Nonparametric Statistics, 2002, Vol. 14(5), pp. 555–568

ON DETECTING CHANGE IN LIKELIHOOD RATIO ORDERING

C. XIONG[a,]* and HAMMOU EL BARMI[b]

[a]*Division of Biostatistics, Washington University in St. Louis, St. Louis, MO 63110, USA;*
[b]*Department of Statistics, Kansas State University, Manhattan, KS 66506, USA*

(Received October 2000; In final form December 2001)

This article studies the problem of testing and locating changepoints in likelihood ratios of two multinomial probability vectors. We propose a binary search procedure to detect the changepoints in the sequence of the ratios of probabilities and obtain the maximum likelihood estimators of two multinomial probability vectors under the assumption that the probability ratio sequence has a changepoint. We also give a strongly consistent estimator for the changepoint location. An information theoretic approach is used to test the equality of two discrete probability distributions against the alternative that their ratios have a changepoint. Approximate critical values of the test statistics are provided by simulation for several choices of model parameters. Finally, we examine a real life data set pertaining to average daily insulin dose from the Boston Collaborative Drug Surveillance Program and locate the changepoints in the probability ratios.

Keywords: Changepoints; Information criterion; Maximum likelihood estimate; Strong consistency

1 INTRODUCTION

Stochastic ordering of distributions is a very important concept in the theory and application of statistical inference. There are many different types of stochastic ordering in the literature. Shaked and Shanthikumar [24] provided a comprehensive source for the theory of stochastic ordering and its applications. Robertson *et al.* [20] gave a systematic and excellent treatment to the general order restricted statistical inference. One of the important types of stochastic ordering is the likelihood ratio ordering. If two distributions F and G possess density functions (or mass functions) f and g, respectively, then F dominates G in the sense of likelihood ratio ordering if $f(x)/g(x)$ is nondecreasing in x. This ordering is closely related to the usual stochastic ordering defined by Lehmann [18]. Keilson and Sumita [14] and Ross [21] have studied many of the properties of the likelihood ratio ordering and pointed out many of its applications in stochastic scheduling, closed queueing network, biology and reliability problems.

We restrict our attention to the case of two multinomial probability vectors. Let $\mathbf{p} = (p_1, p_2, \ldots, p_l)$ and $\mathbf{q} = (q_1, q_2, \ldots, q_l)$ be two multinomial probability distributions with

* Corresponding author.

ISSN 1048-5252 print; ISSN 1029-0311 online
DOI: 10.1080/1048525021000002154

$p_i > 0, q_i > 0, i = 1, 2, \ldots, l$. Dykstra *et al.* [9] studied the likelihood ratio test of the equality of the two multinomial probability vectors against the alternative that they are likelihood ratio ordered. They showed that the limiting distribution of the test statistic is of Chi-bar square type and provided the expression of the weighting values. When the null hypothesis of the equality of the two probability vectors is rejected, it is of interest to locate the changepoints in the ratios of the probabilities. To be specific, let H_0: $p_1/q_1 = p_2/q_2 = \cdots = p_l/q_l$ and H_a: $p_1/q_1 \le p_2/q_2 \le \cdots \le p_l/q_l$. Assume that H_0 is rejected, the problem we consider here is that of estimating the changepoints in the sequence $(p_1/q_1, p_2/q_2, \ldots, p_l/q_l)$. We provide strongly consistent estimators of the changepoints and use an information theoretic approach to test H_0 against the alternative that there are some changepoints in the ratios of probabilities.

The problem of detecting changepoints from the sequence $p_1/q_1 \le p_2/q_2 \le \cdots \le p_l/q_l$ is of interest in several aspects. First, these changepoints not only give the configuration of the probability ratios under likelihood ratio ordering, but also provide important information on the comparison of two probability vectors. For example, if k_1 is the first changepoint, then $p_1/q_1 = p_2/q_2 = \cdots = p_{k_1}/q_{k_1} < 1$, hence, $p_i < q_i$ for $i \le k_1$. Likewise, if k_s is the final changepoint, then $p_{k_s+1}/q_{k_s+1} = p_{k_s+2}/q_{k_s+2} = \cdots = p_l/q_l > 1$, *i.e.*, $p_i > q_i$ for $i > k_s$. Second, in a 2 by l two-way classification contingency table, if we assume that $\mathbf{p}$ and $\mathbf{q}$ are the conditional probability distribution of the column classification variable on row 1 and row 2, respectively, then $p_i/q_i \le p_{i+1}/q_{i+1}$ holds if and only if the local odds ratio $(p_i q_{i+1})/(q_i p_{i+1}) \le 1$. Therefore, the changepoints from the probability ratios provide the information where the local odds ratios are strictly less than 1. Third, in a competing-risks model studied by Dykstra *et al.* [10], the probability ratios correspond to the ratios between two cause-specific hazard rates. Thus, the changepoints from the probability ratios present information on the comparison between cause-specific hazard rates.

The classic change point problem is to estimate the time point where a sequence of random variables changes the distribution. The problem of changepoint has been studied by many authors. Chernoff and Zacks [8] proposed an estimator to estimate the current mean of a normal distribution which is subject to changes in time. Kander and Zacks [15] provided test procedures for possible changes in parameters of distributions. Scariano and Watkins [22] presented three classes of change-point estimators and gave their asymptotic properties. Pettitt [19] considered a nonnormalized version of Wilcoxon statistics for the detection of change points. Antoch and Hušková [3] proposed a class of change-point estimators and studied the limiting distributions of the estimators and their bootstrap versions. Zacks [29, 30] and Brodsky and Darkhovsky [4] surveyed the recent developments in nonparametric and Bayesian approaches to change-point detection. Chatterjee and Bandyopadhyay [6] provided a specific testing procedure for testing a possible change-point in time from a sequence of distributions. Sen [23] incorporated a time-dependent coefficients model in the formulation of change-point models in survival analysis and presented the relevant statistical methodology. Siegmund [25] gave a confidence set estimation of a change-point when observations are from an exponential family. Haccou *et al.* [13] studied the likelihood ratio test for a change-point for exponentially distributed random variables and derived the explicit asymptotic null-distribution of the test statistic. Gardner [11] and Chen and Gupta [7] proposed procedures to search for changepoints in terms of the location and variance parameters in a Gaussian model. Giraitis *et al.* [12] studied the hypothesis test of a change-point when the observations are correlated. Lavielle and Ludeña [17] used some maximal inequalities for quadratic forms and studied the detection of multiple change-points in the spectrum of a second-order stationary random process. Zacks [31, 32] developed a sequential testing procedure for testing reliability systems having a random number of change points in their hazard rate functions. Anruka [2] used an information criterion to detect the correct configuration of a sequence of order-constrained parameters from a sequence of distributions. Xiong and Milliken [28] proposed a testing procedure to detect the correct configuration of two stochastically ordered multinomial probability vectors.

Although the problem of changepoint has been studied by many authors, we are not aware of any estimates and tests in the literature specifically designed to detect the changepoints in the sequence of probability ratios. In this article we address this estimation and testing problem for two multinomial distributions.

More generally, we define the following hypotheses:

$$H_0\colon \frac{p_1}{q_1} = \frac{p_2}{q_2} = \cdots = \frac{p_l}{q_l},$$

$$H_1 - H_0\colon \frac{p_1}{q_1} = \cdots = \frac{p_{k_1}}{q_{k_1}} \neq \frac{p_{k_1+1}}{q_{k_1+1}} = \cdots = \frac{p_{k_s}}{q_{k_s}} \neq \frac{p_{k_s+1}}{q_{k_s+1}} = \cdots = \frac{p_l}{q_l},$$

$$H_2 - H_0\colon \frac{p_1}{q_1} = \cdots = \frac{p_{k_1}}{q_{k_1}} < \frac{p_{k_1+1}}{q_{k_1+1}} = \cdots = \frac{p_{k_s}}{q_{k_s}} < \frac{p_{k_s+1}}{q_{k_s+1}} = \cdots = \frac{p_l}{q_l},$$

where s is the unknown number of changepoints and $k_1 < k_2 < \cdots < k_s$ are unknown positions of the changepoints.

Vostrikova [27] proposed a method, known as binary segmentation procedure, to detect the number of changepoints in a multidimentional random process. We use the similar idea and propose the following similar steps to find the number of changepoints and locate the positions of changepoints in the sequence of likelihood ratios.

Step 1 Test $H_0\colon p_1/q_1 = p_2/q_2 = \cdots = p_l/q_l$ against $H_1 - H_0\colon p_1/q_1 = \cdots = p_{k_1}/q_{k_1} \neq p_{k_1+1}/q_{k_1+1} = \cdots = p_l/q_l$. Notice that k_1 is the only changepoint in the alternative hypothesis. If H_0 is accepted, then there is no changepoint. Otherwise, estimate k_1 and go to Step 2.

Step 2 From the two subsequences, $\{p_1/q_1, \ldots, p_{k_1}/q_{k_1}\}$ and $\{p_{k_1+1}/q_{k_1+1}, \ldots, p_l/q_l\}$, detect a changepoint by using the method in Step 1, respectively. (when $k_1 = 1$ or $l - 1$, only one subsequence should be considered). In doing so, one should condition on the appropriate sub-sample space of the two multinomial distributions. For example, when $l - 1 > k_1 > 1$, to detect a changepoint from the subsequence $\{p_1/q_1, \ldots, p_{k_1}/q_{k_1}\}$, one should test

$$H_0\colon \frac{p_1 q_{(k_1)}}{q_1 p_{(k_1)}} = \frac{p_2 q_{(k_1)}}{q_2 p_{(k_1)}} = \cdots = \frac{p_{k_1} q_{(k_1)}}{q_{k_1} p_{(k_1)}}$$

against

$$H_1 - H_0\colon \frac{p_1 q_{(k_1)}}{q_1 p_{(k_1)}} = \cdots = \frac{p_{k_2} q_{(k_1)}}{q_{k_2} p_{(k_1)}} \neq \frac{p_{k_2+1} q_{(k_1)}}{q_{k_2+1} p_{(k_1)}} \cdots = \frac{p_{k_1} q_{(k_1)}}{q_{k_1} p_{(k_1)}}$$

with unknown position $1 \leq k_2 < k_1$, where $p_{(k_1)} = \sum_{i=1}^{k_1} p_i$ and $q_{(k_1)} = \sum_{i=1}^{k_1} q_i$.

Step 3 The above process continues until there exists no more changepoint in any of the subsequences.

When the likelihood ratio ordering is assumed between $\mathbf{p} = (p_1, p_2, \ldots, P_l)$ and $\mathbf{q} = (q_1, q_2, \ldots, q_l)$, to detect the changepoints, one should then test $H_0\colon p_1/q_1 = p_2/q_2 = \cdots = p_l/q_l$

against $H_2 - H_0$: $p_1/q_1 = \cdots = p_{k_l}/q_{k_l} < p_{k_1+1}/q_{k_1+1} = \cdots = p_{k_2}/q_{k_2} \cdots < p_{k_s+1}/q_{k_s+1} = \cdots = p_l/q_l$. Similar three steps as above can be used to find the number of changepoints and to locate these changepoints.

When data are used to perform the proposed sequential process, the overall significance level for the entire process is always less than the minimum of the significance levels from all the test steps used in the sequential process. To be more specific, let s $(1 \leq s \leq l-1)$ be the (random) number of changepoints detected by the sequential process. Assume that the rejection region for testing H_0 against H_i, $i = 1, 2$, from the tth step in the sequential process is R^t, $1 \leq t \leq l-1$. The whole rejection region for detecting all s changepoints in the entire process is $R_s = (\cap_{t=1}^{s} R^t) \cap (R^{s+1})'$. Therefore, the overall significance level is

$$P(R_s|H_0) \leq \min_{1 \leq t \leq s} P(R^t|H_0).$$

The above three-step process reduces the problem of detecting many changepoints into the problem of detecting a single changepoint in the probability ratio sequence. Therefore, in this article, we should be focusing on testing the following hypothesis: (we still use the same notations H_0, H_1 and H_2)

$$H_0\colon \frac{p_1}{q_1} = \frac{p_2}{q_2} = \cdots = \frac{p_l}{q_l},$$

$$H_1 - H_0\colon \frac{p_1}{q_1} = \cdots = \frac{p_{k_0}}{q_{k_0}} \neq \frac{p_{k_0+1}}{q_{k_0+1}} = \cdots = \frac{p_l}{q_l},$$

$$H_2 - H_0\colon \frac{p_1}{q_1} = \cdots = \frac{p_{k_0}}{q_{k_0}} < \frac{p_{k_0+1}}{q_{k_0+1}} = \cdots = \frac{p_l}{q_l}.$$

where k_0 is the unknown position of the changepoint.

Our approach is based on the information criterion introduced by Akaike [1], which is also closely related to the likelihood principle. In fact, our approach is also similar to that of Anruka [2] and Xiong and Milliken [28]. But there are differences between our approach and that of Anruka [2] and Xiong and Milliken [28]. While Anruka [2] searched for a different bias correction term of information criterion which incorporates the simple order constraint among model parameters of the exponential family, this article looks for the asymptotically unbiased Akaike information criterion up to a certain order. On the other hand, although Xiong and Milliken [28] considered similar asymptotically unbiased Akaike information criterion, this paper deals with the constraint of likelihood ratio ordering which is much stronger than the usual stochastic ordering that they considered. We point out that other approaches such as Schwartz's Bayesian Information Criterion and Hannan/Quinn Information Criterion can also be used to tackle our problem and will pursue the research in that direction in our future work. In Section 2, we derive the maximum likelihood estimates for $\mathbf{p}$ and $\mathbf{q}$ and the asymptotically unbiased information criterion up to the order $(m+n)^{-1}$ under $H_0, H_1 - H_0$ and $H_2 - H_0$ when k_0 is assumed known. In Section 3 we give an estimate for k_0 under $H_i - H_0$ $(i = 1, 2)$ and prove the strong consistency of the estimator. We also test H_0 against $H_i - H_0$, $(i = 1, 2)$, in Section 3 and give approximate critical values for the test statistics under H_0 through a simulation study for selected model parameters. In Section 4, we demonstrate our testing and estimation procedures using a real data set originally studied by Dykstra *et al.* [9] pertaining to average daily insulin dose for patients with and without hypoglycemia.

2 MLE AND ASYMPTOTICALLY UNBIASED INFORMATION CRITERION WITH KNOWN k_0

2.1 Maximum Likelihood Estimates

Throughout Section 2.1 we assume that k_0 is known. We assume that a random sample of size m and a random sample of size n are independently taken from multinomial distributions with probability vectors $\mathbf{p} = (p_1, p_2, \ldots, p_l)$ and $\mathbf{q} = (q_1, q_2, \ldots, q_l)$, respectively. Let $(m_1, m_2, \ldots, m_l)$ and $(n_1, n_2, \ldots, n_l)$ be the corresponding observed frequencies. The likelihood as a function of $\mathbf{p}$ and $\mathbf{q}$ is

$$L(\mathbf{p}, \mathbf{q}) \propto \prod_{i=1}^{l} p_i^{m_i} q_i^{n_i}.$$

The MLEs of $\mathbf{p}$ and $\mathbf{q}$ under H_0 are given by $\hat{\mathbf{p}}^{(0)}$ and $\hat{\mathbf{q}}^{(0)}$ when $\widehat{p_i}^{(0)} = \widehat{q_i}^{(0)} = (m_i + n_i)/(m+n),\ i = 1, 2, \ldots, l$. To find the MLEs under H_1, we introduce the same parametrization as in Dykstra *et al.* [9]. Let $\theta_i = mp_i/(mp_i + nq_i)$ and $\phi_i = mp_i + nq_i$ for $i = 1, 2, \ldots l$, so that $p_i = \theta_i\phi_i/m$ and $q_i = \phi_i(1-\theta_i)/n$. The restrictions on the new parameters are $0 \le \theta_i \le 1$, $\phi_i \ge 0$, $\sum_{i=1}^{l} \phi_i = m + n$ and $\sum_{i=1}^{l} \theta_i\phi_i = m$. In terms of the new parameters, the likelihood function $L(\mathbf{p}, \mathbf{q})$ is proportional to

$$\theta_1^{m_{(k_0)}}(1-\theta_1)^{n_{(k_0)}}\theta_l^{m-m_{(k_0)}}(1-\theta_l)^{n-n_{(k_0)}} \prod_{i=1}^{l} \phi_i^{m_i+n_i}$$

where $m_{(k_0)} = \sum_{i=1}^{k_0} m_i$ and $n_{(k_0)} = \sum_{i=1}^{k_0} n_i$.

Note that the likelihood function factors into two parts, the first part involving only θ_1 and θ_l and the other only ϕ_i's. The two parts can be maximized independently and we get

$$\begin{aligned}
\widehat{\phi_i}^{(1)} &= m_i + n_i,\\
\widehat{\theta_1}^{(1)} &= \frac{m_{(k_0)}}{m_{(k_0)} + n_{(k_0)}},\\
\widehat{\theta_l}^{(1)} &= \frac{m - m_{(k_0)}}{m - m_{(k_0)} + n - n_{(k_0)}}.
\end{aligned}$$

Thus, the MLEs of $\mathbf{p}$ and $\mathbf{q}$ under H_1 are given by $\hat{\mathbf{p}}^{(1)}$ and $\hat{\mathbf{q}}^{(1)}$ when

$$\begin{aligned}
\widehat{p_i}^{(1)} &= \frac{(m_i+n_i)m_{(k_0)}}{m(m_{(k_0)} + n_{(k_0)})}, \quad \text{for } i \le k_0,\\
\widehat{p_i}^{(1)} &= \frac{(m_i+n_i)(m - m_{(k_0)})}{m(m - m_{(k_0)} + n - n_{(k_0)})}, \quad \text{for } i > k_0,\\
\widehat{q_i}^{(1)} &= \frac{(m_i+n_i)n_{(k_0)}}{n(m_{(k_0)} + n_{(k_0)})}, \quad \text{for } i \le k_0,\\
\widehat{q_i}^{(1)} &= \frac{(m_i+n_i)(n - n_{(k_0)})}{n(m + n - m_{(k_0)} - n_{(k_0)})}, \quad \text{for } i > k_0.
\end{aligned}$$

Notice that H_2 is equivalent to $\theta_1 \le \theta_l$. In order to find the MLEs under H_2 with known k_0, we need to maximize $\theta_1^{m_{(k_0)}}(1-\theta_1)^{n_{(k_0)}}\theta_l^{m-m_{(k_0)}}(1-\theta_l)^{n-n_0}$ in $L(\mathbf{p}, \mathbf{q})$ subject to $\theta_1 \le \theta_l$.

This is the bioassay problem discussed by Robertson *et al.* [20] (page 32), and the solution $(\widehat{\theta_1}^{(2)}, \widehat{\theta_l}^{(2)})$ is given by

$$(\widehat{\theta_1}^{(2)}, \widehat{\theta_l}^{(2)}) = \begin{cases} \left(\dfrac{m_{(k_0)}}{m_{(k_0)} + n_{(k_0)}}, \dfrac{m - m_{(k_0)}}{m + n - m_{(k_0)} - n_{(k_0)}}\right), & \text{if } \dfrac{m_{(k_0)}}{m} \le \dfrac{n_{(k_0)}}{n} \\ \left(\dfrac{m}{m+n}, \dfrac{m}{m+n}\right), & \text{otherwise} \end{cases}$$

Thus, the MLEs of $\mathbf{p}$ and $\mathbf{q}$ under H_2 are given by $\hat{\mathbf{p}}^{(2)}$ and $\hat{\mathbf{q}}^{(2)}$ when

$$(\widehat{p_i}^{(2)}, \widehat{q_i}^{(2)}) = \begin{cases} (\widehat{p_i}^{(1)}, \widehat{q_i}^{(1)}), & \text{if } \dfrac{m_{(k_0)}}{m} \le \dfrac{n_{(k_0)}}{n} \\ (\widehat{p_i}^{(0)}, \widehat{q_i}^{(0)}), & \text{otherwise} \end{cases}$$

for $i = 1, 2, \ldots, l$.

2.2 Asymptotically Unbiased Information Criterion

Akaike [1] proposed the Akaike information criterion (AIC) to select and differentiate statistical models. When two competing model M_0 and M_1 are available, the decision to accept M_0 or M_1 is made based on the principle of minimum information criterion. In his derivation for AIC, Akaike [1] used $\log L(\hat{\theta})$ as an estimate of $J = E_{\hat{\theta}} E_Y(\log f(Y|\hat{\theta}))$, where $f(Y|\theta_0)$ is the probability density or mass function of a future sample Y, $\hat{\theta}$ is the MLE of the model parameter θ_0 (may be a vector) from a random sample X with likelihood function $L(\theta)$, Y is of the same size and distribution as X, and X and Y are independent. The expectation $E_{\hat{\theta}}$ is taken under the distribution of X. In fact, Akaike information criterion (AIC) is closely related with the Kullback–Leibler's discrimination information [16] defined as

$$\begin{aligned} D &= E\left(\log \frac{f(Y|\theta_0)}{f(Y|\hat{\theta})}\right) \\ &= E(\log f(Y|\theta_0)) - J. \end{aligned}$$

The Kullback–Leibler's information D has the property that it is always nonnegative and equals to 0 if and only if $\hat{\theta} = \theta_0$ almost surely. Therefore, the best estimator for θ_0 based on the Kullback–Leibler's information is the one which minimizes D. Since the first term in D is an unknown constant, the minimum of D means the minimum of $-J$. Since $-2J$ has to be estimated and $-2 \log L(\hat{\theta})$ is not an unbiased estimator of $-2J$, Akaike used a correction term for the asymptotic bias and defined the Akaike information criterion as

$$\mathrm{AIC} = -2 \log L(\hat{\theta}) + 2\gamma,$$

where γ is the number of free parameters (or the dimension) in the model. Since AIC is still not an unbiased estimator of $-2J$, Sugiura [26] proposed unbiased and asymptotically unbiased estimators of $-2J$ up to a certain order and provided finite corrections of AIC for several important model selection problems. We now give the asymptotically unbiased version of AIC up to the order $(m+n)^{-1}$ under H_0, $H_1 - H_0$ and $H_2 - H_0$ when k_0 is known. Let $(m'_1, m'_2, \ldots, m'_l)$ and $(n'_1, n'_2, \ldots, n'_l)$ be the frequencies of two future samples from $\mathbf{p}$ and $\mathbf{q}$, respectively. Assume that $(m'_1, m'_2, \ldots, m'_l)$ and $(n'_1, n'_2, \ldots, n'_l)$ are independent of $\mathbf{m} = (m_1, m_2, \ldots, m_l)$ and $\mathbf{n} = (n_1, n_2, \ldots, n_l)$. Suppose that $\sum_{i=1}^l m_i = \sum_{i=1}^l m_i' = m$, $\sum_{i=1}^l n_i = \sum_{i=1}^l n_i' = n$, $\lim_{m,n\to\infty} n/m > 0$, and let $C(\mathbf{m}, \mathbf{n}) = m!n!/(m_1!m_2!\cdots m_l!n_1!n_2!\cdots n_l!)$. Sugiura [26] pointed out that if X has a binomial distribution $b(k, p)$, then

$$E\left[\frac{X-kp}{\sqrt{kp(1-p)}}\right]^2 = 1,$$

$$E\left[\frac{X-kp}{\sqrt{kp(1-p)}}\right]^3 = \frac{1-2p}{\sqrt{kp(1-p)}},$$

$$E\left[\frac{X-kp}{\sqrt{kp(1-p)}}\right]^4 = 3 - \frac{6}{k} + \frac{1}{kp(1-p)}.$$

Thus, under H_0,

$$\begin{aligned}
J &= E_{(\hat{\mathbf{p}}^{(0)},\hat{\mathbf{q}}^{(0)})}E_{(\mathbf{m}',\mathbf{n}')}(\log C(\mathbf{m}',\mathbf{n}') + \sum_{i=1}^l (m_i' + n_i')\log \widehat{q_i}^{(0)}) \\
&= E_{(\hat{\mathbf{p}}^{(0)},\hat{\mathbf{q}}^{(0)})}[\log L(\hat{\mathbf{p}}^{(0)}, \hat{\mathbf{q}}^{(0)}) + (m+n)\sum_{i=1}^l (q_i - \widehat{q_i}^{(0)})\log \widehat{q_i}^{(0)}] \\
&= E_{(\hat{\mathbf{p}}^{(0)},\hat{\mathbf{q}}^{(0)})}\Bigg\{\log L(\hat{\mathbf{p}}^{(0)}, \hat{\mathbf{q}}^{(0)}) - (m+n) \\
&\qquad \times \left[\sum_{i=1}^l \frac{(\widehat{q_i}^{(0)} - q_i)^2}{q_i} - \sum_{i=1}^l \frac{(\widehat{q_i}^{(0)} - q_i)^3}{2q_i^2} + \sum_{i=1}^l \frac{(\widehat{q_i}^{(0)} - q_i)^4}{3q_i^3}\right]\Bigg\} + O(m+n)^{-3/2} \\
&= E_{(\hat{\mathbf{p}}^{(0)},\hat{\mathbf{q}}^{(0)})}\log L(\hat{\mathbf{p}}^{(0)}, \hat{\mathbf{q}}^{(0)}) - (l-1) - \sum_{i=1}^l \frac{1-q_i}{2q_i(m+n)} + O(m+n)^{-3/2}.
\end{aligned}$$

Therefore, the asymptotically corrected AIC up to the order $(m+n)^{-1}$ under H_0 is

$$\mathrm{AIC}_0 \equiv -2\log L(\hat{\mathbf{p}}^{(0)}, \hat{\mathbf{q}}^{(0)}) + 2(l-1) + \sum_{i=1}^l \frac{1-q_i}{q_i(m+n)}.$$

Note that $\mathrm{AIC}_0 = \mathrm{AIC} + \sum_{i=1}^l (1-q_i)/[q_i(m+n)]$, where AIC is the classic Akaike information criterion. The estimated asymptotically unbiased AIC_0 up to the order $(m+n)^{-1}$ can be obtained after replacing q_i by its MLE $q_i^{(0)}$.

Next, we find the asymptotically unbiased version of AIC under $H_1 - H_0$ with known k_0. Again let $(m_1', m_2', \ldots, m_l')$ and $(n_1', n_2', \ldots, n_l')$ be the frequencies of two future samples from

$\mathbf{p}$ and $\mathbf{q}$ independent of $(m_1, m_2, \ldots, m_l)$ and $(n_1, n_2, \ldots, n_l)$, respectively. Suppose that $\sum m_i = \sum m_i' = m$ and $\sum n_i = \sum n_i' = n$. Under $H_1 - H_0$,

$$
\begin{aligned}
J &= E_{(\hat{\mathbf{p}}^{(1)},\hat{\mathbf{q}}^{(1)})}E_{(\mathbf{m}',\mathbf{n}')}(\log C(\mathbf{m}',\mathbf{n}') + \sum_{i=1}^{l}(m_i'\log \widehat{p_i}^{(1)} + n_i'\log \widehat{q_i}^{(1)})) \\
&= E_{(\hat{\mathbf{p}}^{(1)},\hat{\mathbf{q}}^{(1)})}\Bigg\{\log L(\hat{\mathbf{p}}^{(1)},\hat{\mathbf{q}}^{(1)}) + \sum_{i=1}^{k_0}(mp_i - m_i)\log\left[\frac{m_{(k_0)}n}{mn_{(k_0)}}\right] \\
&\quad + \sum_{i=k_0+1}^{l}(mp_i - m_i)\log\left[\frac{m - m_{(k_0)}}{n - n_{(k_0)}}\right] \\
&\quad + \sum_{i=1}^{k_0}(mp_i + nq_i - m_i - n_i)\log\left[\frac{n_{(k_0)}(m_i + n_i)}{m_{(k_0)} + n_{(k_0)}}\right] \\
&\quad + \sum_{i=k_0+1}^{l}(mp_i + nq_i - m_i - n_i)\log\left[(n - n_{(k_0)})(m_i + n_i)\right] \\
&\quad - \sum_{i=k_0+1}^{l}(mp_i + nq_i - m_i - n_i)\log(m + n - m_{(k_0)} - n_{(k_0)})\Bigg\} \\
&= E_{(\hat{\mathbf{p}}^{(1)},\hat{\mathbf{q}}^{(1)})}\log L(\hat{\mathbf{p}}^{(1)},\hat{\mathbf{q}}^{(1)}) - 2 - \sum_{i=1}^{l}\frac{mp_i(1-p_i) + nq_i(1-q_i)}{mp_i + nq_i} \\
&\quad + \frac{(m+n)mp_{(k_0)}(1-p_{(k_0)})}{(mp_{(k_0)} + nq_{(k_0)})(m + n - mp_{(k_0)} - nq_{(k_0)})} \\
&\quad + \frac{(m+n)nq_{(k_0)}(1-q_{(k_0)})}{(mp_{(k_0)} + nq_{(k_0)})(m + n - mp_{(k_0)} - nq_{(k_0)})} \\
&\quad + \frac{2q_{(k_0)} - 1}{2nq_{(k_0)}(1-q_{(k_0)})} + \frac{2p_{(k_0)} - 1}{2mp_{(k_0)}(1-p_{(k_0)})} \\
&\quad - [mp_{(k_0)}(1-p_{(k_0)})(1-2p_{(k_0)}) + nq_{(k_0)}(1-q_{(k_0)})(1-2q_{(k_0)})] \\
&\quad \times \left\{\frac{1}{2(m + n - mp_{(k_0)} - nq_{(k_0)})^2} + \frac{1}{2(mp_{(k_0)} + nq_{(k_0)})^2}\right\} \\
&\quad + \sum_{i=1}^{l}\frac{mp_i(1-p_i)(1-2p_i) + nq_i(1-q_i)(1-2q_i)}{2(mp_i + nq_i)^2} \\
&\quad + \frac{[mp_{(k_0)}(1-p_{(k_0)}) + nq_{(k_0)}(1-q_{(k_0)})]^2}{(mp_{(k_0)} + nq_{(k_0)})^3} \\
&\quad + \frac{[mp_{(k_0)}(1-p_{(k_0)}) + nq_{(k_0)}(1-q_{(k_0)})]^2}{(m + n - mp_{(k_0)} - nq_{(k_0)})^3} \\
&\quad - \sum_{i=1}^{l}\frac{[mp_i(1-p_i) + nq_i(1-q_i)]^2}{(mp_i + nq_i)^3} + O(m+n)^{-3/2}.
\end{aligned}
$$

Thus, the asymptotically corrected AIC under $H_1 - H_0$ up to the order $(m+n)^{-1}$ is

$$
\begin{aligned}
\mathrm{AIC}_1(k_0) = \mathrm{AIC}(k_0) - 2\Bigg\{&\frac{2q_{(k_0)} - 1}{2nq_{(k_0)}(1-q_{(k_0)})} + \frac{2p_{(k_0)} - 1}{2mp_{(k_0)}(1-p_{(k_0)})} \\
&- [mp_{(k_0)}(1-p_{(k_0)})(1-2p_{(k_0)}) + nq_{(k_0)}(1-q_{(k_0)})(1-2q_{(k_0)})]
\end{aligned}
$$

$$
\begin{aligned}
&\times \left\{ \frac{1}{2(m+n-mp_{(k_0)}-nq_{(k_0)})^2} + \frac{1}{2(mp_{(k_0)}+nq_{(k_0)})^2} \right\} \\
&+ \sum_{i=1}^{l} \frac{mp_i(1-p_i)(1-2p_i)+nq_i(1-q_i)(1-2q_i)}{2(mp_i+nq_i)^2} \\
&+ \frac{[mp_{(k_0)}(1-p_{(k_0)})+nq_{(k_0)}(1-q_{(k_0)})]^2}{(mp_{(k_0)}+nq_{(k_0)})^3} + \frac{[mp_{(k_0)}(1-p_{(k_0)})+nq_{(k_0)}(1-q_{(k_0)})]^2}{(m+n-mp_{(k_0)}-nq_{(k_0)})^3} \\
&- \sum_{i=1}^{l} \frac{[mp_i(1-p_i)+nq_i(1-q_i)]^2}{(mp_i+nq_i)^3} \Bigg\},
\end{aligned}
$$

where $\mathrm{AIC}(k_0) = -2\log L(\hat{\mathbf{p}}^{(1)}, \hat{\mathbf{q}}^{(1)}) + 2l$ is the classic Akaike information criterion with changepoint k_0. Notice that the MLEs $(\hat{\mathbf{p}}^{(2)}, \hat{\mathbf{q}}^{(2)})$ of $\mathbf{p}$ and $\mathbf{q}$ under $H_2 - H_0$ are exactly the same as the MLEs $(\hat{\mathbf{p}}^{(1)}, \hat{\mathbf{q}}^{(1)})$ under $H_1 - H_0$ when $m_{(k_0)}/m \le n_{(k_0)}/n$. Since $\sum_{i=1}^{k_0} p_i < \sum_{i=1}^{k_0} q_i$ under $H_2 - H_0$, the strong law of large numbers assures that $m_{(k_0)}/m \le n_{(k_0)}/n$ almost surely when m and n are large. Thus, the asymptotically corrected AIC under $H_2 - H_0$, denoted $\mathrm{AIC}_2(k_0)$, has the same form as $\mathrm{AIC}_1(k_0)$, except that $-2\log L(\hat{\mathbf{p}}^{(1)}, \hat{\mathbf{q}}^{(1)})$ should be changed to $-2\log L(\hat{\mathbf{p}}^{(2)}, \hat{\mathbf{q}}^{(2)})$ in the formula of $\mathrm{AIC}_1(k_0)$. Again, in practice, $(\hat{\mathbf{p}}^{(i)}, \hat{\mathbf{q}}^{(i)})$ should be used for the unknown parameters $\mathbf{p}$ and $\mathbf{q}$ in the computation of $\mathrm{AIC}_i(k_0),\ i = 1, 2$.

3 INFERENCES

When the changepoint k_0 is known, testing H_0 *vs.* $H_1 - H_0$ (or $H_2 - H_0$) is a relatively easy task as the asymptotic distribution of likelihood ratio test statistics is well known. We state these results in the following theorem.

THEOREM 1 *When H_0 is true, for any $t > 0$,*

$$\lim_{m,n\to\infty} P(2\log L(\hat{\mathbf{p}}^{(1)}, \hat{\mathbf{q}}^{(1)}) - 2\log L(\hat{\mathbf{p}}^{(0)}, \hat{\mathbf{q}}^{(0)}) \ge t) = P(\chi^2(1) \ge t) \tag{1}$$

$$\lim_{m,n\to\infty} P(2\log L(\hat{\mathbf{p}}^{(2)}, \hat{\mathbf{q}}^{(2)}) - 2\log L(\hat{\mathbf{p}}^{(0)}, \hat{\mathbf{q}}^{(0)}) \ge t) = \frac{1}{2}P(\chi^2(1) \ge t) \tag{2}$$

Proof Equation (1) is from the classical asymptotic theory. To prove (2), we realize that there is only one inequality constraint in the hypothesis H_2 and apply Theorem 3.1 from Dykstra *et al.* [9] to obtain (2).

When the changepoint k_0 in $H_1 - H_0$ (or $H_2 - H_0$) is unknown, we use the information criterion to test H_0 against $H_1 - H_0$ (or $H_2 - H_0$) and to estimate k_0. Let AIC_0 be the information criterion under H_0, and let $\mathrm{AIC}_1(k)$ and $\mathrm{AIC}_2(k)$ be the information criterion under $H_1 - H_0$ and $H_2 - H_0$ when the changepoint is k, respectively. According to the principle of minimum information criterion, we should fail to reject H_0 if $\mathrm{AIC}_0 \le \min_{1\le k\le l-1} \mathrm{AIC}_i(k),\ i = 1, 2$. Hence, H_0 should be rejected when there is at least one k such that $\mathrm{AIC}_i(k) < \mathrm{AIC}_0(k),\ i = 1, 2$. When H_0 is rejected, we appeal to the principle of information criterion again and estimate the real position of the changepoint k_0 in $H_i - H_0$, $i = 1, 2$, by $\widehat{k^i}$ such that

$$\mathrm{AIC}_i(\widehat{k^i}) = \min_{1\le k\le l-1} \mathrm{AIC}_i(k). \tag{3}$$

The following theorem gives the consistency of $\widehat{k^i}$.

THEOREM 2 *Suppose that $H_1 - H_0$ (or $H_2 - H_0$) is the correct model with the correct changepoint at position k_0, and $\lim_{m,n\to\infty} n/(m+n) = \lambda$ with $0 < \lambda < 1$. If $\widehat{k^1}$ ($\widehat{k^2}$) is defined by (3), then $\widehat{k^1}$ (or $\widehat{k^2}$) is a strongly consistent estimator of k_0.*

Proof We prove the consistency of $\widehat{k^1}$. The proof for the consistency of $\widehat{k^2}$ is very similar. Let $\pi_1 = p_1/q_1 = \cdots = p_{k_0}/q_{k_0}$ and $\pi_2 = p_{k_0+1}/q_{k_0+1} = \cdots = p_l/q_l$. We proceed by considering two different situations. When $k > k_0$, as $m, n \to \infty$,

$$\begin{aligned}
&\frac{[\mathrm{AIC}_1(k_0) - \mathrm{AIC}_1(k)]}{(2m)} \\
&= \left(\frac{1}{m}\right)\Bigg\{ \sum_{i=1}^{k} m_i \log\left(\frac{m_{(k)}}{n_{(k)}}\right) - \sum_{i=1}^{k_0} m_i \log\left(\frac{m_{(k_0)}}{n_{(k_0)}}\right) \\
&\quad + \sum_{i=k+1}^{l} m_i \log\left[\frac{m - m_{(k)}}{n - n_{(k)}}\right] - \sum_{i=k_0+1}^{l} m_i \log\left[\frac{m - m_{(k_0)}}{n - n_{(k_0)}}\right] \\
&\quad + \sum_{i=1}^{k} (m_i + n_i) \log\left[\frac{n_{(k)}}{m_{(k)} + n_{(k)}}\right] - \sum_{i=1}^{k_0} (m_i + n_i) \log\left[\frac{n_{(k_0)}}{m_{(k_0)} + n_{(k_0)}}\right] \\
&\quad + \sum_{i=k+1}^{l} (m_i + n_i) \log\left[\frac{n - n_{(k)}}{m + n - m_{(k)} - n_{(k)}}\right] \\
&\quad - \sum_{i=k_0+1}^{l} (m_i + n_i) \log\left[\frac{n - n_{(k_0)}}{m + n - m_{(k_0)} - n_{(k_0)}}\right]\Bigg\} + O\left(\frac{1}{m}\right) \\
&\overset{a.s.}{\to} p_{(k)}\left\{ -\log\left(\frac{q_{(k)}}{p_{(k)}}\right) + \left[1 + \frac{(\lambda/(1-\lambda))q_{(k)}}{p_{(k)}}\right] \log\left[\frac{q_{(k)}}{(1-\lambda)p_{(k)} + \lambda q_{(k)}}\right]\right\} \\
&\quad - p_{(k_0)} \log \pi_1 - (p_{(k)} - p_{(k_0)}) \log \pi_2 + \left[p_{(k_0)} + \frac{\lambda q_{(k_0)}}{1-\lambda}\right] \log[(1-\lambda)\pi_1 + \lambda] \\
&\quad + \left[1 + \frac{\lambda}{(1-\lambda)\pi_2}\right](p_{(k)} - p_{(k_0)}) \log[(1-\lambda)\pi_2 + \lambda].
\end{aligned}$$

Since

$$-\log\left(\frac{q_{(k)}}{p_{(k)}}\right) + \left[\frac{1 + (\lambda/(1-\lambda))q_{(k)}}{p_{(k)}}\right] \log\left[\frac{q_{(k)}}{(1-\lambda)p_{(k)} + \lambda q_{(k)}}\right]$$

is a convex function in $q_{(k)}/p_{(k)}$ and

$$\frac{q_{(k)}}{p_{(k)}} = \left(\frac{p_{(k_0)}}{p_{(k)}}\right)\left(\frac{1}{\pi_1}\right) + \left(1 - \frac{p_{(k_0)}}{p_{(k)}}\right)\left(\frac{1}{\pi_2}\right),$$

it follows that

$$p_{(k)}\left\{-\log\left(\frac{q_{(k)}}{p_{(k)}}\right)+\left[1+\frac{(\lambda/(1-\lambda))q_{(k)}}{p_{(k)}}\right]\log\left[\frac{q_{(k)}}{(1-\lambda)p_{(k)}+\lambda q_{(k)}}\right]\right\}$$
$$< p_{(k_0)}\log(\pi_1)+(p_{(k)}-p_{(k_0)})\log\pi_2-\left[p_{(k_0)}+\frac{\lambda q_{(k_0)}}{1-\lambda}\right]\log[(1-\lambda)\pi_1+\lambda]$$
$$-\left[1+\frac{\lambda}{(1-\lambda)\pi_2}\right][p_{(k)}-p_{(k_0)}]\log[(1-\lambda)\pi_2+\lambda].$$

Thus, when $k > k_0$, $\mathrm{AIC}_1(k)-\mathrm{AIC}_1(k_0) > 0$ almost surely when m and n are sufficiently large. So, almost surely, $\limsup \widehat{k^1} \le k_0$. Similarly, when $k < k_0$,

$$\frac{[\mathrm{AIC}_1(k_0)-\mathrm{AIC}_1(k)]}{2m}\xrightarrow{a.s.}(1-p_{(k)})\left\{-\log\left(\frac{1-q_{(k)}}{1-p_{(k)}}\right)+\left[1+\frac{(\lambda/(1-\lambda))(1-q_{(k)})}{1-p_{(k)}}\right]\right.$$
$$\left.\times\log\left[\frac{1-q_{(k)}}{(1-\lambda)(1-p_{(k)})+\lambda(1-q_{(k)})}\right]\right\}-(1-p_{(k_0)})\log\pi_2$$
$$+\left[1-p_{(k_0)}+\frac{(1-q_{(k_0)})\lambda}{1-\lambda}\right]\log[(1-\lambda)\pi_2+\lambda]+(p_{(k)}-p_{(k_0)})\log\pi_1$$
$$+\left[1+\frac{\lambda}{(1-\lambda)\pi_1}\right](p_{(k_0)}-p_{(k)})\log[(1-\lambda)\pi_1+\lambda].$$

Using the fact that $-\log((1-q_{(k)})/(1-p_{(k)}))$ $+[1+(\lambda/(1-\lambda))(1-q_{(k)})/(1-p_{(k)})]\times\log[(1-q_{(k)})/[(1-\lambda)(1-p_{(k)})+\lambda(1-q_{(k)})]]$ is a convex function of $(1-q_{(k)})/(1-p_{(k)})$ and $(1-q_{(k)})/(1-p_{(k)})=[(1-p_{(k_0)})/(1-p_{(k)})](1/\pi_2)+$ $[1-$ $(1-p_{(k_0)})/(1-p_{(k)})](1/\pi_1)$, we conclude that $\mathrm{AIC}_1(k)-$ $\mathrm{AIC}_1(k_0) > 0$ when m and n are sufficiently large, which then implies that almost surely $\liminf \widehat{k^1} \ge k_0$. This proves the consistency of $\widehat{k^1}$.

Although the principle of information criterion gives us very simple and easy to use decision rule for testing H_0 against $H_i - H_0$, $i = 1, 2$, it does not give the significance level for the test. When the AICs from H_0 and H_i do not have much difference, a decision based on the above simple decision rule may be very risky since the slight difference between the AICs from H_0 and H_i might be caused due to the random noise from the data. In order to control the significance level of the test, we modify our decision rule for testing H_0 against $H_i - H_0$ to give a test of significance level α as follows: rejecting H_0 if $\mathrm{AIC}_0 \ge \min_{1\le k\le l-1}\mathrm{AIC}_i(k)+c_\alpha$, where c_α is chosen such that

$$P(\mathrm{AIC}_0-\min_{1\le k\le l-1}\mathrm{AIC}_i(k)\ge c_\alpha|H_0)=\alpha.$$

The computation of c_α depends on the distribution of $\mathrm{AIC}_0-\min_{1\le k\le l-1}\mathrm{AIC}_i(k)$ under H_0. Since the exact distribution and the asymptotic distribution of $\mathrm{AIC}_0-\min_{1\le k\le l-1}\mathrm{AIC}_i(k)$ under H_0 is not easy to obtain, we use simulations to approximate the critical points. For a set of chosen m, n and l, we simulate 10,000 samples under H_0 when all probabilities are assumed equal, and compute the same number of test statistics $\mathrm{AIC}_0-\min_{1\le k\le l-1}\mathrm{AIC}_i(k)$ and locate the sample approximation of c_α. This process is repeated 30 times, and the mean and the standard deviation of 30 c_α's are reported. Table I presents the approximate critical points c_α for different choices of sample sizes $m = n$ and $l = 5$ when all probabilities

TABLE I Approximate Critical Values of $\mathrm{AIC}_0 - \min_{1 \le k \le l-1} \mathrm{AIC}_i(k)$.

$m = n$	$\alpha = 1\%$	$\alpha = 2.5\%$	$\alpha = 5\%$	$\alpha = 10\%$
50	7.013(0.190)	5.189(0.142)	3.878(0.064)	2.606(0.104)
	5.644(0.222)	3.891(0.097)	2.594(0.123)	1.308(0.042)
60	7.311(0.280)	5.507(0.155)	4.032(0.059)	2.690(0.076)
	5.851(0.223)	4.031(0.103)	2.689(0.097)	1.362(0.055)
70	7.336(0.259)	5.472(0.096)	3.986(0.089)	2.716(0.049)
	5.892(0.185)	4.023(0.114)	2.711(0.101)	1.388(0.075)
80	7.253(0.174)	5.418(0.132)	3.997(0.100)	2.645(0.080)
	5.862(0.193)	3.995(0.139)	2.671(0.104)	1.336(0.079)
90	7.318(0.224)	5.429(0.134)	4.006(0.083)	2.640(0.044)
	5.859(0.223)	3.972(0.083)	2.637(0.071)	1.333(0.048)
100	7.188(0.182)	5.408(0.086)	4.018(0.073)	2.671(0.040)
	5.803(0.226)	4.035(0.116)	2.684(0.062)	1.399(0.060)
200	7.088(0.252)	5.291(0.171)	3.966(0.068)	2.625(0.050)
	5.727(0.209)	3.958(0.088)	2.632(0.086)	1.348(0.052)

in H_0 are assumed equal. The first row in Table I is the approximate critical values for testing H_0 against $H_1 - H_0$ and the second row for testing H_0 against $H_2 - H_0$. The number inside the parenthesis is the standard deviation from 30 replicates.

In practice, since H_0 does not specify the common probability vector, we recommend that one use the MLEs $\hat{\mathbf{p}}^{(0)}(=\hat{\mathbf{q}}^{(0)})$ under H_0 in the simulation of the cut-off points of the test statistics. The strong consistence of $\hat{\mathbf{p}}^{(0)}(=\hat{\mathbf{q}}^{(0)})$ makes the simulated cut-off points close to the true ones when sample sizes are large.

4 EXAMPLE

We use the data set studied by Dykstra *et al.* [9] to demonstrate our estimation and testing procedures. This data set is originally discussed in a report from the Boston Collaborative Drug Surveillance Program [5]. The data set consists of observed values for the mean daily insulin dose from 80 subjects categorized as "hypoglycemia present" and 245 subjects from the population "hypoglycemia absent". The measurements are grouped into five ordered categories and are shown in Table II.

Let $\{p_i\}_{i=1}^5$ be the probabilities of five insulin levels when hypoglycemia is present, and $\{q_i\}_{i=1}^5$ be the probabilities of five insulin levels when hypoglycemia is absent. One would expect that hypoglycemia (low blood sugar) occur when large amounts of glucose are metabolized and hence would be consistent with higher levels of insulin dosage. Dykstra *et al.* [4] tested $p_1/q_1 = p_2/q_2 = \cdots = p_5/q_5$ against $p_1/q_1 \le p_2/q_2 \le \cdots \le p_5/q_5$ and concluded that there is strong evidence supporting the likelihood ratio ordering hypothesis over equality of the distributions. We would like to further locate the changepoints in the likelihood ratios. We use the binary segmentation procedure and first test H_0: $p_1/q_1 = p_2/q_2 = \cdots = p_5/q_5$

TABLE II Mean Daily Insulin Dose.

Insulin level	*1 (<0.25)*	*2 (0.25–0.49)*	*3 (0.50–0.74)*	*4 (0.75–0.99)*	*5 (≥1.00)*
Hypo. present	4	21	28	15	12
Hypo. absent	40	74	59	26	46

TABLE III Computed AIC Under H_0 and H_2.

	H_0	$H_2\ (k_1 = 1)$	$H_2\ (k_1 = 2)$	$H_2\ (k_1 = 3)$	$H_2\ (k_1 = 4)$
AIC	1016.72	1011.07	1012.87	1018.22	1018.78

TABLE IV Computed AIC Under H_0 and H_2 Conditioning on $2 \le i \le 5$.

	H_0	$H_2\ (k_2 = 2)$	$H_2\ (k_2 = 3)$	$H_2\ (k_2 = 4)$
AIC	758.97	757.20	759.01	759.03

against $H_2 - H_0$: $p_1/q_1 = \cdots = p_{k_1}/q_{k_1} < p_{k_1+1}/q_{k_1+1} = \cdots = p_5/q_5$ for some unknown $1 \le k_1 \le 4$. Table III presents the computed AIC under H_i, $i = 0, 2$.

Let $\Delta = \mathrm{AIC}_0 - \min_{1\le k\le l-1} \mathrm{AIC}_2(k) = \max_{1\le k\le 4}(\mathrm{AIC}_0 - \mathrm{AIC}_2(k))$. The computed Δ is 5.65. To find the approximate critical value for Δ, we use the MLE $\hat{\mathbf{p}}^{(0)} = \hat{\mathbf{q}}^{(0)}$ and simulate 10,000 samples under H_0 when $m = 80$ and $n = 245$ and locate the sample percentiles of Δ from the 10,000 observations. This process is repeated 30 times. The mean of the thirty 97.5th percentiles of the test statistics based on AIC is 4.331 with a standard deviation of 0.118. Thus, H_0 is rejected at 2.5% significance level, and the estimate for the changepoint is $k_1 = 1$ as it gives the minimum AIC_2. Following the binary segmentation procedure, we further locate changepoints for the subsequence when $2 \le i \le 5$. We test H_0: $p_2 \sum_{i=2}^5 q_i/(q_2 \sum_{i=2}^5 p_i) = p_3 \sum_{i=2}^5 q_i/(q_3 \sum_{i=2}^5 p_i) = p_4 \sum_{i=2}^5 q_i/(q_4 \sum_{i=2}^5 p_i) = p_5 \sum_{i=2}^5 q_i/(q_5 \sum_{i=2}^5 p_i)$ against $H_2 - H_0$: there is an increase at some $2 \le k_2 \le 4$ for the sequence $\{p_i \sum_{t=2}^5 q_t/(q_i \sum_{t=2}^5 p_t)\}_{i=2}^5$. Conditioning on $2 \le i \le 5$, we compute AIC under H_0 and H_2. Table IV presents these computation results.

The test statistics for testing H_0 *vs*. $H_2 - H_0$ based on AIC is 1.77. We again locate the sample percentiles of the test statistics under H_0 by repeatedly simulating 10,000 samples of sizes 76 and 205 from $\mathbf{p}$ and $\mathbf{q}$ under H_0 and computing the test statistics. The mean of the thirty 70th percentiles of the test statistics based on AIC is 2.202 with a standard deviation of 0.014. Thus we fail to reject H_0 at 30% significance level. In conclusion, among five insulin levels, there is only one significant increase in likelihood ratios and that increase is from level 1 to level 2.

Acknowledgements

The authors wish to express their sincere thanks to the editor, associated editor and referees for their constructive and valuable comments and suggestions which considerably improved the manuscript.

References

[1] Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle. In: Petrov, B. N. and Csaki, F. (Eds.). *Proceedings of the 2nd International Symposium of Information Theory.* Akademiai Kiado, Budapest, pp. 267–281.

[2] Anraku, K. (1999). An information criterion for parameters under simple order restriction. *Biometrika*, **86**, 141–152.

[3] Antoch, J. and Hušková, M. (1995). Change-point problem and bootstrap. *J. Nonparametr. Statist.*, **5**, 123–144.

[4] Brodsky, B. E. and Darkhovsky, B. S. (1993). *Nonparametric Methods in Change-point Problems, Mathematics and its Applications*, Vol. 243. Kluwer Academic Publishers Group, Dordrecht.

[5] Boston Collaborative Drug Surveillance Program (1974). Relations of bodyweight and insulin dose to the frequency of hypoglycemia. *J. Am. Med. Assoc.*, **228**, 192–194.
[6] Chatterjee, S. K. and Bandyopadhyay, U. (1992). On the change-point problem. *Sequential Anal.*, **11**, 101–118.
[7] Chen, J. and Gupta, A. K. (1997). Testing and locating variance changepoints with applications to stock prices. *J. Am. Statist. Assoc.*, **92**, 739–747.
[8] Chernoff, H. and Zacks, S. (1964). Estimating the current mean of a normal distribution which is subjected to changes in time. *Ann. Math. Statist.*, **35**, 999–1018.
[9] Dykstra, R. L., Kochar, S. and Robertson, T. (1995). Inference for likelihood ratio ordering in the two-sample problem. *J. Am. Statist. Assoc.*, **90**, 1034–1040.
[10] Dykstra, R. L., Kochar, S. and Robertson, T. (1998). Restricted tests for testing independence of time to failure and cause of failure in a competing-risks model. *Can. J. Statist.*, **26**, 57–68.
[11] Gardner, L. A. (1969). On detecting change in the mean of normal variates. *Annals of Mathematical Statistics*, **40**, 116–126.
[12] Giraitis, L., Leipus, R. and Surgailis, D. (1996). The change-point problem for dependent observations. *J. Statist. Plann. Inference*, **53**, 297–310.
[13] Haccou, P., Meelis, E. and Van de Geer, S. (1987). The likelihood ratio test for the change point problem for exponentially distributed random variables. *Stochastic Process. Appl.*, **27**, 121–139.
[14] Kander, Z. and Zacks, S. (1966). Test procedures for possible changes in parameters of statistical distributions occurring at unknown time points. *Ann. Math. Statist.*, **37**, 1196–1210.
[15] Keilson, J. and Sumita, U. (1982). Uniform stochastic ordering and related inequalities. *Can. J. Statist.*, **10**, 181–198.
[16] Kullback, S. and Leibler, R. A. (1951). On information and sufficiency. *Annals of Mathematical Statistics*, **22**, 79–86.
[17] Lavielle, M. and Ludeña, C. (2000). The multiple change-points problem for the spectral distribution. *Bernoulli*, **6**, 845–869.
[18] Lehmann, E. L. (1955). Ordered families of distributions. *Annals of Mathematical Statistics*, **26**, 399–419.
[19] Pettitt, A. N. (1979). A nonparametric approach to the change-point problem. *J. Roy. Statist. Soc. Ser. C*, **28**, 126–135.
[20] Robertson, T., Wright, F. T. and Dykstra, R. L. (1988). *Order Restricted Statistical Inference.* John Wiley & Sons, Chichester.
[21] Ross. S. (1983). *Stochastic Processes.* John Wiley & Sons, New York.
[22] Scariano, S. M. and Watkins, T. A. (1988). Nonparametric point estimators for the change-point problem. *Communications in Statistics–Theory and Method*, **17**, 3645–3675.
[23] Sen, P. K. (1994). Some change-point problems in survival analysis: Relevance of nonparametrics in applications. *J. Appl. Statist. Sci.*, **1**, 425–444.
[24] Shaked, M. and Shanthikumar, J. G. (1994). *Stochastic Orders and Their Applications.* Academic Press, San Diego.
[25] Siegmund, D. (1988). Confidence sets in change-point problems. *Internat. Statist. Rev.*, **56**, 31–48.
[26] Sugiura, N. (1978). Further analysis of the data by Akaike's information criterion and the finite corrections. *Communications in Statistics–Theory and Method*, **7**, 13–26.
[27] Vostrikova, L. J. (1981). Detecting 'disorder' in multidimensional random processes. *Soviet Mathematica Doklady*, **24**, 55–59.
[28] Xiong, C. and Milliken, G. A. (2000). Changepoints in stochastic ordering. *Communications in Statistics–Theory and Method*, **29**, 381–400.
[29] Zacks, S. (1983). Survey of classical and Baysian approaches to the change-point problem: Fixed sample and sequential procedures of testing and estimation. *Recent Advances in Statistics.* Academic Press, NY, pp. 245–269.
[30] Zacks, S. (1991). Detection and change-point problems. *Handbook of Sequential Analysis, Statist. Textbooks Monogr., 118.* Dekker, New York, pp. 531–562.
[31] Zacks, S. (1994). Sequential testing of reliability system with change points. *J. Appl. Statist. Sci.*, **1**, 469–483.
[32] Zacks, S. (1995). Sequential testing of reliability system with change points. *Applied Change Point Problems in Statistics.* Nova Sci. Publ., Commack, NY, pp. 147–161.