

1. 序論

通常変化点仮説というとは，分布平均がある時点で階段的に変化し，それを維持する複合仮説を指す．そしてその各成分が，丁度単調仮説を定義する凸錘のコーナーベクトルに対応することから，単調仮説検定の完全類に属する最大対比型検定 **max acc. t** が変化点仮説に対しても **efficient score test** となることが導かれる．興味ある解釈として，検定の完全類は基本的に，観測値ベクトルをコーナーベクトルに射影した各成分について単調増大となる統計量で構成される (Hirotsu, 1982)．さらに，最大対比型検定統計量の引続く成分が 1 階の **Markov Process** に従うことから，指数分布族に対し正確かつ高速な確率計算アルゴリズムが導かれる．単調仮説検定の方法としては，**isotonic regression** (制約付尤度比検定) がよく知られているが，このようにパラメータ空間が制限される場合にはその最適性は自明ではない．

一方，これに対応して正規分布の場合に，スロープ変化点仮説と凸性仮説の間に類似の関係が成立し，最大対比型検定統計量成分の 2 階 **Markov** 性から正確高速な確率計算アルゴリズムが得られることが示されている (Hirotsu & Marumo, 2002)．正規分布の場合，これらの性質は分散行列の形のみから示されるが，本論は **Poisson Process** に関し同様の結果を導こうとするものである．

検定の完全類に属する統計量としては，基準化最大成分を用いる **max acc. t** の他，二乗和を用いる累積 χ 二乗があるが，本論では割愛する．それについては，広津 (1992, 2004) 等を参照されたい．

2. モデル

データ $\mathbf{y} = (y_1, \dots, y_n)$ の各成分が互いに独立に **Poisson** 分布 $\text{Po}(\lambda_i)$ に従うものとする．このとき，スロープ変化モデル **S** は次式で定義される，

$$\mathbf{S}: \begin{cases} \mu_i = \log \lambda_i = \beta_0 + \beta_1 \cdot i, & i = 1, \dots, k, \\ \mu_i = \log \lambda_i = \beta_0^* + \beta_1^* \cdot i, & i = k+1, \dots, n, \quad (\beta_0 + \beta_1 \cdot k = \beta_0^* + \beta_1^* \cdot k). \end{cases}$$

次に，2 階差分行列 L'_n により，凸性仮説 **C** を定義する，

$$\mathbf{C}: L'_n(\mu_1, \dots, \mu_n) \geq 0 \quad (\text{少なくとも 1 個の不等号厳密}).$$

このとき，凸錘 **C** のコーナーベクトル $L_n(L'_n L_n)^{-1}$ がスロープ変化モデルを形成し， \mathbf{y} をコーナーベクトルへ射影した $\mathbf{z} = (L'_n L_n)^{-1} L'_n \mathbf{y}$ の各成分について単調増大な検定統計量を構成すればよいことが分かる (Hirotsu, 1982)．幾つかの統計量が考えられる中，規準化最大成分を用いる方法がスロープ変化点仮説に対する **efficient score test** であると同時に，凸性仮説を検定する適切な統計量となる．以下では \mathbf{z} の引続く成分が 2 階 **Markov** 性を持ち，それに基づいて規準化最大成分に関し，漸化式による確率計算アルゴリズムが得られることを示す．

3. 確率分布の分解とそれを用いた漸化式

モデル **S**, **C** それぞれの帰無仮説の下での完備十分統計量は、 $S_i = iy_1 + (i-1)y_2 + \dots + y_i$, $i = n-1, n$ であり、以下ではこれらを与えた条件付き推測を考える。その場合、 \mathbf{z} の各成分は $S_i, i = 1, \dots, n-2$, の線形関数（+定数）として表され、2階 **Markov** 系列となることが示される。それにより、 $S_i, i = 1, \dots, n-2$, の同時分布は、 $n-2$ 個の条件付き確率の積に分解される。同時分布に基づく確率計算は n が大きくなるとすぐ実行不能になるが、条件付き確率の積を利用した漸化式による計算は多重和を単一和の繰り返しに変換するため、計算時間が大幅に改善される。さらに、帰無仮説の下で S_i を基準化するための平均と分散も漸化式によって計算することが出来る。そこで、基準化した統計量を S_i^* と表し、検定統計量を $W = \max S_i^*$ とする。このとき、 S_i^* の同時分布もまた条件付き確率 $f(S_{k-1}^* | S_k^*, S_{k+1}^*)$ の積に分解され、 W の分布関数の計算に関し、次の漸化式が得られる。

$$F_{k+1}(S_k^*, S_{k+1}^*) = \sum f(S_{k-1}^* | S_k^*, S_{k+1}^*) F_k(S_{k-1}^*, S_k^*)$$

ただし、 $F_k(S_{k-1}^*, S_k^*) = \Pr(S_1^* \leq s_0, \dots, S_k^* \leq s_0)$ は、 S_{k-1}^*, S_k^* を与えた条件付確率である。最終的に、分布関数 $\Pr(W \leq s_0) = 1 - F_n(S_{n-1}^*, S_n^*)$ が得られる。ここで、条件付き分布 $f(S_{k-1}^* | S_k^*, S_{k+1}^*)$ 自体も漸化式で計算されるが、その導出は本質的に、3元分割表で無3因子交互作用モデルの分布を漸化式で求めた **Hirotsu et al. (2001)** の方法と同種で、それを2階に拡張したものとなる。

4. 応用

医薬品機構には医薬品－副作用コンビネーションの自発報告が日々集積されている。その時系列を追い、増加傾向の変化を検出するには単調仮説検定に対する **max acc. t** 法が有用である。それにより変化を検出し、必要なら警告措置が取られるが、その後、増加傾向が減少傾向に転じたことを確認するのに凹性検定が有用と考えられる。その他、凹性は用量反応関係を始め、一般的な入出力関係でもよく仮定される。なお、棄却域の反転で、変化点、およびその前後の変化量の信頼区間が得られるが、ある程度タフな数値計算になることが予想される。

5. 参考文献

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

Hirotsu, C. & Marumo, K. (2002). Changepoint analysis as a method for isotonic inference. *Scandinavian Journal of Statistics* 29, 125-138.

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.

広津千尋(1992). 実験データの解析—分散分析を超えて—. 共立出版, 東京.

広津千尋(2004). 医学・薬学データの統計解析—データの整理から交互作用多重比較まで—東京大学出版会, 東京.

