A unifying approach to the shape and change-point hypotheses in the discrete univariate exponential family

Chihiro Hirotsu\(^{a,\ast}\), Shoichi Yamamoto\(^{b}\), Harukazu Tsuruta\(^{c}\)

\(^{a}\) Collaborative Research Center, Meisei University, Tokyo, Japan
\(^{b}\) RPM Co. Ltd., Tokyo, Japan
\(^{c}\) Medical Informatics, AHS, Kitasato University, Sagamihara, Kanagawa, Japan

**A R T I C L E I N F O**

Article history:
Received 16 June 2015
Received in revised form 10 October 2015
Accepted 25 November 2015
Available online 30 November 2015

**Keywords:**
Cumulative sum statistic
Goodness-of-fit test
Markov property
Maximal contrast test
Recursion formula
Restricted alternative

**A B S T R A C T**

A unifying approach to the shape and change-point hypotheses is extended generally to a discrete univariate exponential family. The maximal contrast type tests are newly proposed for the convexity and sigmoidicity hypotheses based on the complete class lemma of tests for the restricted alternatives. Those tests are also efficient score tests for the slope change-point and inflection point models, respectively. For each of those tests the successive component statistics are the doubly- and triply-accumulated statistics. They have nice Markov properties for the exact and efficient recursion formulae for calculating the \(p\)-value. Further the sum of squares of the component statistics are developed as the cumulative chi-squared statistics for the directional goodness-of-fit tests of the dose–response model. Therefore the interesting applications will be in monitoring of spontaneous reporting of the adverse drug events and the directional goodness-of-fit tests.

© 2015 Elsevier B.V. All rights reserved.

1. Introduction

A general relationship between the shape restrictions and change-point models in the normal means has been discussed in Hirotsu and Marumo (2002). In particular the monotone hypothesis corresponds to the simple step change-point model whose components are the special case of monotony. For this monotone case the maximal standardized cumulative sum statistic, \(\text{max acc. } \text{t}\), has been constructed based on the complete class lemma in Hirotsu (1982), which has been proved to be well behaved compared with an isotonic regression approach by Bartholomew (1959a,b), see Hirotsu et al. (1992, 2011). In fact the restricted maximum likelihood approach has no obvious optimality for a restricted parameter space like this. It is also too complicated to extend to non-normal distributions, to interaction problems and to other shape constraints such as convex and sigmoid. Actually in the book by Miller (1998) a choice of maximin contrast test by Abelson and Tukey (1963) is recommended for the isotonic inference to escape from the complicated calculations of the isotonic regression. However, such one degree of freedom contrast test cannot keep high power against the wide range of the monotone hypothesis even by a careful choice of the contrast. Instead we propose a more robust approach against the wide range of the restricted alternatives, which can be extended in a systematic way to convex and sigmoid hypotheses.

It has been shown that the max acc. \(\text{t}\) test is also an efficient score test for the step change-point hypothesis (Hirotsu, 1997). Then the convex and sigmoid restrictions are closely related to the slope change-point and inflection point models, for which the doubly- and triply-accumulated statistics are newly developed in this paper. Since Page (1954, 1961) cumulative
sum statistics based approach has been widely developed in the statistical process control, see also Montgomery (2009) for various applications. More recently it is also extended to the field of environmental statistics as in Manly and Mackenzie (2003), for example. However, as stated in the review paper by Amiri and Allahyari (2012) most papers are assuming the step, linear trend or monotonic change, and it seems that the slope change-point and inflection point models are not popular in those fields. One should refer also to Hawkins (1977) and Worsley (1986) for the likelihood ratio test approach to the change-point hypothesis as the previous work to the present paper. Thus the present paper gives a unified approach to the shape and change-point hypotheses which have been investigated in two different streams of statistics. The unifying approach is useful also in practice since in the applications it is usually of importance not only to detect a general tendency but also to detect a particular change-point as seen in the examples of Section 11.

More precisely, each corner vector of the convex polyhedral cone defined by the convexity hypothesis corresponds to a component of the slope change-point model. The convexity hypothesis

\[
C : \mathbf{L}_K^\alpha \mathbf{\mu} \geq 0, \quad \text{with at least one inequality strong,}
\]

is defined by the second order differential matrix

\[
\mathbf{L}_K^\alpha = \begin{bmatrix}
\frac{1}{x_2-x_1} & 0 & \cdots & 0 \\
\frac{1}{x_2-x_3} & \frac{1}{x_2-x_4} & \cdots & 0 \\
\vdots & \vdots & \ddots & \vdots \\
0 & 0 & \cdots & \frac{1}{x_{K-1}-x_K}
\end{bmatrix}
\]

where \(x_i\) denotes the time or location of the \(i\)th event, \(\mathbf{\mu} = (\mu_1, \ldots, \mu_K)'\) a vector of unknown parameters and \(\mathbf{A}\) the transpose of a matrix \(\mathbf{A}\). It forms a convex polyhedral cone with the corner vectors \(\mathbf{L}_K^\alpha (\mathbf{L}_K^\alpha \mathbf{L}_K^\alpha)^{-1}\) or in other words each \(\mathbf{\mu}\) satisfying (1) is expressed by a unique and positive linear combination of the columns of \(\mathbf{L}_K^\alpha (\mathbf{L}_K^\alpha \mathbf{L}_K^\alpha)^{-1}\), that is, \(\mathbf{\mu} = \mathbf{L}_K^\alpha (\mathbf{L}_K^\alpha \mathbf{L}_K^\alpha)^{-1} \mathbf{c}\) with \(\mathbf{c} \geq 0\). Thus the matrix \(\mathbf{L}_K^\alpha (\mathbf{L}_K^\alpha \mathbf{L}_K^\alpha)^{-1}\) is more directly concerned with the convexity hypothesis than the defining matrix \(\mathbf{L}_K^\alpha\).

On the other hand as shown in Section 2 each column of \(\mathbf{L}_K^\alpha (\mathbf{L}_K^\alpha \mathbf{L}_K^\alpha)^{-1}\) represents a component of the slope change-point model \(M_k\) at \(x = x_{k-1}\),

\[
M_k : \begin{cases}
\mu_2 - \mu_1 = \frac{\mu_3 - \mu_2}{x_3 - x_2} = \cdots = \frac{\mu_{k+1} - \mu_k}{x_{k+1} - x_k} = \beta_k, \\
\mu_{k+2} - \mu_{k+1} = \frac{\mu_{k+3} - \mu_{k+2}}{x_{k+3} - x_{k+2}} = \cdots = \frac{\mu_K - \mu_{K-1}}{x_K - x_{K-1}} = \beta_k^n, \\
\end{cases} \quad k = 1, \ldots, K - 2.
\]

The null hypothesis \(H_0 : \mathbf{L}_K^\alpha \mathbf{\mu} = 0\) corresponds to \(M_0 : \beta_k = \beta_k^n \equiv \beta_1\) or equivalently to a linear regression model

\[
M_0 : \mu_i = \beta_0 + \beta_1 x_i, \quad i = 1, \ldots, K.
\]

For the normal model, \(\mathbf{y} \in N(\mathbf{\mu}, \sigma^2 \mathbf{I})\), let \(\mathbf{z}\) be \((\mathbf{L}_K^\alpha \mathbf{L}_K^\alpha)^{-1} \mathbf{L}_K^\alpha \mathbf{y}\). Then the maximal standardized element \(z_m\) of \(\mathbf{z}\) has been derived from a complete class lemma for the convexity hypothesis (1) and shown to be an appropriate statistic also for testing the slope change-point model (2) (Hirotsu, 1982; Hirotsu and Marumo, 2002). Further a very efficient algorithm for probability calculation of \(z_m\) for any \(K\) has been proposed based on the second order Markov property of the successive elements of \(\mathbf{z}\). Those properties have been derived only by the covariance structure in case of the normal model since the key vector \(\mathbf{z}\) is a linear function of \(\mathbf{y}\). For a general exponential family the key vector \(\mathbf{z}\) becomes \(\mathbf{z} = (\mathbf{L}_K^\alpha \mathbf{L}_K^\alpha)^{-1} \mathbf{L}_K^\alpha \mathbf{\nu}_0\) with \(\mathbf{\nu}_0\) an efficient score vector evaluated at the null hypothesis as shown in Section 2, and it looks formidable to develop an exact analysis since an explicit expression of \(\mathbf{z}\) is so complicated. In this paper, however, the ideas for the normal model are extended generally to the discrete univariate exponential family and an exact analysis is developed. Firstly in Section 3 it is shown that the component statistics forming \(\mathbf{z}\) are essentially the doubly-accumulated statistics and the maximal standardized component of them is proposed as the test statistic. This is a very important finding for extending the normal theory to a general discrete univariate exponential family. In Section 4 an approach to constructing the joint conditional distribution of the component statistics is developed given the complete sufficient statistics under the null model for developing a similar test. Such a conditional distribution given plural conditional variables is quite unfamiliar for the discrete random variables. The second order Markov property of the successive component statistics and the factorization of their joint distribution into the products of conditional probabilities are shown. In Section 5 a bottom up procedure is developed for determining the probability function definitely and some inequalities are introduced for executing the recursion formula efficiently. Then Section 6 is devoted for calculating the tail probability and moments of the maximal contrast test statistic. As a promising statistic for a directional goodness-of-fit test of a dose–response model the cumulative chi-squared is introduced in Section 7, which is defined as the sum of squares of the standardized elements of the key vector. In Section 8 some power comparisons are made to show the robustness of the proposed methods. In Section 9 these ideas are extended to testing the sigmoicidal and inflection point model. Then the basic statistics are triply-accumulated statistics and the related statistical property is the third order Markov. In Section 10 a more direct method for calculating \(p\)-value is given for the non-explosive sequence for both of the maximal and cumulative chi-squared statistics.
An interesting application of the convexity test will be in the monitoring of spontaneous reporting of the adverse drug events. The proposed method is a natural extension of the max acc t which has been proved to have high power for detecting an increasing tendency, and should therefore be useful for verifying that the increasing tendency has turned to decreasing after an appropriate action. A real example of this is given in Section 11.1. Another important application is a directional goodness-of-fit test for linearity of a dose–response model. Information concerning the shape whether it is linear, convex or sigmoid is useful in predicting the behaviour of the dose–response after starting up, and also influencing choice of an appropriate model or test for the dose–response (Schmoyer, 1984; Damgaard et al., 2002). It gives useful information also for extrapolating the dose–response curve to extremely low dose level (Hirotu and Srivastava, 2000). Examples of testing the linearity of a logit linear model against convex and sigmoid departures are given in Sections 11.2 and 11.3, respectively. Some concluding remarks are mentioned in Section 12, which includes an extension to the two-way interaction problem. Finally Appendices are for the efficient score test of the slope change-point model \( M_k(2) \) and inequalities for determining the range of doubly-accumulated statistics.

2. The key vector for testing the convexity hypothesis

Let \( y_i \) be distributed independently as a discrete univariate exponential family

\[
a(\mu_i) b(y_i) \exp(\mu_i y_i), \quad i = 1, \ldots, K.
\]

A general case of unequal spacing of events is considered and the convexity hypothesis \((1)\) is defined in the natural parameter \( \psi = (\psi_1, \ldots, \psi_K)' \). Then according to a complete class lemma in Hirotu (1982) an appropriate test for the convexity hypothesis should be increasing in every element of \( z = (L^*_K L^*_K)^{-1} L^*_K \tau \), where \( \tau = y - \tilde{E}_0(y) \) is an efficient score vector with respect to \( \psi \) evaluated at the null model \( M_0(3) \) with \( y = (y_1, \ldots, y_K)' \) and \( \tilde{E}_0(y) \) the maximal likelihood estimator of the expectation of \( y \) under \( M_0 \). \( \tilde{E}_0(y) \) is obviously a function of the complete sufficient statistics \( Y_K = y_1 + y_2 + \cdots + y_K \) and \( T_K = x_1 y_1 + x_2 y_2 + \cdots + x_K y_K \), where the cumulative sum \( Y_K \) corresponds to the parameter \( \beta_0 \) and the weighted sum \( T_K \) to \( \beta_1 \). It should be noted that the statistics \( Y_K \) and \( T_K \) have been introduced for the statistical analysis of the log linear model \( \mu_i = \log \lambda_i = \beta_0 + \beta_1 x_i \) in the Poisson sequence with mean \( \lambda_i \) in Cox and Lewis (1966). Here it is a null model and the \( T_K \) is a conditioning variable as well as \( Y_K \) for developing a similar test.

Very interestingly each corner vertex of the convex polyhedral cone \((1)\) represented by \( L^*_K (L^*_K L^*_K)^{-1} \) corresponds to the slope change-point model

\[
M^*_k : \quad \mu = B \beta_0 \beta_1' + b_k \tau_k,
\]

\[
B = \begin{bmatrix}
1 & 1 & \cdots & 1 \\
x_1 & x_2 & \cdots & x_K
\end{bmatrix}, \quad \Pi_\theta = B (B' B)^{-1} B',
\]

\[
b_k = (I - \Pi_\theta) (0, 0, x_{k+2} - x_{k+1}, \ldots, x_K - x_{k+1})'.
\]

By a simple algebra \( M^*_k \) is shown to be another expression of \( M_k(2) \) with the amount of slope change \( \tau_k \). The matrix \( I - \Pi_\theta \) is equal to \( \Pi_\theta L^*_K = L^*_K (L^*_K L^*_K)^{-1} L^*_K \) since \( \Pi_\theta \) and \( \Pi_\theta L^*_K \) are projection matrices of order \( 2 \) and \( K - 2 \) and orthogonal to each other. It is also easy to verify \( L^*_K (0, 0, x_{k+1} - x_{k+1}, \ldots, x_K - x_{k+1}) \) to be equal to \( (0 \cdots 0 \cdots 0)' \) with a unit element as its kth component, which implies the equality \( B_1 b_2 \ldots b_{K-1} = L^*_K (L^*_K L^*_K)^{-1} \) suggesting a relationship between the convexity hypothesis and a slope change-point model. Among the tests belonging to the complete class a maximal contrast test is proposed below. Then as shown in Appendix A it also gives the efficient score test for the slope change-point hypothesis \( M_k, k = 1, \ldots, K - 2 \).

3. The basic variable extracted from \( (L^*_K L^*_K)^{-1} L^*_K \tilde{v}_0 \)

Since a similar test is developed conditionally on the sufficient statistics \((Y_K, T_K)\) the essential part \( t = (L^*_K L^*_K)^{-1} L^*_K y \) of \( z = (L^*_K L^*_K)^{-1} L^*_K \tilde{v}_0 \) is kept by discarding \( \tilde{E}_0(y) \) from \( \tilde{v}_0 \). After standardization under the null distribution given \((Y_K, T_K)\) the maximal elements based on \( t \) and \( z \) will coincide. However, this form of \( t \) is still not very convenient for dealing with the discrete model since the elements of \( t \) do not take integer values because of the complicated coefficient matrix. Therefore a more convenient basic variable for dealing with the discrete distribution is searched for. First we have the equation \( L^*_K t = L^*_K (L^*_K L^*_K)^{-1} L^*_K y = (I - \Pi_\theta) y \) so that \( y = L^*_K t + \Pi_\theta y \), where \( \Pi_\theta y \) is a function of the sufficient statistics \((Y_K, T_K)\) and can be discarded for developing a similar test. Then the similar test should be based on \( S = (S_1, \ldots, S_{K-2})' \) that satisfies \( y = L^*_K S \) instead of \( t \), where we ignore the last two rows of this equation. By accumulating both sides of \( y = L^*_K S \) the equations \( Y_k = y_1 + \cdots + y_k = (S_k - S_{k-1})/(x_{k+1} - x_k) \), \( k = 1, \ldots, K - 2 \), are obtained, where \( S_0 \) is defined to be zero. By accumulating \( Y_k \) further after multiplying \((x_{k+1} - x_k)\) the equation

\[
S_k = \sum_{i=1}^{k} (x_{i+1} - x_i) Y_i = (x_{k+1} - x_k) Y_1 + \cdots + (x_{k+1} - x_k) Y_k, \quad k = 1, \ldots, K - 2,
\]

C. Hirotu et al. / Computational Statistics and Data Analysis 97 (2016) 33–46
is obtained. The statistics $S_k$ (5) reduce to the doubly-accumulated statistics of $y_k$ in case of equal spacing like $S_k = \sum Y_k = ky_1 + (k - 1)y_2 + \cdots + y_k$. The extended Eq. (5) of $S_k$ is called also the doubly-accumulated statistic hereafter and the maximal standardized element $s_m^*$ of $S_k$ is employed as the test statistic. This finding of the basic variables makes an exact analysis of the discrete model possible extending the normal theory of Hirotsu and Marumo (2002).

4. Markov property and factorization of the conditional distribution in terms of $S_k$

In case of the normal distribution the conditional distribution given the complete sufficient statistics and its factorization are very simple resulting in the respective normal distributions defined by the conditional mean and variance. In case of the Poisson and binomial distributions the conditional distributions given $Y_k$ only and their factorization into the products of successive conditional distributions are also well known. However, an additional conditioning variable $T_k$ is here and several steps are required to obtain the exact conditional distribution in a tractable form. The conditional null distribution given $(Y_k, T_k)$ is in the form $G(y|Y_k, T_k) = C_{k-1}^{-1}(Y_k, T_k) \prod_{y} b(y|k)$, where the constant $C_{k-1}(Y_k, T_k)$ is determined so that the total probability of $y$ is unity conditionally given $Y_k$ and $T_k$. Then by the relationship

$$S_k = S_{k-1} + (x_{k+1} - x_k)Y_k, \quad k = 1, \ldots, K - 2,$$

the equation

$$G(y|Y_k, T_k) = C_{k-1}^{-1}(Y_k, T_k) \prod_{y} b\left(\frac{S_k - S_{k-1}}{x_{k+1} - x_k} - \frac{S_{k-1} - S_{k-2}}{x_k - x_{k-1}}\right)$$

is obtained, where $S_0$ and $S_1$ are defined to be zero and $S_{k+1} = x_k Y_k - T_k$, $S_k = S_{k-1} + (x_{k+1} - x_k)Y_k$ with $x_k$ an arbitrary number. It should be noted that the $S_{k-1}$ is an extension of the definition (5) to $k = K - 1$ but it cannot be done for $S_K$ without introducing a hypothetical value $x_K$. For notational convenience anyway $S_{k-1}$ and $S_k$ are employed as conditioning variables instead of $Y_k$ and $T_k$. Also $S_{k-1}$ and $Y_k$ can be taken as conditioning variables. The one to one correspondence among the sets of variables $(Y_k, T_k)$, $(S_{k-1}, S_k)$ and $(S_{k-1}, Y_k)$ is obvious. Eq. (7) implies the second order Markov property of the sequence $S_1, \ldots, S_{K-2}$. The proof is essentially the same as that of Lemma 3.2 in Hirotsu et al. (1992) and omitted. Because of this property there is available a very efficient and exact algorithm for calculating the $p$-value of the maximal statistic $s_m^*$ as well as the calculation of $C_{k-1}(S_{k-1}, S_k)$ and the moments.

Now by virtue of the second order Markov property the null distribution $G(y)$ can be factorized in terms of $S_k$ like $G(y) = \prod_{k=1}^{K-2} f_k(S_k|S_{k+1}, S_{k+2})$, where $f_k(S_k|S_{k+1}, S_{k+2})$ is the conditional distribution of $S_k$ given $S_{k+1}$ and $S_{k+2}$. Then the $k$th conditional distribution should be in the form of

$$f_k(S_k|S_{k+1}, S_{k+2}) = C_{k+1}^{-1}(S_{k+1}, S_{k+2})C_k(S_k, S_{k+1})b\left(\frac{S_{k+2} - S_{k+1}}{x_{k+3} - x_{k+2}} - \frac{S_{k+1} - S_k}{x_{k+2} - x_{k+1}}\right), \quad k = 1, \ldots, K - 2,$$

where $C_{k+1}$ is the normalizing constant and the initial constant is defined as

$$C_1(S_1, S_2) = b\left(\frac{S_1}{x_2 - x_1}\right) b\left(\frac{S_2 - S_1}{x_3 - x_2} - \frac{S_1}{x_2 - x_1}\right).$$

It should be noted that in Eq. (8) the random variable $S_k$ is included also in the normalizing constant $C_k(S_k, S_{k+1})$ in the previous step as the conditioning variable. Starting from $C_1$ all the $C_k$ can be calculated recursively by the equation

$$C_{k+1}(S_{k+1}, S_{k+2}) = \sum_{S_k} C_k(S_k, S_{k+1})b\left(\frac{S_{k+2} - S_{k+1}}{x_{k+3} - x_{k+2}} - \frac{S_{k+1} - S_k}{x_{k+2} - x_{k+1}}\right).$$

Then at the final step the overall normalizing constant $C_{K-1}(S_{K-1}, S_K)$ is obtained and $G(y)$ is determined in terms of $S_k$, $k = 1, \ldots, K - 2$. It should be noted that the $C_{k-1}(S_{k-1}, S_k)$ is well defined since $(S_k - S_{K-1})/(x_k - x_{k-1})$ is simply $Y_k$; however, in the following the notation $C_{k-1}(S_{k-1}, Y_k)$ is employed instead of $C_{k-1}(S_{k-1}, S_k)$. This factorization of the simultaneous distribution is following the same idea as employed in Hirotsu et al. (2001) for obtaining an exact factorization of the null distribution of no three factor interaction in a $2 \times J \times K$ contingency table.

5. Constructing the conditional probabilities by a bottom up procedure

The construction of the conditional probabilities $f_k(S_k|S_{k+1}, S_{k+2})$ should be bottom up since $C_k(S_k, S_{k+1})$ is necessary for calculating $C_{k+1}(S_{k+1}, S_{k+2})$. For executing the recursion formula (10) efficiently several inequalities for the range of $S_k$ are introduced below.

**Absolute**:

$$S_{k-1} - (x_k - x_{k+1})Y_k \leq S_k \leq x_{k+1} - x_k S_{k-1}, \quad k = 1, \ldots, K - 2.$$  \hspace{1cm} (11)

**Relative**:

$$\frac{x_{k+1} - x_k}{x_k - x_1} S_{k-1} \leq S_k \leq \frac{1}{x_k - x_k} (x_k - x_{k+1}) S_{k-1} + (x_{k+1} - x_k) S_{k-1}, \quad k = 2, \ldots, K - 2.$$  \hspace{1cm} (12)

$$(S_{k+1} - S_k)/(x_{k+2} - x_{k+1}) - (S_k - S_{k-1})/(x_k - x_{k+1}) \geq 0, \quad k = 2, \ldots, K - 2.$$  \hspace{1cm} (13)
The derivation of these inequalities are given in Appendix B. Further by inverting the inequality (12) the inequality (14) is defined,

\[
\frac{1}{X_k - x_{k+1}} \{ (x_k - x_k)S_k - (x_{k+1} - x_k)S_{k-1} \} \leq S_{k-1} \leq \frac{x_k - x_1}{x_{k+1} - x_1}, \quad k = 2, \ldots, K - 2.
\]

These inequalities are really useful since the sample space of \( S = (S_1, \ldots, S_{K-2}) \) is unknown and explosive for large \( K \) as shown in Table 3.

First the possible combinations of \((S_1, S_2)\) satisfying the inequalities (11) and (12) are found and \( C_1(S_1, S_2) \) is calculated by (9). Then for those combinations of \( S_1 \) and \( S_2 \) the range of the variable \( S_3 \) can be determined by the inequalities (11)–(13). Then the coefficient \( C_2(S_2, S_3) \) can be calculated by Eq. (10) like

\[
C_2(S_2, S_3) = \sum_{S_1} C_1(S_1, S_2)b \{ (S_3 - S_2)/(x_4 - x_3) - (S_2 - S_1)/(x_3 - x_2) \},
\]

where the summation is with respect to \( S_1 \). In executing the recursion formula, however, a difficulty arises since the \( S_k \) satisfying Eq. (5) does not necessarily take successive integers differently from \( Y_k \). It does not occur in the equal spacing where \( x_k \) can be defined as \( k \) without any loss of generality. Therefore the inequalities are converted for \( Y_k \) which takes successive integers and generate conformable \( S_k \) through Eq. (6). The inequality for \( Y_k \) is as follows.

\[
\max \left\{ \frac{S_{k-1} - S_{k-1} - (x_k - x_{k+1})Y_k}{x_{k+1} - x_k}, \frac{S_{k-1} - x_1}{x_k - x_1}, Y_{k-1} \right\} \leq Y_k
\]

\[
\leq \min \left\{ \frac{S_{k-1} - S_{k-1}}{x_k - x_k}, \frac{1}{x_{k+1} - x_k} \left( \frac{x_{k+1} - x_1}{x_k - x_1} S_{k-1} - S_{k-1} \right) \right\}.
\]

A simple example is given below to construct the conformable sequence.

**Example 1.** Generating the sequence \( S \) conformable to \( y = (1, 1, 1, 1)' \) with \( x = (1, 2, 4, 7, 8)' \).

In this example the sufficient statistics are \( Y_5 = 1 + 1 + 1 + 1 + 1 + 1 = 5 \) and \( S_4 = 8Y_5 - T_5 = 7Y_1 + 6Y_2 + 4Y_3 + Y_4 = 18 \). Then by the absolute inequality (11) the range of \( S_1 = Y_1 \) is obtained as \( -12 \leq S_1 = Y_1 \leq \frac{18}{7} \) and therefore \( S_1 \) and \( Y_1 \) take 0, 1 and 2. Then by the inequality (15) the conformable \( Y_2 \) for \((Y_1, S_1) = (0, 0), (1, 1) \) and \[(2, 2) \) are \( 0 \leq Y_2 \leq 3, 1 \leq Y_2 \leq 2 \) and \( Y_2 = 2 \), respectively. Therefore the respective conformable \( S_2 \) are \((0, 2, 4, 6), (3, 5) \) and \((6) \) by \( S_2 = S_1 + 2Y_2 \). Then from the pair \((Y_2, S_2)\) the conformable \( Y_3 \) can be determined by the inequality (15) and then \( S_3 \) from \( S_3 = Y_3 \) by (6). The process is summarized in Table 1, where NG implies there is no conformable \( Y_3 \) satisfying the inequality.

In this example the number of partitions of integer 5 into five integers is 126 among which only five sequences in Table 1 are conformable also to \( S_4 = 18 \). On the other hand under the equal spacing \((x_k = k)\) 12 sequences are found to be conformable to the fixed sufficient statistics \( S_4 = 10 \) and \( Y_5 = 5 \).

Given the underlying distribution the initial constant \( C_1(S_1, S_2) \) can be calculated by Eq. (9) at stage 2. At stage 3 the normalizing constant \( C_2(S_2, S_3) \) and the conditional probability \( f(S_1|S_2, S_3) \) can be calculated by Eqs. (10) and (8), respectively. At stage 4 \( C_4 \) and \( f_2 \) are calculated, at stage 5 \( C_4 \) and \( f_3 \) are calculated and finally the total probability \( G(y|S_4, Y_5) \) is obtained by the products of \( f_1, f_2 \) and \( f_3 \).

As an example Poisson model with mean \( \lambda_i \) can be dealt with taking \( \mu_i = \log \lambda_i, a(\mu_i) = \exp(-\lambda_i) = \exp(-\exp(\mu_i)) \) and

\[
b(y_i) = (y_i!)^{-1} = \left\{ \frac{S_i - S_{i-1}}{x_{i+1} - x_i} - \frac{S_{i-1} - S_{i-2}}{x_i - x_{i-1}} \right\}^{-1}.
\]

Starting from the initial constant (9) the normalizing constants and the conditional probabilities are calculated recursively by (10) and (8), respectively. To continue the current example at stage 2 we obtain

\[
C_1(0, 2) = \left\{ \frac{0}{2-1} \right\}^{-1} \left\{ \frac{2 - 0}{4 - 2} - \frac{2}{2-1} \right\} = 1
\]
and similarly $C_1(0, 4) = \frac{1}{3}$, $C_1(0, 6) = \frac{1}{6}$, $C_1(1, 5) = 1$, $C_1(2, 6) = \frac{1}{2}$. At stage 3 noting that the pair $(S_2, S_3) = (2, 14)$ has only one root $S_1 = 0$ the $C_2(2, 14)$ is obtained as

$$C_2(2, 14) = C_1(0, 2) \times \left( \left( \frac{14 - 2 - 2 - 0}{7 - 4 - 4 - 2} \right) \right)^{-1} = \frac{1}{6}$$

and the conditional probability $f_1(0|2, 14)$ is equal to unity. On the other hand the pair $(S_2, S_3) = (6, 15)$ has two roots $S_1 = 0$ and 2 so that

$$C_2(6, 15) = \frac{1}{6} \times \left( \left( \frac{15 - 6 - 6}{3 - 2} \right) \right)^{-1} + \frac{1}{2} \times \left( \left( \frac{15 - 6 - 2}{3 - 2} \right) \right)^{-1} = \frac{2}{3}.$$ 

Then the conditional probabilities are calculated as

$$f_1(0|6, 15) = \left( \frac{2}{3} \right)^{-1} \times \frac{1}{6} \times \left( \left( \frac{15 - 6 - 6}{3 - 2} \right) \right)^{-1} = \frac{1}{4},$$

$$f_1(2|6, 15) = \left( \frac{2}{3} \right)^{-1} \times \frac{1}{2} \times \left( \left( \frac{15 - 6 - 6}{3 - 2} \right) \right)^{-1} = \frac{3}{4}.$$ 

It should be noted that inequalities (13) and (14) are useful for finding out the roots $S_{k-1}$ for the pair $(S_k, S_{k+1})$. Other $C_2$’s are $C_2(4, 13) = \frac{1}{4}$, $C_2(5, 14) = 1$ and the conditional probabilities $f_1(0|4, 13), f_1(1|5, 14)$ are both unity since those pairs $(S_2, S_3)$ have only one root $S_1 = 0$ and 1, respectively. The calculation of $C_3(S_3, S_4), C_4(S_4, S_5)$ and the conditional probabilities $f_2(S_2|S_3, S_4), f_3(S_3|S_4, S_5)$ goes similarly as $C_2(S_2, S_3)$ and $f_1(S_1|S_2, S_3)$, respectively, and the result is summarized in Table 2. In this small example all the conformable sequences might be written down one by one but it is generally impossible to figure out all the sequences even for a moderate $K$ and $\lambda$, and the bottom up procedure utilizing those inequalities is inevitable. In Table 3 we give the number of conformable sequences and the computing time of probabilities by the methods with or without recursion formula to the outcome $y_i \equiv 2, i = 1, \ldots, K$, for some $K$ assuming a Poisson model. It is easily explosive at around $K = 15$ and the naive calculation method becomes soon impossible.
6. Recursion formulae for moments and tail probability

6.1. Calculating the moments

For standardizing the test statistics it is necessary to calculate the mean and variance of $S_k$, $k = 1, \ldots, K - 2$. They are calculated also recursively by the formula (16).

$$E(S_k^*|S_{k-1}, S_K) = \sum_{S_{k-2}} \sum_{S_{k+1}} \sum_{S_k} S_k^* f_k(S_k|S_{k+1}, S_{k+2}) \times f_{k-1}(S_{k-1}|S_{k+2}, S_{k+3}) \times \cdots \times f_{K-2}(S_{K-2}|S_{K-1}, Y_K).$$

(16)

6.2. Calculating the $p$-value of the convexity test

Let $S_k^*$ be the standardized version of $S_k$, $S_k^* = \{S_k - E(S_k)/\text{var}^{1/2}(S_k)\}$, $k = 1, \ldots, K - 2$. Then the test statistic is $s_m^* = \max_{k=1,\ldots,K-2} S_k^*$. For the recursion formula define the conditional probability

$$F_k(S_{k-1}, S_k) = \Pr(S_k^* < d, \ldots, S_k^* < d|S_{k-1}, S_k) = \Pr(S_k < d_1^*, \ldots, S_k < d_k^*|S_{k-1}, S_k), \quad k = 2, \ldots, K,$$

where $d_k^* = E(S_k) + \text{var}^{1/2}(S_k) d$, $k = 1, \ldots, K$, and $S_k^*$ and $S_k^*$ are defined to be $-\infty$ although their conditional variances are zero so that the inequality holds always. Correspondingly $d_{k-1}^*$ and $d_k^*$ are set as $S_{k-1} + \delta$ and $S_k + \delta$, respectively, with $\delta$ a positive small number. Then a recursion formula for $F_k$ is obtained as

$$F_{k+1}(S_k, S_{k+1}) = \sum_{S_{k-1}} \Pr(S_k < d_1^*, \ldots, S_k < d_k^*|S_{k-1}, S_k, S_{k+1}) \times f_{k-1}(S_{k-1}|S_k, S_{k+1})$$

(17)

$$= \left\{ \begin{array}{ll}
\sum_{S_{k-1}} \Pr(S_k < d_1^*, \ldots, S_k < d_k^*|S_{k-1}, S_k) \times f_{k-1}(S_{k-1}|S_k, S_{k+1}) & \text{if } S_k < d_k^*, \\
0 & \text{otherwise.}
\end{array} \right.$$

(18)

Eq. (17) is due to the law of total probability and Eq. (18) is due to the second order Markov property of $S_k$. Thus essentially the recursion formula is obtained as

$$F_{k+1}(S_k, S_{k+1}) = \sum_{S_{k-1}} F_k(S_{k-1}, S_k) \times f_{k-1}(S_{k-1}|S_k, S_{k+1}).$$

(19)

It is straightforward to extend the formula to $F_k$ by taking $S_{k-1}$ and $Y_k$ as fixed values satisfying the inequality. Then the $p$-value of the observed maximum $s_m^*$ is obtained at the final step by $p = 1 - F_k(S_{k-1}, Y_k)$ at $d = s_m^*$. It should be noted that the procedure is converting the multiple summations which are computationally very awkward into the repetitions of single summation so that the calculation is feasible for large $K$. This is an extension of the methods by Hawkins (1977), Worsley (1986) and Hirotsu et al. (2001).

Again the procedure is explained via an example employed in Section 5 with the observed sequence $y = (1, 1, 1, 1, 1)'$. First the moments of $S_k$ are calculated by the method of Section 6.1 to obtain

$$E(S_1) = 6/7, \quad E(S_2) = 100/21, \quad E(S_3) = 295/21, \quad \text{var}(S_1) = 20/7^2, \quad \text{var}(S_2) = 500/21^2, \quad \text{var}(S_3) = 146/21^2.$$

Then the standardized statistics are obtained as $S_1^* = 0.22361$, $S_2^* = 0.22361$, $S_3^* = 0.08276$ and therefore $s_m^* = 0.22361$, which gives $d_1^* = 1.000$, $d_2^* = 5.000$, $d_3^* = 14.1762$. Of course $d_2^*$ and $d_3^*$ are set as $18 + \delta$ and $S_k + \delta$. The calculation of the $p$-value by the recursion formula goes as follows.

First $F_2(S_1, S_2)$ is set to unity for possible combination of $(S_1, S_2)$ with $S_1 < 1.000$ and $S_2 < 5.000$, namely for $F_2(0, 2)$ and $F_2(0, 4)$ while it is set to zero for other combinations. Then the possible combinations of $(S_2, S_3)$ with $S_2 < 5.000$ and $S_3 < 14.176$ are found to be, $(2, 14)$ and $(4, 13)$. Since those combinations have only one root $(S_1, S_2) = (0, 2)$ and $(0, 4)$, respectively, $F_3(2, 14)$ and $F_3(4, 13)$ are found to be unity while $F_3$ is set to zero for other combinations. Then the $F_k's$ are calculated as $F_k(13, 18) = 1 	imes 1 = 1$ and $F_k(14, 18) = 1 	imes 1/7 = 1/7$. Finally the $F_k(18, 5)$ is calculated as $F_2(18, 5) = 1 \times 1/7 + 1/7 \times 2/3 = 5/21$ and therefore the $p$-value is obtained as $p = 1 - 5/21 = 16/21 = 0.762$. This is the sum of the probabilities of the last three sequences of Table 2.

This approach based on $S_k$ might be called as max acc. t2 method in contrast to max acc. $t$ based on $Y_k$ since $S_k$ is an accumulation of $Y_k$ thus doubly accumulating $Y_k$’s. It is naturally expected to inherit many of good properties of max acc. $t$ which have been proved in various situations.
6.3. Calculating the p-value of the concavity test

For testing the concavity hypothesis \( \tilde{C} : \mathbf{L}^*_K \mu \leq 0 \), with at least one inequality strong, the test should be based on \( -S_k \), and the maximal statistic \( s_{1n}^* = \max_{k=1, \ldots , K-2} S_k^{**} \) is calculated from the standardized statistics \( S_k^{**} = (-S_k - E(-S_k))/\sqrt{K \delta} \), \( k = 1, \ldots , K - 2 \). Then only a slight modification is necessary from the previous section. Define the conditional probability \( F_k(S_{k-1}, S_k) = \text{pr}(S_k^{**} < d, \ldots , S_k^{**} < d|S_{k-1}, S_k) = \text{pr}(d_k < \delta_k, \ldots , S_k > |S_k|_{\infty}, S_k) = \text{pr}(S_k^{**} < d_k, \ldots , S_k^{**} < S_k - |S_k|_{\infty}, S_k) = 2, \ldots , K \), where \( d_k^{**} = E(S_k - \delta^2)_{\infty} | S_k, \delta > 0 \). The recursion formula (19) and p-value calculation are exactly the same.

The p-value for the two-sided alternative \( \tilde{C} : \mathbf{L}^*_K \mu \geq 0 \) or \( \mathbf{L}^*_K \mu \leq 0 \) with at least one inequality strong, can be obtained by defining \( s_{1n}^* = \max\{S_k^*\} = \max\{S_k^{**}\} \) and \( F_k(S_{k-1}, S_k) = \text{pr}(S_k^{**} < d, \ldots , |S_k^{**}| < d|S_{k-1}, S_k) = \text{pr}(d_k^{**} < S_k - \delta, d_k^{**} < S_k - \delta \delta > 0 \). The recursion formula and p-value calculation are exactly the same.

The p-value for the two-sided alternative \( \tilde{C} : \mathbf{L}^*_K \mu \geq 0 \) or \( \mathbf{L}^*_K \mu \leq 0 \) with at least one inequality strong, can be obtained by defining \( s_{1n}^* = \max\{S_k^*\} = \max\{S_k^{**}\} \) and \( F_k(S_{k-1}, S_k) = \text{pr}(S_k^{**} < d, \ldots , |S_k^{**}| < d|S_{k-1}, S_k) = \text{pr}(\hat{d}_k^{**} < S_k - \delta, d_k^{**} < S_k - \delta \delta > 0 \). The recursion formula (19) and p-value calculation are exactly the same.

7. Cumulative chi-squared statistic

An important application of convexity and concavity test is a directional goodness-of-fit test of the linearity of a dose–response model. For the purpose the cumulative chi-squared statistic is proposed. It is denoted by \( \chi^2 \) and defined by the sum of squares of the standardized components of the key vector: \( \chi^2 = \sum_k K - 2 S_k^2 \). Its null distribution is well approximated by the constant times chi-squared distribution of \( d \chi^2 \), where constants \( d \) and \( f \) are determined adjusting first two cumulants, \( df = E\left\{\sum_{i=1}^{K-2} S_i^2\right\} = K - 2 \), \( 2d f = E\left\{\sum_{i=1}^{K-2} S_i^4 + 2 \sum \sum_{1 \leq k < l \leq K-2} (S_k^2 S_l^2)\right\} - (K - 2)^2 \). The calculation of joint moments can be carried out extending the formula in Section 6.1. If necessary an improved formula has been obtained in Hirotsu (1979) which adjusts the first three cumulants.

8. Power comparisons

The max acc. \( t \) for the monotone hypothesis has been verified to keep high power for the wide range of the simple ordered alternative. For the normal model max acc. \( t \) is also shown to keep relatively high power in the wide range of the convexity hypothesis in Hirotsu and Marumo (2002). In the following we compare the power of max acc. \( t \) and the cumulative chi-squared \( \chi^2 \) with the test of the quadratic term of the logistic regression model by simulation of 10,000 replications at the significance level 0.05 assuming the binomial distribution with \( n_i = 20 \). We call the last test as a polynomial test according to Hirotsu and Marumo (2002). The comparisons are made as the two-sided tests in the direction of corner vectors and also for the logistic quadratic pattern for each of \( K = 6 \) and 10. The corner vectors and the quadratic pattern are as follows:

For \( K = 6 \):

- Corner vector 1: \(-10, 8, 5, 2, -1, -4\)
- Corner vector 2: \(-20, 2, 24, 11, -2, -15\)
- Corner vector 3: \(-15, -2, 11, 24, 2, -20\)
- Corner vector 4: \(-4, -1, 2, 5, 8, -10\)
- Quadratic pattern: \(-5, 1, 4, 4, 1, -5\)

For \( K = 10 \):

- Corner vector 1: \(-36, 16, 13, 10, 7, 4, 1, -2, -5, -8\)
- Corner vector 2: \(-168, -28, 112, 87, 62, 37, 12, -13, -38, -63\)
- Corner vector 3: \(-189, -70, 49, 168, 122, 76, 30, -16, -62, -108\)
- Corner vector 4: \(-36, -17, 2, 21, 40, 26, 12, -2, -16, -30\)
- Corner vector 5: \(-30, -16, -2, 12, 26, 40, 21, 2, -17, -36\)
- Corner vector 6: \(-108, -62, -16, 30, 76, 122, 168, 49, -70, -189\)
- Corner vector 7: \(-63, -38, -13, 12, 37, 62, 87, 112, -28, -168\)
- Corner vector 8: \(-8, -5, -2, 1, 4, 7, 10, 13, 16, -36\)
- Quadratic pattern: \(-6, -2, 1, 3, 4, 4, 3, 1, -2, -6\)

Then the polynomial test should be asymptotically most powerful for the quadratic pattern among all the tests. The non-centrality parameter \( \sum \mu_i^2 \) is fixed for the power of the polynomial test to be around 0.80 for the quadratic pattern. The simulation results are shown in Table 4 also for the null model, where Const. means \( \mu_i = \text{const} \) and Slope means the regression model \( \mu_i = \beta_0 + \beta_i \). There is observed some conservatism for max acc. \( t \) but it is inevitable for the exact discrete test and in the acceptable range. On the other hand the other tests become sometimes slightly aggressive. The similarity between the corner vectors at the symmetric location suggests the reliability of the simulation. It is verified that the max acc. \( t \) keeps relatively high power in the wide range of the convexity hypothesis. The polynomial test looks very good when the change-point is located in the middle but too bad when it is in the end. It is because the test does not belong to the complete class and it cannot be recommended without any prior information on the location of the mean vector. The cumulative chi-squared \( \chi^2 \) looks to be located between the max acc. \( t \) and polynomial test. It is expected from that the
9. Testing the sigmoidicity and inflection-point hypotheses

The sigmoidicity hypothesis is defined by

\[ S : \frac{1}{x_3 - x_1} \left( \frac{\mu_3 - \mu_2}{x_3 - x_2} - \frac{\mu_2 - \mu_1}{x_2 - x_1} \right) \geq \cdots \geq \frac{1}{x_K - x_{K-2}} \left( \frac{\mu_K - \mu_{K-1}}{x_K - x_{K-1}} - \frac{\mu_{K-1} - \mu_{K-2}}{x_{K-1} - x_{K-2}} \right) \]  

(20)

extending the convexity hypothesis of Section 2. The null hypothesis is defined by all the equalities in (20), which is equivalent to \( H_0 : \mu_i = \beta_0 + \beta_1 x_i + \beta_2 x_i^2 \). The inequalities (20) can be expressed in the matrix form as \( Q_{n}^\mu \geq 0 \) just like Eq. (1) for the convexity hypothesis. The explicit form of \( Q_{n}^\mu \) and its relationship with the inflection point model are given in Hirotsu and Marumo (2002). The triply-accumulated statistics are derived as the basic variables after a similar but somewhat more complicated algebra as in deriving the doubly-accumulated statistics for the convexity hypothesis in Section 3. They are denoted by \( W_k \) and expressed in terms of \( S_k \) as \( W_k = \sum_{i=1}^{K} (x_{i+2} - x_i) S_{i+k} \), \( k = 1, \ldots, K - 2 \). Then the maximal contrast test is defined by \( w_{n,k}^\alpha = \max_k W_k^\alpha \), where \( W_k^\alpha \) is the standardized version of \( W_k \), \( k = 1, \ldots, K - 3 \). The sufficient statistics under the null model are \( Y_k, S_{K-1} \) and \( W_{K-2} \) and we can derive the factorization of the conditional probability \( f(W_1, \ldots, W_{K-3}|Y_K, S_{K-1}, W_{K-2}) \) into the products of conditional probabilities \( \Pi_{1}^{K-3}f_k(W_k|W_{k+1}, W_{k+2}, W_{k+3}) \). Here we employed the notation \( W_k \) and \( W_{K-1} \) for convenience but actually they are expressed by the sufficient statistics \( (Y_k, S_{K-1}, W_{K-2}) \). The calculation is essentially the same as given in Section 5 except the inequalities for \( W_k \).

### Table 4

Power comparisons of max acc. \( t_2 \), cumulative chi-squared and the polynomial tests.

<table>
<thead>
<tr>
<th>Alternative hypothesis (( \mu ))</th>
<th>Max acc. ( t_2 )</th>
<th>Cumulative chi-squared</th>
<th>Polynomial test</th>
</tr>
</thead>
<tbody>
<tr>
<td>Const.</td>
<td>0.047</td>
<td>0.053</td>
<td>0.053</td>
</tr>
<tr>
<td>Slope</td>
<td>0.046</td>
<td>0.052</td>
<td>0.046</td>
</tr>
<tr>
<td>Corner vector 1</td>
<td>0.681</td>
<td>0.649</td>
<td>0.590</td>
</tr>
<tr>
<td>Corner vector 2</td>
<td>0.712</td>
<td>0.749</td>
<td>0.750</td>
</tr>
<tr>
<td>Corner vector 3</td>
<td>0.715</td>
<td>0.744</td>
<td>0.752</td>
</tr>
<tr>
<td>Corner vector 4</td>
<td>0.678</td>
<td>0.650</td>
<td>0.593</td>
</tr>
<tr>
<td>Quadratic pattern</td>
<td>0.751</td>
<td>0.804</td>
<td>0.802</td>
</tr>
<tr>
<td>Const.</td>
<td>0.048</td>
<td>0.053</td>
<td>0.050</td>
</tr>
<tr>
<td>Slope</td>
<td>0.045</td>
<td>0.047</td>
<td>0.052</td>
</tr>
<tr>
<td>Corner vector 1</td>
<td>0.644</td>
<td>0.519</td>
<td>0.418</td>
</tr>
<tr>
<td>Corner vector 2</td>
<td>0.692</td>
<td>0.682</td>
<td>0.625</td>
</tr>
<tr>
<td>Corner vector 3</td>
<td>0.707</td>
<td>0.738</td>
<td>0.726</td>
</tr>
<tr>
<td>Corner vector 4</td>
<td>0.715</td>
<td>0.760</td>
<td>0.763</td>
</tr>
<tr>
<td>Corner vector 5</td>
<td>0.713</td>
<td>0.755</td>
<td>0.772</td>
</tr>
<tr>
<td>Corner vector 6</td>
<td>0.709</td>
<td>0.742</td>
<td>0.734</td>
</tr>
<tr>
<td>Corner vector 7</td>
<td>0.691</td>
<td>0.684</td>
<td>0.626</td>
</tr>
<tr>
<td>Corner vector 8</td>
<td>0.652</td>
<td>0.529</td>
<td>0.422</td>
</tr>
<tr>
<td>Quadratic pattern</td>
<td>0.738</td>
<td>0.799</td>
<td>0.808</td>
</tr>
</tbody>
</table>

leading term of \( \chi^2 \) is the chi-squared for the quadratic pattern in the expansion in the independent chi-squared series in the balanced normal model (Hirotsu, 1986). The comparisons between max acc. \( t_2 \) and polynomial test coincide very well with those of the normal case given in Hirotsu and Marumo (2002). Another advantage of the max acc. \( t_2 \) is that it can suggest a change-point.

for efficient execution of the bottom up procedure and details are omitted. The calculation of moments and tail probability for the sigmoidicity test are essentially the same as given in Section 6 for the convexity test although the conditioning variables...
Table 5
Spontaneous reporting of adverse events by a month at PMDA.

<table>
<thead>
<tr>
<th>k</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
<th>13</th>
<th>14</th>
<th>15</th>
</tr>
</thead>
<tbody>
<tr>
<td>y_k</td>
<td>1</td>
<td>4</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>3</td>
<td>0</td>
<td>4</td>
<td>1</td>
<td>3</td>
<td>0</td>
<td>2</td>
<td>4</td>
<td>3</td>
</tr>
<tr>
<td>k</td>
<td>16</td>
<td>17</td>
<td>18</td>
<td>19</td>
<td>20</td>
<td>21</td>
<td>22</td>
<td>23</td>
<td>24</td>
<td>25</td>
<td>26</td>
<td>27</td>
<td>28</td>
<td>29</td>
<td>30</td>
</tr>
<tr>
<td>y_k</td>
<td>3</td>
<td>2</td>
<td>4</td>
<td>1</td>
<td>4</td>
<td>1</td>
<td>4</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>k</td>
<td>31</td>
<td>32</td>
<td>33</td>
<td>34</td>
<td>35</td>
<td>36</td>
<td>37</td>
<td>38</td>
<td>39</td>
<td>40</td>
<td>41</td>
<td>42</td>
<td>43</td>
<td>44</td>
<td>45</td>
</tr>
<tr>
<td>y_k</td>
<td>1</td>
<td>4</td>
<td>1</td>
<td>4</td>
<td>2</td>
<td>3</td>
<td>7</td>
<td>3</td>
<td>3</td>
<td>4</td>
<td>1</td>
<td>5</td>
<td>4</td>
<td>5</td>
<td>6</td>
</tr>
<tr>
<td>k</td>
<td>36</td>
<td>37</td>
<td>38</td>
<td>39</td>
<td>40</td>
<td>41</td>
<td>42</td>
<td>43</td>
<td>44</td>
<td>45</td>
<td>46</td>
<td>47</td>
<td>48</td>
<td>49</td>
<td>50</td>
</tr>
<tr>
<td>y_k</td>
<td>2</td>
<td>4</td>
<td>9</td>
<td>3</td>
<td>1</td>
<td>3</td>
<td>3</td>
<td>4</td>
<td>3</td>
<td>3</td>
<td>5</td>
<td>2</td>
<td>2</td>
<td>0</td>
<td>4</td>
</tr>
<tr>
<td>k</td>
<td>51</td>
<td>52</td>
<td>53</td>
<td>54</td>
<td>55</td>
<td>56</td>
<td>57</td>
<td>58</td>
<td>59</td>
<td>60</td>
<td>61</td>
<td>62</td>
<td>63</td>
<td>64</td>
<td>65</td>
</tr>
<tr>
<td>y_k</td>
<td>1</td>
<td>4</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>1</td>
<td>3</td>
<td>4</td>
<td>3</td>
<td>5</td>
<td>2</td>
<td>2</td>
<td>0</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>k</td>
<td>66</td>
<td>67</td>
<td>68</td>
<td>69</td>
<td>70</td>
<td>71</td>
<td>72</td>
<td>73</td>
<td>74</td>
<td>75</td>
<td>76</td>
<td>77</td>
<td>78</td>
<td>79</td>
<td></td>
</tr>
<tr>
<td>y_k</td>
<td>4</td>
<td>2</td>
<td>2</td>
<td>4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

become three making the computation a little harder because of large memories for a large sequence. The maximal contrast test might be called max acc. t3.

For a directional goodness-of-fit test the cumulative chi-squared statistic $\chi^{22} = \sum_{k=1}^{K-3} W_k^2$ is proposed again as a promising test. The chi-squared approximation for its distribution is obtained in exactly the same way as given in Section 7 just by replacing $S_k^*$ by $W_k^*$ in calculating the cumulants. In Section 11.3 the method is applied to detecting a sigmoid departure from a logit linear model and improving the upper confidence limit of the risk at the lowest dose level.

10. A more direct method for calculating the $p$-value in the non-explosive sequence

If it is possible to write down all the conformable sequences in the case of non-explosive sample space, a simpler method for calculating $p$-value is available. First search for the conformable sequences by the bottom up procedure utilizing efficiently the inequalities and attach them the probabilities obtained by the products of conditional probabilities. Then sum up those probabilities for the sequences whose test statistics are equal to or larger than the observed test statistic. The procedure is applicable to both of the maximal contrast and the cumulative chi-squared tests and particularly useful for a directional goodness-of-fit test of the dose–response curve where the sequence is usually not large. This method is employed in Section 11.3.

11. Real examples

11.1. Detecting a change in the Poisson sequence

As a real example a sequence of spontaneous reporting of adverse events caused by the administration of a compound drug for interstitial pneumonia is analysed. The data in Table 5 are the number of reporting per a month since November 2003 to May 2010 at PMDA (Pharmaceutical and Medical Device Agency in Japan) and an independent Poisson sequence is assumed. It is a very serious problem at PMDA how to detect a significant change in a time series in as short time as possible. First applying the classical max acc. $t$ for the Poisson model a shift of means has been detected at November 2006 ($k = 37$) with $p$-value 0.001. Therefore it is interesting to confirm whether it became a decreasing tendency at some point after that time by an appropriate action. The max acc. $t2$ for the concavity hypothesis can be applied with the function $b(y_k) = \{S_k - 2S_{k-1} + S_{k-2}\}^{-1}$ since the data are taken every one month interval. A downturn is detected at October 2007 ($k = 48$) with observed maximum $S_k^* = 2.858$ and a $p$-value of 0.0093 calculated by the formula in Section 6.3.

For dealing with the multiple change-point models the binary segmentation algorithm is described as a standard procedure in Eckley et al. (2011). Therefore we apply the max acc. $t2$, two-sided, to the segments $(1, \ldots, 47)$ and $(49, \ldots, 79)$ to obtain the $p$-value 0.77 and 0.94, respectively. Thus we conclude that $k = 48$ is the only one downturn point.

11.2. A directional goodness-of-fit test for a logit linear model

The data in Table 6 are taken from Allen et al. (1949). Erythroblastosis is a disease in certain newborn infants, and can sometimes be fatal. It is caused by transmission of anti-Rh antibody from an Rh− mother into the blood of an Rh+ baby. One form of treatment is an exchange transfusion, in which as much of the infant's blood as possible is replaced by a donor's blood that is free of anti-Rh antibody. Of a total of 179 cases in which this treatment was used in a Boston hospital, no infant deaths occurred in the 42 cases in which a female donor was used whereas 27 infant deaths in the 137 cases in which a male donor was used. Cochran (1954) carried out a detailed analysis of the apparent difference between the male and female donors, and proved the difference to be highly significant. Therefore a separate analysis is necessary for each of male and female. In this case, however, the interpretation for the female donor is obvious and a detailed analysis of association
between the deaths and the degree of disease is required only for male donor. If the interest is in the dependence of death rates on the degree of disease, a common procedure to deal with this type of data is to assume a logit linear model

\[ \log[p_i/(1 - p_i)] = \beta_0 + \beta_1 x_i \]  

(21)

for the binomial distribution \( \left( \frac{n_i}{y_i} \right) p_i^{y_i} (1 - p_i)^{n_i - y_i} \) since the linear model in \( p_i \) easily goes out of the range \([0, 1]\). However, the goodness-of-fit of the model (21) should be confirmed before applying. Then the convexity and/or the concavity test is useful as a directional goodness-of-fit test for the linearity of model (21). The problem can be dealt with by taking \( \mu_i = \log[p_i/(1 - p_i)], a(\mu_i) = n_i! [1 + \exp(\mu_i)]^{-n_i} \) and

\[
b(y_i) = \{y_i!(n_i - y_i)!\}^{-1} = \left[ \left( \frac{S_i - S_{i-1}}{x_i+1 - x_{i}} - \frac{S_{i-1} - S_{i-2}}{x_i - x_{i-1}} \right)! \left\{ n_i - \left( \frac{S_i - S_{i-1}}{x_i+1 - x_{i}} - \frac{S_{i-1} - S_{i-2}}{x_i - x_{i-1}} \right) \right\} \right]^{-1}.
\]

In this case the initial value of \( C(S_1, S_{1+1}) \) is

\[
C_1(S_1, S_2) = \left[ \left( \frac{S_i}{x_2 - x_1} \right)! \left\{ n_1 - \frac{S_i}{x_2 - x_1} \right\} \right]^{-1} \left[ \left( \frac{S_2 - S_1}{x_3 - x_2} - \frac{S_1}{x_2 - x_1} \right)! \left\{ n_2 - \left( \frac{S_2 - S_1}{x_3 - x_2} - \frac{S_1}{x_2 - x_1} \right) \right\} \right]^{-1}
\]

and the inequality (13) is changed to (B.5) in Appendix B.

The degree of disease is given by an ordinal qualitative measure in Table 6 and tentatively defined as \( x_i = i \) here. The two-sided p-value \( 3.8 \times 10^{-3} \) is obtained by the maximal contrast test suggesting the logit linear model will be inappropriate. The cumulative chi-squared is obtained as \( \chi^2 = 91.84 \) with \( d = 3.604 \) and \( f = 0.554 \). The p-value is \( 1.0 \times 10^{-5} \) by the chi-squared approximation and the same conclusion is obtained. Then as a non-parametric test the common goodness-of-fit \( \chi^2 \) is 29.25 with the degrees of freedom 3 and highly significant at \( p = 1.98 \times 10^{-6} \) but it cannot tell any relationship between the death rates and the degree of disease. Bartholomew (1963) suggested the use of \( \chi^2 \)-test as an appropriate statistic for the monotone alternative. However, an exact analysis by \( \chi^2 \) is possible only in a restricted number of cases. Hirotsu and Kuriki (1990) proposed max acc. \( t \) for the monotone hypothesis for the binomial data which is directly derived from the complete class lemma of Hirotsu (1982) and characterized by an exact and efficient algorithm for probability calculation for any configuration of \( n_i \) and \( y_i \). For Table 6 it gives one-sided p-value \( 1.2 \times 10^{-6} \) suggesting a step change-point between moderate and severe. Thus it is one of the merits of max acc. \( t \) that it can suggest an outline of the response curve without assuming any rigid parametric model.

11.3. Estimating the upper confidence bound at the lowest dose of an inverted-sigmoidally constrained response curve

The data in Table 7 are originally from an experiment performed by Dalbey and Lock (1982) and have been analysed by Schmoyer (1984) and Hirotsu and Srivastava (2000). Schmoyer gave a smoothed estimate of the dose–response curve under the sigmoid assumption and an upper confidence bound 0.055 of the risk at the lowest dose under an additional assumption that the response rate at zero dose is zero, which improves a naive estimate 0.095 (= 1 − 0.05^1/30) at significance level 0.05. Hirotsu and Srivastava obtained estimates 0.056 under the monotone assumption of the dose–response curve and 0.035 under the same assumption with Schmoyer. The proof of sigmoidicity has been, however, based on the normal approximation of the binomial distribution, which might be difficult to assume for the extremely low and high doses of the experiment. Therefore we apply the exact test of Sections 9 and 10 to the logit linear model to obtain the p-values 0.025 for inverted sigmoidicity, 0.814 for sigmoidicity and 0.051 for two-sided by max acc. \( t3 \) and 0.056 by the cumulative chi-squared \( \chi^2 \). This suggests an inverted sigmoidal depart from the response from the logit linear model. Then the logistic linear regression will give a higher response rate than the true response rate at the lowest dose giving a conservative (larger) estimate of the risk. Thus we have a conservative upper confidence bound 0.055 of the risk at the lowest dose at significance level 0.05 by a standard software for logistic regression analysis. It is interesting to see that several different methods suggest a rather similar upper bound. Then the present method should be appealing by an exact test for the sigmoidicity and the

<table>
<thead>
<tr>
<th>Degree of disease</th>
<th>Sex of donor</th>
<th>Number of deaths</th>
<th>Surv.</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>None</td>
<td>M</td>
<td>2</td>
<td>21</td>
<td>23</td>
</tr>
<tr>
<td></td>
<td>F</td>
<td>0</td>
<td>10</td>
<td>10</td>
</tr>
<tr>
<td>Mild</td>
<td>M</td>
<td>2</td>
<td>40</td>
<td>42</td>
</tr>
<tr>
<td></td>
<td>F</td>
<td>0</td>
<td>18</td>
<td>18</td>
</tr>
<tr>
<td>Moderate</td>
<td>M</td>
<td>6</td>
<td>33</td>
<td>39</td>
</tr>
<tr>
<td></td>
<td>F</td>
<td>0</td>
<td>10</td>
<td>10</td>
</tr>
<tr>
<td>Severe</td>
<td>M</td>
<td>17</td>
<td>16</td>
<td>33</td>
</tr>
<tr>
<td></td>
<td>F</td>
<td>0</td>
<td>4</td>
<td>4</td>
</tr>
</tbody>
</table>
Table 7
Results of diesel fuel aerosol experiment.

<table>
<thead>
<tr>
<th>Dose $x_i$</th>
<th>$n_i$</th>
<th>$y_i$</th>
<th>$y_i/n_i$</th>
</tr>
</thead>
<tbody>
<tr>
<td>8</td>
<td>30</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>16</td>
<td>40</td>
<td>1</td>
<td>0.025</td>
</tr>
<tr>
<td>24</td>
<td>40</td>
<td>2</td>
<td>0.05</td>
</tr>
<tr>
<td>28</td>
<td>10</td>
<td>5</td>
<td>0.5</td>
</tr>
<tr>
<td>32</td>
<td>30</td>
<td>12</td>
<td>0.4</td>
</tr>
<tr>
<td>48</td>
<td>20</td>
<td>16</td>
<td>0.8</td>
</tr>
<tr>
<td>64</td>
<td>10</td>
<td>6</td>
<td>0.6</td>
</tr>
<tr>
<td>72</td>
<td>10</td>
<td>10</td>
<td>1.0</td>
</tr>
</tbody>
</table>

maximum likelihood estimation of only two parameters utilizing the whole data without any additional assumption like zero response rate at zero dose. Other methods are utilizing only a part of data.

12. Discussion

While accumulated statistics are very popular in the statistical applications the use of doubly- and triply-accumulated statistics is rather novel and expected to have new applications. One of the important applications of the proposed method will be in the monitoring of time series of spontaneous reporting of the drug and adverse event combinations. In those situations, the max acc. $t$ method is useful for detecting an increasing tendency of reporting and the max acc. $t2$ should be useful for verifying that an increasing tendency changes to a decreasing tendency after some appropriate action. From the experience at PMDA one month is considered to be appropriate as a unit of time interval. In such situations, the number of points sometimes becomes as large as $K = 100$. Therefore a fast computer program has been developed for realizing the algorithm proposed in this paper, which is available on request.

The convexity and sigmoidicity tests are useful also as the directional goodness-of-fit tests for linearity of the assumed model in the dose–response analysis. If the linearity fails and convexity or sigmoidicity is suggested it provides useful information for choosing an appropriate directional test and also for improving a naive estimate of the response rate. For example, in the case of normal distribution, Hirotsu and Srivastava (2000) improved the estimate of the response rate for extremely low dose under the assumption of convexity, see also Example 11.3.

A particular power comparison has not been performed in this paper for the sigmoidicity hypothesis since it is time consuming and also there is no well known alternative method. However, since the angle of the convex polyhedral cone defined by the sigmoidicity hypothesis is narrower than that of the convex polyhedral cone defined by the monotone or convexity hypothesis the behaviour of max acc. $t3$ as a directional test is expected even better than the max acc. $t$ and the max acc. $t2$. The excellent behaviour of max acc. $t$ has been shown in Hirotsu et al. (1992, 2011), for example.

Although the application to the interaction effects was not mentioned here one of the merits of this approach is the ease of its extension to the analysis of two-way data. The interaction contrasts can be defined by the Kronecker product of two vectors representing the contrasts in the rows and columns. The ordinal effects along the rows and/or columns can be reflected by replacing the usual orthogonal contrasts with the change-point type contrasts corresponding to the ordered rows and/or columns. The application of the step change-point contrasts to the monotone effects along the rows and/or columns has been developed in Hirotsu (1982, 1983, 1997, 1998, 2009). The application of slope change-point contrasts to interaction effects has been obtained only for the normal model (Hirotsu et al., 2003). The theory and application of slope change-point and inflection-point contrasts to the contingency table are now investigated.

Acknowledgements

The research of the first author was supported by Grant-in-Aid for scientific research of the Japan Society for Promotion of Sciences (23500362). The authors thank Mr. Y. Nakamura of RPM for his assistance in preparing this article. Their thanks are also due to the two anonymous reviewers who have read the manuscript carefully and given very useful and constructive comments for revision.

Appendix A. Efficient score test for $M_k$, $k = 1, \ldots, K - 2$

This Appendix gives the proof that the maximal contrast test for the convexity hypothesis is simultaneously the efficient score test for the slope change-point hypothesis $M_k (2)$, $k = 1, \ldots, K - 2$. The model $M_k$ is rewritten as $\mu = \beta_0^* + \beta_k x_i$, $i = 1, \ldots, k + 1$; $\beta_0^* + \beta_k x_i$, $i = k + 2, \ldots, K$ with $\beta_0 + \beta_k x_{k+1} = \beta_0^* + \beta_k^* x_{k+1}$. Then by introducing a parameter $\tau_k = \beta_k^* - \beta_k$ the model can be rewritten further as

$$
\mu_i = \begin{cases}
\beta_0^* + \tau_k x_{k+1} + (\beta_k^* - \tau_k) x_i, & i = 1, \ldots, k + 1, \\
\beta_0^* + \beta_k^* x_i, & i = k + 2, \ldots, K.
\end{cases}
$$
Then assuming Poisson sequence an efficient score with respect to \( \tau_k \) is obtained as the derivative of the log likelihood function
\[
\frac{\partial \log L}{\partial \tau_k} = \sum_{i=1}^{k+1} (x_{k+1} - x_i)y_i + \text{const.,}
\]
where const. is free from the observation \( y_i \). Therefore the essential part of the efficient score evaluated at \( \tau_k = 0 \) is
\[
\sum_{i=1}^{k+1} (x_{k+1} - x_i)y_i = (x_{k+1} - x_1)y_1 + \cdots + (x_{k+1} - x_k)y_k, \quad k = 1, \ldots, K - 2,
\]
and coincides with the equations \( k = 1, \ldots, K - 2 \) of (5). The remaining part is a constant when considering the conditional distribution given the sufficient statistics \( S_{K-1} \) and \( Y_K \).

**Appendix B. Inequalities to define the range of \( S_k \)**

First there is an obvious inequality \( S_k \leq (x_{k+1} - x_1)Y_k = (x_{k+1} - x_1)(S_k - S_{k-1})/(x_{k+1} - x_k) \), which leads to
\[
(x_{k+1} - x_1)S_{k-1} \leq (x_k - x_1)S_k, \quad k = 2, \ldots, K - 1.
\]
It should be noted that the inequality (B.1) is stricter than an obvious inequality \( S_{k-1} \leq S_k \). Then from (B.1) Eq. (B.2) follows,
\[
\frac{S_1}{x_2 - x_1} \leq \cdots \leq \frac{S_k}{x_{k+1} - x_1} \leq \cdots \leq \frac{S_{K-1}}{x_K - x_1}.
\]
By virtue of a relation \((x_K - x_1)Y_k + S_{k-1} = (x_K - x_1)y_1 + \cdots + (x_K - x_k)y_k \) an inequality \((x_K - x_k)(S_k - S_{k-1})/(x_{k+1} - x_k) + S_{k-1} \leq S_{k-1} \) holds, namely
\[
(x_K - x_k)S_k \leq (x_k - x_{k+1})S_{k-1} + (x_{k+1} - x_k)S_{k-1}.
\]
Further to achieve \( S_{K-1} - (x_K - x_1)y_1 + \cdots + (x_K - x_k)y_k \) by \( y_k 's \) after k a relation
\[
(x_K - x_{k+1})(Y_K - Y_k) \geq S_{K-1} - (x_K - x_1)y_1 + \cdots + (x_K - x_k)y_k
\]
is required or equivalently
\[
S_k \geq S_{k-1} - (x_K - x_{k+1})Y_k.
\]
Then the absolute inequality (11) is obtained from (B.2) and (B.4), and gives restriction on \( S_k \) in terms of \( S_{K-1} \) and \( Y_K \) which are constants. The relative inequality (12) is obtained from (B.1) and (B.3), and gives the restriction on \( S_k \) by \( S_{k-1} \), which is useful for the bottom up procedure for constructing the conditional probabilities in Section 5. The inequality (13) follows immediately from \( Y_K \leq Y_{k+1} \). These inequalities are general for the exponential family (4) and sometimes additional inequalities are required. For example, for the independent binomial distribution \( B(n_i, p_i) \) in Sections 11.2 and 11.3 the inequality (13) is changed to
\[
0 \leq (S_{k+1} - S_k)/(x_{k+2} - x_{k+1}) - (S_k - S_{k-1})/(x_{k+1} - x_k) \leq n_k + 1.
\]