A unifying approach to the shape and changepoint hypotheses

Chihiro, Hirotsu
Collaborative Research Center, Meisei University
Hino-city, 191-8506 Tokyo, Japan
E-mail: hirotsu@ge.meisei-u.ac.jp

Summary Usually a changepoint model assumes a step-type change at some point of a time series. Then as an interesting thing it has been shown that the max acc. t test for the isotonic hypothesis is also appropriate for detecting the changepoint. It comes from a relationship that each corner vector of the polyhedral cone defined by the isotonic hypothesis corresponds to a component of the changepoint model. The max acc. t is essentially the maximal component of the projections of the observation vector on to the corner vectors of the polyhedral cone and a very efficient algorithm for the probability calculation is obtained based on the Markov property of the component statistics. The idea has been extended to the concavity hypothesis and the slope change model for the normal model in Hirotsu and Marumo (Scand. J. Statist. 2002). In this paper we extend the idea further to a Poisson model.

Keywords: Doubly accumulated statistic, Exponential family, Maximal contrast type test, Recursion formula, Second order Markov property.

1. Introduction

A general relationship between the shape restrictions and the changepoint models has been discussed in Hirotsu and Marumo (2002). In particular the monotone increase corresponds to the simple step-type changepoint model whose components are the special case of monotony. For this monotone case the maximal contrast type statistic, max acc. t, has been constructed based on the complete class lemma in Hirotsu (1982) and proved in several occasions to be well behaved as compared with an isotonic regression approach, see Hirotsu (1993, 1998) and Hirotsu *et al.* (1992, 2010), for example. It can be shown that the test is also an efficient score test for the changepoint hypothesis. To be exact define the monotone hypothesis by

(1) $$\mathrm{C} : D_K'\boldsymbol{\mu} \geqq 0$$

with a differential matrix

$$D_K' = \begin{bmatrix} -1 & 1 & 0 & 0 & 0 & \dots & 0 & 0 & 0 \\ 0 & -1 & 1 & 0 & 0 & \dots & 0 & 0 & 0 \\ & & & & \dots\dots\dots\dots & & & & \\ 0 & 0 & 0 & 0 & 0 & \dots & 0 & -1 & 1 \end{bmatrix}_{(K-1)\times K},$$

where = $(\mu_1, \dots, \mu_K)'$ is a vector of means and A' denotes the transpose of a matrix A. Then the corner vectors of cone C are given by $D_K(D_K'D_K)^{-1}$ or in other words each $\boldsymbol{\mu}$ satisfying (1) is expressed by a positive linear combination of the columns of $D_K(D_K'D_K)^{-1}$, that is, $\boldsymbol{\mu} = D_K(D_K'D_K)^{-1}\mathbf{c}$ with $\mathbf{c} \geqq 0$. On the other hand the matrix is given explicitly by

(2) $$D_K(D_K'D_K)^{-1} = \frac{1}{K}\begin{bmatrix} -(K-1) & -(K-2) & & -1 \\ 1 & -(K-2) & & -1 \\ 1 & 2 & \cdots & -1 \\ \vdots & \vdots & & \vdots \\ 1 & 2 & & (K-1) \end{bmatrix},$$

and it is seen that each column of the right-hand side of (2) represents a component of the step-type changepoint model

(3) $$\mathrm{C}_k : \mu_1 = \cdots = \mu_k \le \mu_{k+1} = \cdots = \mu_K,\ k = 1, \cdots, K-1.$$

This clarifies the relationship between the monotone hypothesis and the step-type changepoint model. Similarly we define the concavity hypothesis by

(4) $$\mathrm{C} : L_K'\boldsymbol{\mu} \geqq 0$$

with the second order differential matrix L_K. Then the corner vectors of cone C are given by $L_K(L_K'L_K)^{-1}$ and it can be shown that each column of $L_K(L_K'L_K)^{-1}$ represents a component of the slope changepoint model

(5) $$\mathrm{M}_k : \begin{cases} \mu_2 - \mu_1 = \mu_3 - \mu_2 = \cdots = \mu_{k+1} - \mu_k = \beta_k, \\ \mu_{k+2} - \mu_{k+1} = \mu_{k+3} - \mu_{k+2} = \cdots = \mu_K - \mu_{K-1} = \beta_k^*, \end{cases} \quad k = 1, \cdots, K-2.$$

The null hypothesis $\mathrm{C}_0 : L_K'\boldsymbol{\mu} = 0$ corresponds to

$$\mathrm{M}_0 : \mu_i = \beta_0 + \beta_1 i,\ i = 1, \cdots, K.$$

For $\boldsymbol{y} \in N(\boldsymbol{\mu}, \sigma^2 I)$ let z be $D(H_K'H_K)^{-1}H_K'\boldsymbol{y}$ with D a diagonal matrix for standardization and H_K stands for D_K and L_K for each of the monotone and concavity hypotheses, respectively. Then the maximal element z_m of z has been shown to be an appropriate statistic for testing the shape hypothesis (1)(4) and/or the corresponding changepoint model (3)(5), see Hirotsu (1982) and Hirotsu and Marumo (2002). Further a very efficient algorithm for probability calculation of z_m for any K has been proposed based on the Markov property of the serial elements of z. Those properties have been derived only by the covariance structure in case of the normal model. In this paper we extend those ideas to an independent Poisson sequence.

2. Poisson model

Let y_i be distributed as an independent Poisson distribution $P_o(\lambda_i)$ with mean $\lambda_i, i = 1, \ldots, K$. We assume a log linear model $\mu_i = \log \lambda_i$ and define the shape hypothesis C (1) or (4) in $\boldsymbol{\mu}$ and a set of changepoint models by the equations (3) or (5). Then according to a complete class lemma in Hirotsu (1982) an appropriate test should be increasing in every element of $(H_K'H_K)^{-1}H_K'\hat{\boldsymbol{v}}_0$, where

$$\hat{\boldsymbol{v}}_0 = (y_1 - e^{\hat{\mu}_1}, \cdots, y_K - e^{\hat{\mu}_K})'$$

is the efficient score vector with respect to $\boldsymbol{\mu}$ evaluated at the null model with $\hat{\mu}_i$ the maximum likelihood estimator and function of the complete sufficient statistics, which are $Y_K = y_1 + y_2 + \cdots y_K$ for the monotone hypothesis and $Y_K = y_1 + y_2 + \cdots + y_K$ and $T_K = y_1 + 2y_2 + \cdots + Ky_K$ for the concavity hypothesis. In particular in this paper we propose the maximal standardized element of $(H_K'H_K)^{-1}H_K'\hat{\boldsymbol{v}}_0$ and denote it by z_m. It can be shown that the k th element of $(H_K'H_K)^{-1}H_K'\hat{\boldsymbol{v}}_0$ is essentially the efficient score for testing the changepoint hypothesis C_k or M_k.

3. Markov property of the serial elements of $(H_K'H_K)^{-1}H_K'\hat{v}_0$ and the recursion formula for calculating the *p*-value

By considering the conditional probability given the complete sufficient statistics under the null model we may take the essential part of $(H_K'H_K)^{-1}H_K'\hat{v}_0$ as the basic variables. They are $Y_k = y_1 + y_2 + \cdots + y_k$ for the monotone hypothesis and $S_k = ky_1 + (k-1)y_2 + \cdots + y_k$ for the concavity hypothesis. For the monotone case it can be shown that the sequence $Y_1, \cdots, Y_{K-1}$is a Markov process and the conditional null distribution is factorized into

$$G(\boldsymbol{Y}) = \prod_{k=2}^{K} g(Y_{k-1}|Y_k),\ g(Y_{k-1}|Y_k) = \prod_{k=2}^{K} \binom{Y_k}{Y_{k-1}} (\frac{k-1}{k})^{Y_{k-1}} (\frac{1}{k})^{Y_k - Y_{k-1}}.$$

We consider the maximal contrast test, the max acc. t, which is defined by

$$\max_{k=1,\cdots,K-1} t_k^*,\ t_k^*(Y_k) = -\{\frac{k(K-k)Y_K}{K^2}\}^{-1/2}(Y_k - \frac{k}{K}Y_K).$$

It should be noted that this is an efficient score test for the step-type changepoint hypothesis for which Hawkins (1977) derived a likelihood ratio test in case of the normal model.

For the calculation of the p-value we define the conditional probability

$$F(Y_k) = \Pr\{t_1^*(Y_1) < t_0, \cdots, t_k^*(Y_k) < t_0 | Y_k\}.$$

Then by the Markov property of Y_k we have a recursion formula

$$F(Y_k) = \begin{cases} \sum_{Y_{k-1}} F(Y_{k-1}) f(Y_{k-1}|Y_k),\ t_k^*(Y_k) < t_0 \\ 0, \text{otherwise.} \end{cases}$$

and the p-value is obtained at the final step as $p = 1 - F(Y_K)$.

For the concavity hypothesis the maximal contrast test, the max acc. $t2$, is defined by the maximal standardized statistic of $(-S_k), k = 1, \cdots, K-2$, and a recursion formula for the p-value is also obtained based on the second order Markov property of S_k.

4. Application

An interesting application will be in the monitoring of the spontaneous reporting of the drug-adverse events combination. The max acc. t is useful for detecting the increasing tendency and/or the changepoint. However, after detecting a change and taking an appropriate action it should be interesting to verify that the increasing tendency is changed to decreasing and for the purpose the concavity test should be useful. A real example will be given at the presentation. Another interesting application will be the non-parametric input-output analysis.

5. Discussion

The shape and the changepoint hypotheses have been discussed in two different streams. However, in Hirotsu and Marumo (2002) it is revealed that there is a close relationship between those two research subjects. Then in the normal distribution case the same maximal contrast type statistic is shown to be useful for both hypotheses. In this paper we extended the idea to the Poisson distribution based on the accumulated statistics rather than the contrast statistics.

For the shape hypothesis the most popular approach will be the maximum likelihood ratio test known as an isotonic regression for the monotone case. However, the approach has no obvious optimality for the restricted parameter space like this and actually the maximal contrast test based on the complete class lemma has been proved to behave better on several occasions.

References

Hawkins, D. M. (1977). Testing a sequence of observations for a shift in location. *J. Am. Statist. Assoc.*, 72, 180-186.

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

Hirotsu, C., Kuriki, S. and Hayter, T. (1992) Multiple comparison procedures based on the maximal component of the cumulative chi-squared statistic. *Biometrika*, 79, 381-392.

Hirotsu, C. (1993) Beyond analysis of variance techniques : Some applications in clinical trials. *Int. Statist. Rev.*, 61, 183-201.

Hirotsu, C. (1998) Isotonic inference. In *The Encyclopedia of Biostatistics* (eds P. Armitage and T. Colton), pp. 2107-2115. New York: John Wiley & Sons, .

Hirotsu, C. and Marumo, K. (2002) Changepoint analysis as a method for isotonic inference. *Scandinavian J. Statist.*, 29, 125-138.

Hirotsu, C., Ohta, E., Hirose, N. and Shimizu, K. (2003) Profile analysis of 24-hours measurements of blood pressure. *Biometrics*, 59, 907-915.

Hirotsu, C., Yamamoto, S. and Hothorn, L. (2010) Estimating the dose-response pattern by the maximal contrast type test approach. *Statistics in Biopharmaceutical Research*, 3, 40-53.