滑化事前分布が全く効いておらず，時点ごとにクロスセクション回帰して得られた係数列と全く違いがないため，予測モデルとしては不適当である．Elton-Grüber タイプの共分散行列を与えた場合は，業種によっては滑らかなファクター係数が得られるものの，一般に銘柄数の多い業種になればなるほど OLS の結果に近づいてゆく傾向がある．これは，観測方程式における共分散行列の非対角項をたったひとつのパラメータで表現しようとする方法の限界が，観測次元の拡大に伴って顕著になっているものと考えられる．Temporal Effect モデルは，最も flexible な形である分パラメータも Elton-Grüber 型より多いが，推定に要する時間は殆ど同じである．情報量規準の値を見ても，概して Temporal Effect モデルが支持される．ファクターごとに Kalman gain の時間変化を眺めてみると，共分散構造のモデリングの違いは，予測誤差の振り分け方の違いに表れていることがわかる．リバーサルファクターの係数変化は日本の株式市場における投資家の期待の変化を示唆しており興味深い．

参考文献

[1] Anderson, B. D. O. and J. B. Moore (1979) *Optimal Filtering*, Englewood Cliffs, New Jersey, Prentice-Hall.

[2] Kawasaki, Y., S. Sato and S. Tachiki (1998) Smoothness prior approach to estimate large scale multifactor models, *ISM Research Memorandum No. 714*, The Institute of Statistical Mathematics, Tokyo.

[3] Kawasaki, Y., S. Sato and S. Tachiki (2000) Vector-valued multiple regression model with time varying coefficients and its application to predict excess stock returns, *Proceedings of IEEE/IAFE/INFORMS Conference on Computational Intelligence for Financial Engineering*, pp. 162–165.

Geometric Consideration on Simple Loop Alternative: 非正則多次元正規片側検定の実用化を目指して

Akio Kudô *,Yoshiro Yamamoto†and SUDIARSA, I Made ‡

初めに原点を帰無仮説 H_0 とし，原点を頂点とする閉凸多面錘 I を対立仮説 H_1 とする検定の概説を行う．

1. **The formulation:** $\boldsymbol{x}$ を p 次元正規確率変数で，分散行列は，一般性を失うことなく，単位行列とし，平均ベクトル $\boldsymbol{\theta}$ については，$\boldsymbol{\theta}=\mathbf{0}$ underH_0 and $\boldsymbol{\theta}\in I$ underH_1 とする．

2. **The maximum likelihood estimate and the test statistic.** $H_0\cup H_1$ のもとでの MLE $\hat{\boldsymbol{\theta}}$ は $\boldsymbol{X}\in I\Longrightarrow\hat{\boldsymbol{\theta}}=\boldsymbol{X}$ $\quad\boldsymbol{X}\notin I\Longrightarrow\hat{\boldsymbol{\theta}}=\hat{\boldsymbol{X}}$ ここで，$\hat{\boldsymbol{X}}$ は $\boldsymbol{X}(\notin I$ から I への最短距離をみたす I の点で，これは I の境界集合 (Boundary) に属する．

3. **Classification of boundary points** 閉凸錐 I は一次元の稜，二次元の面，さらには高い次元の部分空間に含まれる境界がある．今の場合は 2^{k-1} 個の可能性があるがその全部ではない例がある．

   I の内点は $I^{o}=\{\boldsymbol{\theta}\,|(\boldsymbol{\theta},\boldsymbol{a}_i)>0,\text{ for }i=1,2,\cdots,k\}$ で定まる．，不等式の幾つかが等式になるのが境界で全部が等式になるのは頂点である原点である．部分空間にはそれと I との共通部分は原点 $\mathbf{0}$ のみである場合もあり，ある部分空間に入る $\boldsymbol{a}_i$ の個数がその空間の次元よりも多い場合もある．

4. **The test statistic.** 検定統計量は $\|\hat{\boldsymbol{X}}\|^2$ で記号 $\bar{\chi}^2$ で表す．

5. **The null distribution of the test statistic.** カイ自乗分布の混合であり，その係数は I の形状により定まる．

$$
\begin{aligned}
&\Pr\left(\bar{\chi}^2\le a\right)\\
&=\Pr\left(\boldsymbol{X}\in I^{*}\right)+\sum_{l=1}^{k-1}P(l,k)\ \Pr\left(\chi^2\le a\right)+\Pr\left(\boldsymbol{X}\in I\right)\Pr\left(\chi^2_p\le a\right)\\
&=\Pr\left(\boldsymbol{X}\in I^{*}\right)+\sum_{M}\Pr\left(\hat{\boldsymbol{\theta}}\in B_M\right)\Pr\left(\chi^2_{dim(B_M)}\le a\right)+\Pr\left(\boldsymbol{X}\in I\right)\Pr\left(\chi^2_p\le a\right)
\end{aligned}
$$

   ここで，$\sum_M$ は $k(>p)$ 個のベクトル $\boldsymbol{a}_1,\boldsymbol{a}_2,\ldots,\boldsymbol{a}_k$ から出来る R^p の真部分空間についての和で，B_M は部分空間内の I の境界で，$\chi^2_{dim(B_M)}$ は部分空間 B の次元と等しい自由度を持つカイ自乗分布に従う確率変数で，I^{*} は I の双対 (dual) 錐である．

6. **History** 最初の論文は [2] であるが論旨不明な部分が多い．[1] では B. J. Bartholomew の論文と辻褄があうので認めている．[3] 論旨不明な部分は改善されたが記号に誤植が多い．一番新しい結果は [5] である．[6] の文献リストには [2], [3], [5] のすべてが記載されているが，著者達は読んでいないで引用した様である．

*Faculty of Economics and Information Science, Hyogo University, 2301, Shinzaike, Hiraoka-Cho, Kakogawa 675-01, Japan

†Graduate School of Engineering, Hokkaido University, kita 13 Nishi 8, Kita-ku, Sapporo 060-8628, Japan

‡Graduate School of Economics and Information Science, Hyogo University, 2301, Shinzaike, Hiraoka-Cho, Kakogawa 675-01, Japan

上の式中の $\Pr(\boldsymbol{X} \in I)$, $\Pr\left(\hat{\theta} \in B_M\right)$ 及び $\Pr(\boldsymbol{X} \in I^*)$ をすべて計算して和が 1 になること，及び [4] に与えられた等式より，計算精度を調べた．その計算の経過で明らかになった実用上の問題点について報告する．

Matrix Order とは行列の要素からなる空間である [6].

$$X = \begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1q} \\ x_{21} & x_{22} & \cdots & x_{2q} \\ \vdots & \vdots & \vdots & \vdots \\ x_{p1} & x_{p2} & \cdots & x_{pq} \end{pmatrix}, \ \Theta = \begin{pmatrix} \theta_{11} & \theta_{12} & \cdots & \theta_{1q} \\ \theta_{21} & \theta_{22} & \cdots & \theta_{2q} \\ \vdots & \vdots & \vdots & \vdots \\ \theta_{p1} & \theta_{p2} & \cdots & \theta_{pq} \end{pmatrix} \tag{1}$$

この行列の各行と列に順序関係 $\geq$ と $\rotatebox{90}{$\leq$}$ とを導入する．

$$\left\{ \begin{array}{ccccccc} \theta_{11} & \geq & \theta_{12} & \geq & \cdots & \geq & \theta_{1q} \\ \rotatebox{90}{$\leq$} & & \rotatebox{90}{$\leq$} & & & & \rotatebox{90}{$\leq$} \\ \theta_{21} & \geq & \theta_{22} & \geq & \cdots & \geq & \theta_{2q} \\ \rotatebox{90}{$\leq$} & & \rotatebox{90}{$\leq$} & & & & \rotatebox{90}{$\leq$} \\ \vdots & & \vdots & & \cdots & & \vdots \\ \rotatebox{90}{$\leq$} & & \rotatebox{90}{$\leq$} & & & & \rotatebox{90}{$\leq$} \\ \theta_{p1} & \geq & \theta_{p2} & \geq & \cdots & \geq & \theta_{pq} \end{array} \right. \tag{2}$$

Matrix Order の場合でも，多次元片側検定の形に変換することは容易である．上の (p, q) 行列の場合では，$(p-1) \times (q-1)$ 次元の非正則正規分布になる．

$p = q = 2$ の場合は simple loop altgernative と呼ばれる．この最も簡単な場合から出発して実用化を目指した研究を行っている．

References

[1] Barlow, R. E., Bartholomew, D. J., Bremner, J. M. and Brunk, H . D. (1972). Statistical Inference under Order Restrictions, 2nd ed. New York:, John Wiley.

[2] Kudô, A. (1963). A multivariate analogue of one-sided test. Biometrika 50, 403–18.

[3] Kudô, A. and Choi, J. R.(1975). A generalized multivariate analogue of the one-sided test. Mem. Fac. Sci., Kyushu Univ. Ser. A 29, 303–328.

[4] mcMullen, P. (1975). Non-linear angle-sum relations for polyhedral cones and polytopes. Math. Proc. Camb. Phil. Soc., 78, 247–61.

[5] Nomakuchi, K. (1983). The likelihood ratio test of normal mean with hypothesis determined by a convex polyhedral cone and the monotonicity of its power function. Mem. Fac. Sci. Kyushu Univ. Ser.A 37, 195–204.

[6] Robertson, T., Wright, F. T. and Dykstra, R. L. (1988). Order Restricted Statistical Inference, John Wiley. New York.

[7] Shi, N.-Z and Kudô, A. (1987). The most stringent somewhere most powerful one sided test of the multivariate normal mean. Memoirs of Faculty of Science, Kyushu Univ. Ser. A 41, 37–44.

[8] Sun, H.-J. (1988). A Fortran subroutine for computing normal orthant probabilities of dimensions up to nine. Commun. Statist.-Simula. 17(3), 1097–1111.

[9] Yamamoto, Y., A. Kudô, A. and Ujiie, K. (1997). Computation of the Statistics and the Significance Probability in the Multivariate analogue of the one-sided test. Journal of the Japanese Society of Computational Statistics 10, 89–97.

順序制約の下での多次元正規分布の平均ベクトルの均一性の検定について

九州芸術工科大学　笹渕 祥一

島根県庁　斎藤 寛之

鳥栖高校　田中 孝司

日立ソフトウェア　塚本 武

1. 共分散行列が既知で次元が３以上の場合の種々の検定法の比較

$\boldsymbol{X}_i$ $(i=1,2,\ldots,k)$ を互いに独立に平均ベクトル$\boldsymbol{\mu}_i$, 分散共分散行列Σのp次元正規分布に従う確率ベクトルとする. Σは既知として, 次の検定問題を考える.

帰無仮説H : $\boldsymbol{\mu}_1=\boldsymbol{\mu}_2=\cdots=\boldsymbol{\mu}_k$

対立仮説K : $\boldsymbol{\mu}_1\le\boldsymbol{\mu}_2\le\cdots\le\boldsymbol{\mu}_k$　　(ただし少なくとも１つは等号でない.)

ここで, $\boldsymbol{\mu}_i\le\boldsymbol{\mu}_j$ は $\boldsymbol{\mu}_j-\boldsymbol{\mu}_i$ の各成分がいずれも非負であることを意味する.

Sasabuchi, Inutsuka and Kulatunga (1983) は, この問題に対する尤度比検定統計量 $\chi^2_{k,p}$ を導いた. $p=1$の場合と同様に, 尤度比検定は優れた検出力をもつことが予想されるが, その帰無分布が特別な場合にしか得られていないため, 与えられたデータの有意確率や与えられた有意水準に対する棄却点を得るためには, シミュレーションを行なわなければならない.

この検定問題に対する最も簡単な検定法としては, Abelson and Tukey(1963) のcontrast検定あるいはSchaafsma and Smid(1966) の最近迫局所最強力検定を適用することが考えられる. これは, 検定統計量も, その分布(１次元正規分布!)も, 計算が極めて簡単であるという長所がある.

Nomakuchi and Shi(1988) は, contrast検定とKudô(1963) の多次元片側検定とを組み合わせた検定を提案し, $p=2$の場合に, シミュレーションによって, その検出力を尤度比検定および最近迫局所最強力検定と比較した.

この節では, pが３以上の場合に, 尤度比検定, 最近迫局所最強力検定, およびNomakuchi − Shi の検定を比較検討した結果を報告し, またこれらの検定の性質に関する若干の理論的考察を与えた. この「講演要旨報告」では, 詳細は省略する.

2. 共分散行列が共通で未知の場合

$\boldsymbol{X}_{i1},\ldots,\boldsymbol{X}_{iN_i}$ を平均ベクトル $\boldsymbol{\mu}_i$, 分散共分散行列 Σ の p 次元正規分布からの大きさ N_i の無作為標本とする $(i=1,\ldots,k)$. ただし, Σ は未知とし, $N_1+\cdots+N_k>p+k$ とする. 前節と同じ検定問題を考える.

Σ が未知の場合, 検定問題 H vs. K に対する尤度比検定は今のところ求められていない. そこで, Σ が既知の場合の尤度比検定統計量 $\chi^2_{k,p}$ において Σ をその推定量で置き換えた検定統計

量Tによる次の検定を提案する.

$$\mathrm{T} = \sum_{i=1}^{k} N_i(\widehat{\mu}_i - \overline{X})' S^{-1} (\widehat{\mu}_i - \overline{X}) \geq c \Rightarrow \text{ 帰無仮説 H を棄却}$$

ここで, $\overline{X}_i = \frac{1}{N_i}\sum_{j=1}^{N_i} X_{ij}, i = 1, \ldots, k$, $\overline{X} = \left(\sum_{i=1}^{k} N_i\right)^{-1} \sum_{i=1}^{k} N_i \overline{X}_i$,
$S = \sum_{i=1}^{k}\sum_{j=1}^{N_i}(X_{ij} - \overline{X}_i)(X_{ij} - \overline{X}_i)'$, $(\widehat{\mu}_1, \ldots, \widehat{\mu}_k)$ は $N_1 S^{-1}, \ldots, N_k S^{-1}$ を重みとする $\overline{X}_1, \ldots, \overline{X}_k$ の多変量単調回帰, c は有意水準によって定まる正の定数.

$\boldsymbol{\Sigma}$が未知の場合に, 統計量Tを用いて H vs. K の検定を行うとき, 与えられたデータに対する有意確率を求めるため, あるいは, 有意水準 α の検定の棄却点を求めるためには, 定数cに対して

$$\sup_{\boldsymbol{\Sigma}} \sup_{\mathrm{H}} \mathrm{P}_{\boldsymbol{\mu}, \boldsymbol{\Sigma}}(\mathrm{T} \geq c) \qquad (1)$$

の値を求める必要がある. ここで, $\mathrm{P}_{\boldsymbol{\mu}, \boldsymbol{\Sigma}}$ は母数が$\boldsymbol{\mu} = (\mu_1, \ldots, \mu_k)$と$\boldsymbol{\Sigma}$であるときの確率, $\sup_{\mathrm{H}}$ は $\mu_1 = \cdots = \mu_k$ なる $\mu_1, \cdots, \mu_k$ に関する上限を表わす.

Tの分布に関して, 次の定理が成り立つ.

<u>定理1</u>　帰無仮説Hの下でのTの分布は, $\boldsymbol{\mu}$に依存しない.

この定理により, Hの下でのTの分布に関する計算では, $\boldsymbol{\mu} = 0$とおいてよいが, 未知の共分散行列$\boldsymbol{\Sigma}$には依存するので, このままでは(1)の計算は困難である.

そこで, 次の統計量T*を導入する.

$$\mathrm{T}^* = \sum_{i=1}^{k} N_i(\overline{X}_i - \overline{X})' S^{-1} (\overline{X}_i - \overline{X}) - \frac{1}{s_{11}} \sum_{i=1}^{k} N_i (\overline{X}_{i1} - \widehat{\mu}_{i1})^2$$

ここで, $\overline{X}_{i1}$は$\overline{X}_i$の第1成分, s_{11}はSの$(1,1)$成分であり, $(\widehat{\mu}_{11}, \ldots, \widehat{\mu}_{k1})$ は $N_1, \ldots, N_k$を重みとする $\overline{X}_{11}, \ldots, \overline{X}_{k1}$の単調回帰である.

T*の分布に関しては, 次の定理が成り立つ.

<u>定理2</u>　帰無仮説Hの下でのT*の分布は, $\boldsymbol{\mu}$にも$\boldsymbol{\Sigma}$にも依存しない.

この定理により, 帰無仮説Hの下でのT*の分布を考える際には, $\boldsymbol{\mu} = 0$, $\boldsymbol{\Sigma} = \mathrm{I}_p$ とおいてよい.

帰無仮説Hの下での検定統計量Tの上側確率の上限は, T*を用いて, 次の定理で与えられる.

<u>定理3</u>

$$\sup_{\boldsymbol{\Sigma}} \sup_{\mathrm{H}} \mathrm{P}_{\boldsymbol{\mu}, \boldsymbol{\Sigma}}(\mathrm{T} \geq c) \;=\; \sup_{\boldsymbol{\Sigma}} \mathrm{P}_{0, \boldsymbol{\Sigma}}(\mathrm{T} \geq c) \;=\; \mathrm{P}_{0, \mathrm{I}_p}(\mathrm{T}^* \geq c)$$

この定理により, (1)を求めるには, $\mathrm{P}_{0,\mathrm{I}_p}(\mathrm{T}^* \geq c)$ を求めればよい. T*の分布はわかっていないが, T*そのものの計算は簡単であり, $\boldsymbol{\mu} = 0$, $\boldsymbol{\Sigma} = \mathrm{I}_p$の場合の確率であるから, 標準正規乱数を発生させてシミュレーションを行うことにより, $\mathrm{P}_{0,\mathrm{I}_p}(\mathrm{T}^* \geq c)$ の近似値を求めることができる.

Ψ-DIVERGENCE and Ψ-MAXIMUM LIKELIHOOD

Ψ-ダイバージェンス と Ψ-最尤法

統計数理研究所 江口真透,　大阪大学・人間科学 狩野 裕

A new class of divergence functionals, including Kullback-Leibler divergence, over probability densities is proposed by generative functions $\Psi(z)$. The idea is closely related with that of deviance in generalized linear regression. The maximum Ψ-likelihood estimator is defined by minimizing a Ψ-divergence functional based on data. If Ψ is an identity function, or $\Psi(z) = z$, then it reduces to the relation of the usual maximum likelihood estimator and the Kulback-Leibler divergence. From the information-geometric point of view, it is shown that a Ψ-divergence leads to a simple dualistic geometry similar to the α-geometry developed by Amari.

It is shown that the Ψ-likelihood method is a robust estimation procedure for a general statistical model $f(y, \theta)$ under some simple conditions on the generative function $\Psi(z)$. A typical example of $\Psi(z)$ to guarantee robustness is a log-sigmoidal function. The estimation based on the Ψ-likelihood method is made on the likelihood principle, where each observation is weighted with the magnitude of likelihood evaluated at the observation. Samples with low likelihood are likely to be regarded as outliers and are downweighed. Use of the likelihood enables us to easily construct a robust estimation procedure for many types of statistical models. Several informative examples are provided, and relationships with existing robust estimation methods are studied. In principle the Ψ-likelihood method can be applied to almost all models in addition to the Ψ-information criterion. We take logistic regression classifications and principal/independent component analyses as examples to show the usefulness of the new method proposed. We end with a small simulation study.

Assume a statistical model $\{f(y,\theta)|\theta \in \Theta\}$ for given observations $\{y_j : j = 1, \ldots, n\}$, where $f(y, \theta)$ is a density function with respect to some carrier measure $d\nu(y)$, and Θ is a parameter space of θ. The log-likelihood function is defined by the sum of log-densities over the observations:

$$\ell(\theta) = \frac{1}{n}\sum_{i=1}^{n} \ell(y_j, \theta) \tag{1}$$

with $\ell(y, \theta) = \log f(y, \theta)$. The likelihood principle has been explored as a universal methodology for analyzing data since the beginning of the history of statistics. Without loosing the universality, we would like to make an improvement of the likelihood method to cover the situation where some outliers may be included in the data set to be statistically analyzed.

Let Ψ be an increasing and convex function defined on the R^1. We use the function Ψ

to propose a class of variants of the maximum likelihood estimator (MLE) as follows:

$$\hat{\theta}_\Psi = \underset{\theta\in\Theta}{\operatorname{argmax}} \left\{ \frac{1}{n} \sum_j \Psi(\ell(y_j, \theta)) - \int \Psi^*(\ell(y,\theta)) d\nu(y) \right\}, \tag{2}$$

where

$$\Psi^*(z) = \int_0^z \exp(s)\psi(s) ds \tag{3}$$

with $\psi = (\partial/\partial s)\Psi$. The class is characterized as a collection of all increasing and convex functions Ψs. We will refer to Ψ as a generative function and to the estimation procedure as a Ψ-likelihood method. The second term in the right hand side of (2) comes from the unbiased condition of the corresponding estimating equation, as will be discussed later. It is observed that if $\Psi(z) = z$, the Ψ-likelihood method reduces to the method of maximum likelihood. In a subsequent discussion we will study performance of the Ψ-likelihood method with $\Psi(z) \neq z$.

It is known that the method of maximum likelihood is the sample counterpart of minimization of the Kulback-Leibler distance between a statistical model and a true distribution. The Ψ-likelihood estimatition also have such a distance. For frequencies $f(y)$ and $g(y)$ (densities with respect to $d\nu(y)$), we consider the following distance between the frequencies:

$$\delta_\Psi(f, g) = \int_f^g \frac{g-\mu}{\mu} \psi(\log \mu) d\mu. \tag{4}$$

The distance is closely related with the deviance function introduced by Wedderburn (1974) if we read $\mu/\psi(\log\mu)$ as a variance function in the context of generalized linear model. The essential difference is that we view δ_Ψ not as a function of a response and the corresponding mean but as a fuctional of frequencies. The Ψ-divergence functional is then given as

$$D_\Psi(f, g) = \int_{\mathcal{Y}} \delta_\Psi(f(y), g(y)) d\nu(y) \tag{5}$$

By definition $D_\Psi(f, g)$ is nonnegative and equals 0 if and only if $g = f$. In particular if $\Psi(z) = z$, then D_Ψ reduces to the Kullback-Leibler divergence

$$KL(f, g) = \int_{\mathcal{Y}} g(y) \log \frac{g(y)}{f(y)} d\nu(y). \tag{6}$$

The Ψ-likelihood estimation is the empirical version of (4) after omiting terms involoving only g.

References

Wedderburn, R. W. M. (1974). Quasi-likelihood functions, generalized linear models, and the Gauss-Newton method. *Biometrika,* **61,** 439–447.

スケール変換をしたブートストラップを用いる領域の検定

統計数理研究所　下平英寿

1 はじめに

なめらかな境界をもつ領域を検定する問題を考察する。議論の便宜上簡単のためデータ y は K 次元ベクトルとし、これが多次元正規分布

$$y \sim N_K(\mu, I) \tag{1}$$

に従うとする。帰無仮説は未知母数 μ がある領域 $\mathcal{R}$ に含まれること、すなわち $H_0 : \mu \in \mathcal{R}$ とし、$\mathcal{R}$ のインデックス関数 $I_{\mathcal{R}}(y)$ だけしか利用できないものとする。これは classification やモデル選択にしばしば出てくる問題である。$\mathcal{R}$ の境界 $\partial\mathcal{R}$ が十分滑らかのとき、任意の $\mathcal{R}$ の形状に簡単に適用できる方法を提案する。この方法は近似的に不偏な検定になっており境界上でほぼ相似である (Lehmann p. 135, 1986)。検定の p-値を $p = p(y)$ と書くと、$\mu \in \partial\mathcal{R}$ のとき

$$\Pr\{p(y) < \alpha\} = \alpha, \quad 0 < \alpha < 1 \tag{2}$$

を近似的に満たすことが目標である。

データ y は実はサイズ n のオリジナルデータからパラメタ推定などによって得た統計量を $\sqrt{n}$ 倍するスケール変換したものと考える。十分大きな n において (1) のモデルは良い近似になることが多い。いわゆるノンパラメトリックブートストラップは

$$y_1^*, \ldots, y_B^* \sim N_K(y, I) \tag{3}$$

となる。素朴な p-値構成法として、次のものがしばしば用いられる (Felsenstein 1985):

$$\tilde{p}(y) = B^{-1} \sum_{j=1}^{B} I_{\mathcal{R}}(y_j^*).$$

ところがこの方法は (2) の誤差が first-order $O(n^{-1/2})$ であまり良くない。Efron et al (1996) はこれをバイアス補正することにより seconr-order $O(n^{-1})$ の検定を構成した。我々はそれより簡単な計算法で、かつ third-order $O(n^{-3/2})$ の方法を与える。

2 Signed distance と Curvature

データ y を $\partial\mathcal{R}$ へ射影した点を $\hat{\mu}_0$ とする。これは境界上で y に最も近い点であり、$\hat{\mu}_0$ から y への符号付距離 ($\mathcal{R}$ の外側のとき正) を signed distance と呼び x_0 と書く。$\hat{\mu}_0$ の近傍で $\partial\mathcal{R}$ を 2 次曲面 $u = v'\hat{\mathbf{d}}v$ で近似し、局所的な尤度比検定を構成する。ただし $\hat{\mathbf{d}}$ は曲率を表す $(K-1)\times(K-1)$ 行列とし、$\hat{d}_1 = \mathrm{trace}(\hat{\mathbf{d}})$, $\hat{d}_2 = \mathrm{trace}(\hat{\mathbf{d}}^2)$, $c := \hat{d}_1 - x_0\hat{d}_2$ とおく。Efron and Tibshirani (1996) は bias 補正した p-value として

$$\hat{p}(y) = 1 - \Phi\left(\frac{x_0 - \hat{d}_1}{1 - \hat{d}_2}\right) = 1 - \Phi(x_0 - c) + O_p(n^{-3/2}), \tag{4}$$

を与え、これが (2) を $O(n^{-3/2})$ の誤差で満たすことを示した。一方、素朴な p-value は

$$\tilde{p}(y) = 1 - \Phi(x_0 + c) + O_p(n^{-3/2}) + O_p(B^{-1/2}) \tag{5}$$

であることも示される。c は $O_p(n^{-1/2})$ だから $\tilde{p}=\hat{p}+O_p(n^{-1/2})$ であり、$\tilde{p}$ は (2) を $O(n^{-1/2})$ の誤差で満たすこともわかる。

問題はどうやって x_0 と c を計算するかである。Efron et al (1996) は $\hat{\mu}_0$ を数値的に求めた上で $N_K(\hat{\mu}_0, I)$ に従うブートストラップサンプルを生成し、それから $\hat{d}_1 \approx \Phi^{-1}(1-\tilde{p}(\hat{\mu}_0))$ を推定し $c=\hat{d}_1+O_p(n^{-1})$ の代わりに用いた。これだと (2) は $O(n^{-1})$ の誤差を持つ。我々は、$\hat{\mu}_0$ を明示的に計算することなく、かつ $O(n^{-3/2})$ の誤差の方法を与える。キーになるアイデアは、スケール変換したブートストラップを用いることである。各リサンプルにおけるデータサイズを m にとり $\sigma=\sqrt{n/m}$ とおくとブートストラップは $N_K(y, \sigma^2 I)$ に従う。すると $x_0 \to \sigma^{-1}x_0$, $\hat{\mathbf{d}} \to \sigma\hat{\mathbf{d}}$ の置き換えが起こり、(5) は

$$\Phi^{-1}(1-\tilde{p}(y)) \approx \sigma^{-1}x_0+\sigma c \tag{6}$$

と書ける。これより直ちに x_0 と c を計算する方法が得られる。

3 レシピ

(Step 1) 整数 $k \geq 2, \sigma_1, \ldots, \sigma_k$ と $B_1, \ldots, B_k$ を決める。**(Step 2)** k 組のスケール変換したブートストラップを $y^*_{i,j} \sim N_K(y, \sigma_i^2 I)$, $i=1,\ldots,k$, $j=1,\ldots,B_i$ より生成する。**(Step 3)** $y^*_{i,j} \in \mathcal{R}$ となった回数を $i=1,\ldots,k$ についてそれぞれ数え、$\tilde{p}_{\sigma_i}=B_i^{-1}\sum_{j=1}^{B_i} I_{\mathcal{R}}(y^*_{i,j})$ とする。**(Step 4)** k 個の点 $(\sigma_i^{-1}, \Phi^{-1}(1-\tilde{p}_{\sigma_i}))$ をプロットし、(6) を当てはめる。

$$\mathrm{RSS}=\sum v_i^{-1}\left(\sigma_i^{-1}x_0+\sigma_i c-\Phi^{-1}(1-\tilde{p}_{\sigma_i})\right)^2$$

を最小にするように x_0 と c を推定する。ただし $v_i=\tilde{p}_{\sigma_i}(1-\tilde{p}_{\sigma_i})/(\phi(\Phi^{-1}(\tilde{p}_{\sigma_i}))^2 B_i)$ である。**(Step 5)** 修正 p-値を $\hat{p}=1-\Phi(x_0-c)$ より計算する。**(Step 6)** 漸近論が break-down していないかの診断をする。漸近論による近似が正当化されるなら RSS は自由度 $k-2$ の χ^2 分布に従うはずなので、もし RSS が有意に大きければ $\hat{p}$ はあまり信用できない。

4 補足事項

近似的に不偏な検定はこれまでにもいろいろ提案されている。とくに凹錘を帰無仮説とする問題については Perlman and Wu (1999) に批判的な良いレビューがある。Shimodaira (2000a) では凸錘が議論されている。これらの方法で現れる棄却域の形状の「ふくらみやへこみ」は、我々の方法によって得られたものと定性的には似ていて興味深い。

我々の方法の詳細と数値例は Shimodaira (2000b) を参照。利用したプログラムを得るには shimo@ism.ac.jp に請求してください。

- EFRON, B., HALLORAN, E. AND HOLMES, S. (1996). Bootstrap confidence levels for phylogenetic trees. *Proc. Natl. Acad. Sci. USA* **93**, 13429–13434.
- EFRON, B. AND TIBSHIRANI, R. (1998). The problem of regions. *Ann. Statist.* **26**, 1687–1718.
- FELSENSTEIN, J. (1985). Confidence limits on phylogenies: an approach using the bootstrap. *Evolution* **39**, 783–791.
- LEHMANN, E. L. (1986). *Testing Statistical Hypotheses.* Wiley, New York, second edition.
- PERLMAN, M. D. AND WU, L. (1999). The emperor's new tests. *Statistical Science* **14**, 355–381.
- SHIMODAIRA, H. (2000a). Approximately unbiased one-sided tests of the maximum of normal means using iterated bootstrap corrections. Technical Report No. 2000-7, Stanford University.
- SHIMODAIRA, H. (2000b). Another calculation of the p-value for the problem of regions using the scaled bootstrap resamplings. Technical Report No. 2000-35, Stanford University.