傾向性仮説と変化点モデルの様々な応用

広津千尋（明星大学連携研究センター）

1．序論

用量反応解析においては，通常，厳密な反応曲線を想定することが難しいために単調性，凸性，Ｓ字性のような形状制約がよく設定される．それらのうち正規分布モデルに対する単調仮説については，isotonic regression がよく知られている．しかしながら，それは Bartholomew によってやや直観的に導入され，とくにこのような制約のある母数空間に対する最適性は自明ではない．さらに制約付き最小二乗法は，計算及び分布論が複雑で，凸性やＳ字性問題，正規分布以外の確率モデル，あるいは２元配置交互作用問題への拡張には困難が伴う．一方，著者等のアプローチは Hirotsu (1982)で導かれた一般制約仮説に対する検定の完全類を基にしており，その意味での最適性を持っている．それによると単調性，凸性，Ｓ字性仮説それぞれに対し，単純, 2 重, 3 重累積和に基づく単調増大な統計量が示唆される．本稿ではそのうち規準化最大対比，および規準化二乗和を用いる方法について論ずる．

一方，これらの形状制約は変化点モデルと密接な関係がある．すなわち，単調性，凸性，Ｓ字性仮説はそれぞれ，段差変化点，スロープ変化点，変曲点モデルと対応する(Hirotsu and Marumo, 2002)．例えば，段差変化点モデルを表す対比は単調対比の典型であり，逆にすべての単調対比は段差変化点対比の一意正係数線形結合で表される．例えばＰＭＤＡでは副作用自発報告が収集され，その経年変化が解析されている．その場合，単調増加傾向をいち早く検出すると同時に，増加傾向の生じた時点を推測することは応用上も大変有意義である．同様のことが凸性，Ｓ字性問題についても示される．最大対比統計量は，これら変化点モデルに対する efficient score 検定を与える．すなわち，従来統計学の二つの異なる流れの中で研究されてきた制約仮説と変化点問題を統合的に扱う事が出来る．

以上のように，累積和に基づく方法は様々な確率モデル，様々な実質科学上の問題を理論的に一貫した方法で扱い，効率の良い計算アルゴリズムを与えることが出来る．正規分布，2 項分布，Poisson 分布についてはこれまでいろいろな機会に発表しているので，本論では最近の研究である独立な 2×2表系列への応用について述べる．重要な例としてケース・コントロール研究がある．

2．問題の概要

二つの 2 項分布母集団をK層上で比較する問題を考える．オッズ比の一様性が仮定出来る場合は例えば Gart (1970)により共通オッズ比に関する推論を行えば良い．一様性検定については最尤法や，Breslow & Day 検定が良く知られている他，Zelen (1971)の正確検定も利用出来る．正確推論に関しては Agresti (1992)のサーベイが参考になる．本論はとくにK層

に自然な順序がある場合に，単調性，および凸性を対立仮説とする検定について論ずる．これらは上記の論文では論じられていないが，単調性に対する制約付き最小二乗法が El Barmi によって導かれている．単調性に対する最大対比検定は Hirotsu 他(2001)の特殊ケースに当たり，その特殊ケースはまた太田他(2003)で詳しく論じられている．本論ではまずこの検定について，累積和統計量の規準化に用いられている漸近分散を正確分散に置き換える改良を行い，Fleiss 他(2003)のロジット線形モデルによる交互作用解析と比較する．その結果は良く整合する．次にオッズ比に関する凸性仮説検定の定式化を行い，新たに最大対比正確検定を提案する．これら二つの問題について累積χ^2乗検定も与える．

3. 定式化

第K層のデータをy_{ijk}，その確率を$p_{ijk}(p_{i\cdot k}=1)$で表す，ただし，$i=1,2$ は二つの 2 項分布母集団，$j=1,2$は成否の反応，そして$k=1,\cdots,K$は層を表す．この時，y_{ijk}を並べたベクトルを$\boldsymbol{y}$として，確率分布は$g(\boldsymbol{y})=\prod_{k=1}^{K}\prod_{i=1}^{2}\left\{y_{i\cdot k}!\prod_{j=1}^{2}\left(p_{ijk}^{y_{ijk}}/y_{ijk}!\right)\right\}$で与えられる．ここで 対数線形模型：$\log p_{ijk}=\gamma_k+\varphi_{ik}+\tau_{jk}+\omega_{ijk}$を想定した上，周辺和$y_{i\cdot k}$および$y_{\cdot jk}$を与えた相似検定を考えると，条件付き確率分布は$g(\boldsymbol{y},\boldsymbol{\omega})=\prod_{k=1}^{K}C^{-1}(\omega_k)b(y_{11k})\exp(\omega_k y_{11k})$という簡単な形になる．ただし，$\boldsymbol{y}$は改めて$y_{11k}$を並べたベクトルとし，$\omega_k$が推論対象である対数オッズ比である．この先，$\omega_k$に関する単調性，および凸性検定は 2 項分布や，Poisson 分布等の 1 母数指数分布族に対する方法(Hirotsu 他，2016; Hirotsu and Tsuruta, 2017; Hirotsu，2017)にならい，カーネルおよび漸化式遂行の不等式の変更を適切に行うことにより実行出来る．なお，Hirotsu (2017)では単調対比の同時信頼区間やパターン推測，さらに交互作用のパターン推測問題も扱われている．

4. 単調仮説問題のポイント

単調仮説問題を扱う際にキーとなるのは，累積和統計量$Y_k=\sum_{i=1}^{k}y_{11i}, k=1,\ldots,K-1$である．その規準化最大統計量として$\max \mathrm{acc}.t1$，規準化二乗和として累積$\chi^2(\chi^{*2})$が定義される．前者には効率の良い正確法，後者には良好なχ二乗近似を提案できる．最大統計量の確率計算は一般的には多重積分や多重和分の計算となり，Kがある程度大きいと直ぐ困難になる．しかしこの場合はY_kのマルコフ性から，極めて効率の良い確率計算のアルゴリズムが得れる．その方式はχ^{*2}のχ^2近似のためのモーメント計算にも利用される．キーとなるのはY_kの同時分布のマルコフ性を用いた分解：$\prod_{k=1}^{K-1}f_k\left(Y_k \mid Y_{k+1}\right)$である．一様性の帰無仮説の下でこの分解は，正規，Poisson，2 項分布についてはよく知られており，それぞれ正規，2 項，超幾何分布の積になる．今回の$2\times2\times K$表の場合に一般的に知られた結果は無いが，漸化式を用いて効率よく生成できる．それは漸化式を用いた同時分

布関数やモーメント計算に応用される．なお，対立仮説の下では一様性の下と異なり一般的な結果は無いが，数値的に計算する方法に特別な問題が生じることは無く，漸化式によるp値や検出力の計算は all in one program に纏められる．

5．凸性仮説問題のポイント

凸性仮説問題を扱う際にキーとなるのは，2 重累積和統計量$S_k = \sum_{i=1}^{k} Y_i = ky_{111} + \cdots + y_{11k}, k = 1, \ldots, K-2$である．その規準化最大統計量として$\max \mathrm{acc}.t2$，規準化二乗和として累積$\chi^2(\chi^{\dagger 2})$が定義される．この場合も前者に効率の良い正確法，後者には良好なχ^2近似を提案できる．この場合もS_kの 2 階マルコフ性から，同時分布の分解：$\prod_{k=1}^{K-2} g_k\,(S_k | S_{k+1}, S_{k+2})$が得られるが，成分$g_k$について一般的に知られた結果は無い．そこで，漸化式によって数値的に生成するが，基本的なアイディアは，対立仮説の場合も含め，$f_k(Y_k \mid Y_{k+1})$の生成と同じである．

以上に付き具体的な計算式，および応用例は当日与える．

参考文献

Agresti, A. (1992). A survey of exact inference for contingency tables. *Statistical Science* 7, 131-177.

El Barmi, H. (1997). Testing for or against a trend in odds ratios. *Communications in Statistics - Theory and Methods,* 26, 1877–1891.

Fleiss, J. L., Levin, B. and Paik, M. C. (2003). *Statistical methods for rates and proportions.* Wiley Series in Probability and Statistics, New York.

Gart, J. J. (1970). Point and interval estimation of the common odds ratio in the combination of 2 x 2 tables with fixed marginals. *Biometrika* 57, 471-475.

Hirotsu C. (1982). Use of cumulative efficient scores for testing ordered alternatives in discrete models. *Biometrika* 69, 567-577.

Hirotsu, C. (2017). *Advanced analysis of variance*. Wiley Series in Probability and Statistics, New York.

Hirotsu, C., Aoki, S., Inada, T. & 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. and Marumo, K. (2002). Change point analysis for isotonic inference. Scand. J. Statit. 29, 125-138.

Hirotsu, C., Yamamoto, S. and Tsuruta, H. (2016). A unifying approach to the shape and change-point hypotheses in the discrete univariate exponential family. *Computational Statistics and Data Analysis*, 97, 33-46.

Hirotsu, C, and Tsuruta, H. (2017). An algorithm for a new method of change-point analysis in the independent Poisson sequence. *Biometrical Letters* 54, 1-24.

太田絵里，青木敏，広津千尋(2003). $2 \times 2 \times K$分割表における単調仮説の検定．応用統計学 32, 107－126.

Zelen, M. (1971). The analysis of several 2 x 2 contingency tables. *Biometrika* 58, 129-137.