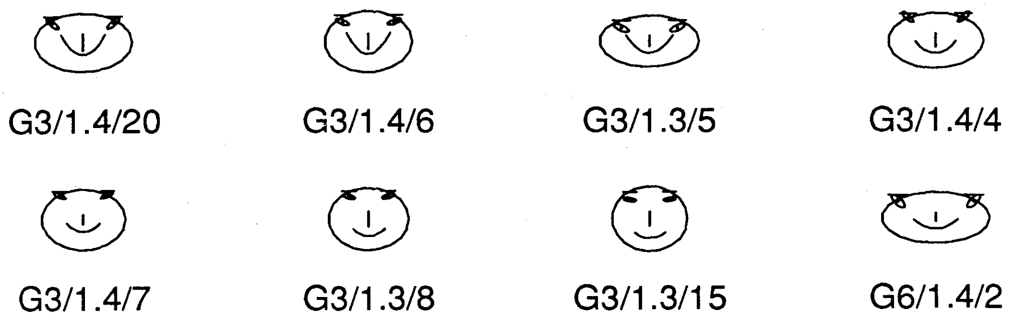

図１．（a）塩分摂取の多い地方のチャーノフの顔．

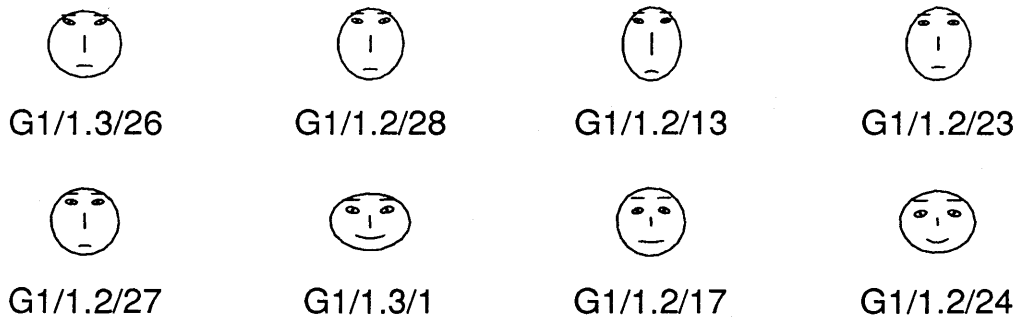

図１．（b）塩分摂取の少ない地方のチャーノフの顔．

と言っても過言ではないと思われる．

年度研究報告会においては，このような統計グラフィックスの現状の一部と使用に際しての注意事項についての報告を行なった．講演で用いた図の一部を表示しておく．厚生省の栄養調査の結果である都道府県別の栄養摂取の充足率の違いをチャーノフの顔を用いて表わそうとした図である．塩分摂取の多い地方（図1(a)）と，比較的少ない地方（図1(b)）でかなりの表情の違いが読みとれると思う．ただし，主観からどれだけ自由になるか，すなわち主観によらない判断ができるかが，チャーノフの顔の重要性を左右する問題となっているが，これに対する回答はないように思われる．しかし，この回答がないことはチャーノフの顔の有効性を否定するものではない．慎重に用いればかなり有効な道具になると言うことができる．

帰無仮説下の周辺尤度の振舞いとパラメータ空間の次元

伊　庭　幸　人

データ数が多いときの周辺尤度（TYPE II 尤度，ABIC）の漸近的な振舞いはどうなるだろうか．この場合，データ数が多くなると同時にパラメータ数（状態ベクトルの要素数）N が多くなり，(データ数)/(パラメータ数) が一定になるような極限を考えるのが自然だろう．このような場合も，通常の i.i.d. の漸近理論の結果が常に成り立つのだろうか．

柳本らは，パラメータ $\{\mu_i\}$ の尤度と“事前分布”がそれぞれ，

(1) $$L_{\sigma^2}(\{\mu_i\}\mid\{y_i\})=\frac{1}{(2\pi\sigma^2)^{N/2}}\exp\left(-\sum_i\frac{(\mu_i-y_i)^2}{2\sigma^2}\right)$$

(2) $$\pi_{\tau^2}(\{\mu_i\})=\frac{1}{Z}\exp\left(-\frac{1}{2\tau^2}\vec{\mu}^t D_2\vec{\mu}\right)$$

で与えられるような1次元の平滑化問題を，仮説検定の枠組で考察した（柳本武美・柳本正勝（1983), Yanagimoto, T. and Yanagimoto, M. (1987)). ただし，D_2 は2次形式 $(x_{i+1}-2x_i+x_{i-1})^2$ に対応する行列を，Z は分布の規格化定数をそれぞれ意味する．具体的な手順としては，周辺尤度

(3) $$U(\sigma^2,\tau^2)=\int\cdots\int\left(\prod_i d\mu_i\right)L_{\sigma^2}(\{\mu_i\}\mid\{y_i\})\pi_{\tau^2}(\{\mu_i\})$$

を σ^2, τ^2 に関して最大化した統計量の分布を，$\{y_i\}$ が直線 $\alpha+\beta i$ のまわりで独立に正規分布するという仮説のもとで求め，それにもとづいて検定を行うわけである．柳本らのシミュレーションの結果は，この分布が通常の i.i.d. の漸近理論で期待される形（χ^2 分布と δ 分布の1:1の混合）にならないことを示唆している．この異常の主な原因はパラメータ数 $N\to\infty$ のとき行列 D_2 の固有値 $\{\lambda_i\}$ が零に集積することにある．この事実はふだんは無害であるが，ここで論じたような仮説検定の状況では $U(\sigma^2,\tau^2)$ の最大値を与える τ^2/σ^2 が正の確率で零になるので，異常な現象を引き起こすことになる．

本報告では，この現象がパラメータ空間の次元 dim を高くしたときにどのように変わるかを論じる（ここでパラメータ空間の次元と呼んでいるのは，1次元なら curve fitting, 2次元なら surface fitting というような意味の次元である). 筆者の予想は，問題の異常は次元が高くなるにつれて弱くなり，ある（かなり高い）次元以上では全く消滅し，通常の i.i.d. の漸近理論の結果が回復されるというものである．これは統計物理での Lower Critical Dimension（それより高次元では赤外発散がなくなる次元）の存在に対応している．

以下では，行列 D_2 の高次元版の代わりに，2次形式 $(x_{i+1}-x_i)^2$ に対応する行列 D_1 の高次元版 D_1^{dim} を使った場合を扱う．D_2 の高次元版の場合については最後に触れる．

柳本らの論文によれば，$\tilde{y}_i$ を独立な正規乱数，$\{\lambda_i\}$ を行列 D_1^{dim} の（非零の）固有値とするとき，

(4) $$\frac{1}{\sqrt{N}}\left\{\sum_i\left(\frac{1}{\lambda_i}\tilde{y}_i^2\right)-\frac{1}{N}\left(\sum_i\frac{1}{\lambda_i}\right)\left(\sum_i\tilde{y}_i^2\right)\right\}$$

の分布が $N\to\infty$ で正規分布になるか否かが問題の本質と考えられる．そのためには，$N\to\infty$ で

(5) $$\frac{\max_i\frac{1}{\lambda_i^2}}{\sum_i\frac{1}{\lambda_i^2}}\to 0$$

となることが必要十分である．

境界条件が簡単な場合には固有値 $\{\lambda_i\}$ を陽に求めることができる．以下では，有限個を除いて固有値が dim 個の整数 $\{m_j\}$ $(0<m_j<L-1)$ で番号付けられ，小さい $\{m_j\}$（小さい固有値に対応する）について，

(6) $$\lambda_{\{m_j\}}\sim\sum_j\left(\frac{m_j}{L}\right)^2$$

となることを利用するだけで十分である．実際にはさらに弱い仮定でよい．L はモデルの“1辺の長さ” (linear dimension) であり，パラメータの数 N とは

(7) $$N=L^{\mathrm{dim}}$$

の関係がある．

まず，(5) の分子を考える．(6) より，Lを大きくしたとき，dim によらず

$$\max_i \frac{1}{\lambda_i^2} \sim (L^2)^2 \tag{8}$$

である．次に分母を考えると，L が大きいとき，和

$$\frac{1}{L^{\dim}} \sum_i \left(\frac{1}{\lambda_i}\right)^2 = \frac{1}{L^{\dim}} \sum_{\{m_i\}} \frac{1}{\lambda_{\{m_i\}}^2} \tag{9}$$

は $k_j = \pi m_j / L$ に関する積分になおして評価することができて，

$$\frac{1}{L^{\dim}} \sum_i \frac{1}{\lambda_i^2} \sim \int_\varepsilon^\pi \cdots \int_\varepsilon^\pi \left(\prod_{j=1}^{j=\dim} dk_j\right) \frac{1}{f(|\vec{k}|)} \sim \int_\varepsilon^\pi \frac{1}{f(|\vec{k}|)} |\vec{k}|^{\dim-1} d|\vec{k}| \tag{10}$$

となる．ただし，$f(|\vec{k}|)$ は $|\vec{k}|$ の小さいところで，

$$f(|\vec{k}|) \sim \frac{1}{|\vec{k}|^4} \tag{11}$$

のように振舞う $|\vec{k}|$ の関数であり，積分の下端の“cut off” ε は，

$$\varepsilon \sim \frac{1}{L} \tag{12}$$

である．積分の $(\dim-1)$ 乗を含む因子は $(\dim-1)$ 次元超球面の面積から出てくる．

(10)，(11)，(12) より，L が大きいときの和の振舞いは，
$\dim < 4$ では，

$$\frac{1}{L^{\dim}} \sum_i \frac{1}{\lambda_i^2} = L^{4-\dim} \tag{13}$$

$\dim > 4$ では，

$$\frac{1}{L^{\dim}} \sum_i \frac{1}{\lambda_i^2} = \text{const.} \tag{14}$$

となる．(8)，(13)，(14) を条件 (5) にあてはめると，$\dim > 4$ では $L \to \infty$ で

$$\frac{\max_i \frac{1}{\lambda_i^2}}{\sum_i \frac{1}{\lambda_i^2}} \sim L^{4-\dim} \to 0 \tag{15}$$

となり，中心極限定理が成り立つ条件が満たされるのに対し，$\dim < 4$ では (5) の分母と分子が同じオーダー (L^4) になり，条件が満たされないことがわかる．

以上のような議論から，$D_1^{\dim}$ に関しては，

$$\dim^c = 4 \tag{16}$$

が通常の漸近理論が成り立つかどうかの境界になるのではないかと推測される．また，行列 $D_1^{\dim}$ の代わりに D_2 の高次元版を考えた場合，高次元への拡張の仕方に任意性があるが，基本的には同様の議論が成り立つと考えられる．この場合の境界は，固有値の振舞いが定性的に異なるため，

$$\dim^c = 8 \tag{17}$$

となる．

なお，あとで知ったことであるが，本報告に関連した事項については，Wahba(1990)の第6章でも論じられているようである．

親身に議論をして下さった柳本武美教授に感謝します．また，土谷 隆氏と鈴木淳史氏には貴重なコメ

ントを頂きました.

参考文献

Wahba, G. (1990). *Spline Models for Observational Data*, SIAM, Philadelphia, Pennsylvania.

柳本武美，柳本正勝 (1983). 系列データに対する単回帰モデルの適合性を診断する手法とその適用，統計数理研究所彙報，**31**，117-127.

Yanagimoto, T. and Yanagimoto, M. (1987). The use of marginal likelihood for a diagnostic test for the goodness of fit of the simple linear regression model, *Technometrics*, **29**, 95-101.

統計基礎研究系

多変量尺度混合分布の漸近展開

清 水 良 一

分布 G に従う p 次元確率ベクトル Z の尺度混合 $X=\Sigma^{1/2}Z$ の分布関数を $F(x)$ とする．ただし，Σ は Z と独立で，単位行列 I_p の近傍で変動する確率行列であるとする．問題は，F を G の周りで展開すること，そして展開を有限の項で打ち切ったときの誤差を評価することである．昨年度に報告したものの改良について述べた.

F, G の確率密度関数をそれぞれ f, g とし，f を

$$f_k(x)=\left\{1+\sum_{j=1}^{k-1}(\cdots)\right\}g(x)$$

という形の関数で近似する：$f(x)=g_k(x)+\delta_k(x)$.

問題は … の部分の決定と誤差項 $\delta_k(x)$ の評価である．具体的には

$$\Delta_k\equiv\int_{R^p}|\delta_k(x)|\,dx$$

を評価したい．昨年度は $\Sigma=\mathrm{diag}(\sigma_1^2, \sigma_2^2, \ldots, \sigma_p^2)$ という特殊な場合について述べたが，これを一般の Σ に拡張する．$k=2$ のとき，

$$g_2(x)=\{1+2^{-1}(x^tE(\Sigma-I)x-(E\mathrm{tr}(\Sigma-I)))\}\phi(x)$$

という展開が得られるが，その誤差評価が次の様に与えられる.

確率1で $\Sigma-I>0$ であれば

$$(*)\qquad \Delta_2\leqq 0.26\cdot(E|\Sigma|)^{1/2}\cdot(E(\mathrm{tr}(\Sigma-I))^4)^{1/2}.$$

確率1で $I-\Sigma>0$ のときには（$*$）において Σ を Σ^{-1} に置き換えればよい．一般の場合には

$$\Delta_2\leqq 2.04\cdot p^3\cdot\{E(\lambda_+^p+\lambda_-^p)\}^{1/2}\cdot\{E(\mathrm{tr}(\Sigma-I)^4+\mathrm{tr}(\Sigma^{-1}-I)^4)\}^{1/2}$$

である．ただし，λ_+ と λ_- はそれぞれ Σ および Σ^{-1} の最大固有値を表す.

統計モデルとしての確率分布とその応用

平 野 勝 臣

本年度の研究：(1) 正値連続分布の典型であるスケール分布族の代表的な分布についての考察をまとめた（岩瀬・平野 (1990)）．(2) オーダー k の離散分布の研究のうち，K.D. Ling 氏の訪問を機会に彼の研究について調べ，これに関するいくつかの結果をまとめた（Hirano et al. (1991)）．当日は以上の共