CHANGE POINT PROBLEMS IN THE MODEL OF LOGISTIC REGRESSION [1]

By G. Gurevich and A. Vexler

Tecnion-Israel Institute of Technology, Haifa, Israel

and

The Central Bureau of Statistics, Jerusalem, Israel

SUMMARY

The paper considers generalized maximum likelihood asymptotic power one tests which aim to detect a change point in logistic regression when the alternative specifies that a change occurred in parameters of the model. A guaranteed non-asymptotic upper bound for the significance level of each of the tests is presented. For cases in which the test supports the conclusion that there was a change point, we propose a maximum likelihood estimator of that point and present results regarding the asymptotic properties of the estimator. An important field of application of this approach is occupational medicine, where for a lot chemical compounds and other agents, so-called threshold limit values (or TLVs) are specified.

We demonstrate applications of the test and the maximum likelihood estimation of the change point using an actual problem that was encountered with real data.

[1] AMS 2000 *subject classifications.* Primary 62F03, 62H15, 62F10; Secondary 62F05, 62L10.

Key words and phrases. change point, logistic regression, logit, maximum likelihood estimation, martingale.

This research was supported by the Israel Science Foundation and by the Marcy Bogen Chair of Statistics at the Hebrew University of Jerusalem.

1 Introduction

We consider a finite sequence of independent observations $(Y_i, x_i), i \geq 1$ where $x_i = [x_{i1}, \ldots, x_{id}]^T, i \geq 1$ are fixed d-vector explanatory variables and $Y_i = [Y_{i1}, \ldots, Y_{im_i}]^T, i \geq 1$, where Y_{ij} are Bernoulli variates whose exact distribution depends on the predictor x_i. We assume that

$$\begin{aligned} P\{Y_{ij} = 1|x_i\} &= (1 + \exp(-x_i^T \beta_0))^{-1} I\{i < \nu\} \\ &\quad + (1 + \exp(-x_i^T \beta_1))^{-1} I\{i \geq \nu\}, \\ \beta_k = [\beta_{k1}, \ldots, \beta_{kd}]^T, &\qquad k = 0, 1,\ j = 1, \ldots, m_i,\ i = 1, \ldots, n,\ \beta_0 \neq \beta_1, \end{aligned} \tag{1.1}$$

where $I\{\cdot\}$ is the indicator function and $\beta_0,\ \beta_1,\ \nu$ are parameters of the model. The model corresponds to a situation where up to an unknown change point $\nu > 0$, the observations satisfy the logistic regression model with parameter β_0. Beyond that (if $\nu \leq n$), the parameter changes to β_1.

Formally, the problem examined in this study is a change point problem of hypothesis testing, where

$$\mathbf{H_0} \ :\ \nu \notin [1, n], \quad \text{versus} \quad \mathbf{H_1} \ :\ 1 \leq \nu \leq n; \quad \nu > 0 \text{ is unknown.} \tag{1.2}$$

One of the main problems considered in the change point literature relates to testing, which focuses on a sequence of previously obtained independent observations in an attempt to determine whether they all share the same distribution. It is usually assumed that if a change in the distribution did occur, then it is unique and the observations after the change all have the same distribution, which differs from the distribution of the observations before the change (e.g. Page (1954), Sen and Srivastava (1975), Pettitt (1980), Worsley (1983), Gombay and Horvath (1994), Ferger (1994)). A number of studies have dealt with independent observations which are not identically distributed (e.g. Quandt (1960), Brown, Durbin and Evans (1975), Kim and Siegmund (1989), Ulm (1991), Pastor and Guallar (1998)). Kim and Siegmund (1989) considered likelihood ratio tests which aim to detect a change point in simple linear regression when the alternative specifies that the intercept and the slope change. They obtained approximations for the significance level of the tests based on asymptotic analysis. Ulm (1991) proposed a test for detecting a change in logistic regression parameters based on a quasi one-sided χ_1^2 distribution. Except for the simulation study, the author has not represented rigorous theoretical substantiation of his method. Note that many epidemiological studies are based on a two-segmented logistic regression model, in which the linear term associated with a continuous exposure in standard logistic regression is replaced by a two-segmented polynomial function with an unknown change point. In these studies, the parameters of the model (including the change point) are often estimated without previous formal testing of the existence of the change point (e.g. Pastor and Guallar (1998), Gossl and Kuchenhoff (2001)). It is reasonable to assume that estimation of a change point without testing its existence may be an erroneous procedure.

The main goal of the paper is to present maximum likelihood asymptotic power one tests with bounded significance levels for the problem under consideration. In this way, we obtain non-asymptotic bounds for the significance levels of our tests that do not depend on explanatory variables x. Therefore, no assumptions are made about behaviour of x. We represent Monte Carlo simulations to demonstrate the accuracy of suggested bounds for the significance levels and powers of our tests. Our approach is based on methods developed through a sequential testing of a change in the distribution's parameters of independent identically distributed ($i.i.d$) observations in the ranges before and after the possible change. In this context, Robbins and Siegmund (1973) constructed a power one class of tests based on estimation of the unknown parameters that appeared after the change point occurred (prior to that point the parameters were known). In order to preserve an H_0-martingale property of the sequence of likelihood ratios, they proposed that the estimator of the parameter used in conjunction with the n-th observation be a function of the first $n-1$ observations only. Based on this idea they produced an upper bound for the level of significance. This method was used by Yakir, Krieger and Pollak (1999) as well as Lorden and Pollak (2004) in the sequential approach for two-phase linear regression and i.i.d observations. We propose retrospective tests for the non-linear model (1.1), which are also based on a statistic that has a martingale structure.

Note that if the application of the test ends with the acceptance of hypothesis H_1 then we propose the maximum likelihood estimator (MLE) of the change point parameter and present results for its asymptotic properties.

The paper is organized as follows. The next sections presents a test for the simple case of the problem (when β_0 and β_1 are known), and the MLE of the change point is defined. Section 3 considers a more complicated case, where β_0 is known and β_1 is unknown. Section 4 presents a situation where both β_0 and β_1 are unknown. Section 5 represents the results of Monte Carlo simulations. An example of real data for the application of the test and the MLE is demonstrated in Section 6.

2 Procedure for both initial and final parameters known

We consider the model (1.1) and test the hypothesis (1.2), where β_0, β_1 are known. The analysis is relatively clear, and has the basic ingredients for more general cases. We posit that this simple case can be practically associated with the following examples: a technical tool manual will indicate β_0 and β_1, if a functional change in the tool's smooth work is expected by the manufacturer; if there are large tails of the sample, i.e., numerous observations are made before and after a possible change in the distribution; in many studies of occupational medicine, the distribution function of "being ill" is known a priori and the dependence on some agents varies before and after some possible TLVs.

Let P_{H_0} denote the distribution of the sequence $Y_1, \ldots, Y_n$ under H_0 and $P_{\nu=k}$ denote the distribution of the sequence $Y_1, \ldots, Y_n$ under H_1 with $\nu = k$. Likewise, let E_{H_0} and $E_{\nu=k}$ denote expectation under P_{H_0} and $P_{\nu=k}$, respectively. Denote

$$\begin{aligned}\Lambda_k^n\left(\beta_0, \beta_1 | x_k, \ldots, x_n\right) &= \frac{\prod_{i=k}^n \prod_{j=1}^{m_i} P_{\nu=k}\{Y_{ij}|x_i\}}{\prod_{i=k}^n \prod_{j=1}^{m_i} P_{H_0}\{Y_{ij}|x_i\}} \\ &= \frac{\prod_{i=k}^n \prod_{j=1}^{m_i} P_{\nu=i}\{Y_{ij}|x_i\}}{\prod_{i=k}^n \prod_{j=1}^{m_i} P_{H_0}\{Y_{ij}|x_i\}} \\ = \exp\left(\sum_{i=k}^{n}\sum_{j=1}^{m_i}(x_i^T\beta_1 - x_i^T\beta_0)Y_{ij}\right) & \prod_{i=k}^{n}\left(\frac{1+\exp(x_i^T\beta_0)}{1+\exp(x_i^T\beta_1)}\right)^{m_i},\end{aligned} \tag{2.3}$$

where

$$\begin{aligned}P_{\nu=k}\{Y_{ij}|x_i\} &= P_{\nu=k}\{Y_{ij}=1|x_i\}^{Y_{ij}}\left(1 - P_{\nu=k}\{Y_{ij}=1|x_i\}\right)^{1-Y_{ij}}, \\ P_{H_0}\{Y_{ij}|x_i\} &= P_{H_0}\{Y_{ij}=1|x_i\}^{Y_{ij}}\left(1 - P_{H_0}\{Y_{ij}=1|x_i\}\right)^{1-Y_{ij}}.\end{aligned}$$

We propose the following test: reject H_0 iff

$$\max_{1\le k\le n} \Lambda_k^n\left(\beta_0,\beta_1|x_k,\ldots,x_n\right) \;>\; C. \tag{2.4}$$

Significance level of the test. In (2.4) we proposed a generalized maximum likelihood ratio test for the problem (1.2). It is widely known in change point literature that such tests have high power. Therefore, evaluation of their significance level is a major issue. Most results deal with the significance level of the generalized maximum likelihood ratio tests (even in the simplest cases of i.i.d observations before and after the change) are asymptotic ($n\to\infty$) (e.g. Gombay and Horvath (1994)). In Proposition 2.1 we obtain a guaranteed non-asymptotic upper bound for the significance level of the proposed test.

Proposition 2.1 *The significance level α of the test satisfies:*

$$\alpha = P_{H_0}\left\{\max_{1\le k\le n} \Lambda_k^n\left(\beta_0,\beta_1|x_k,\ldots,x_n\right) > C\right\} \le 1/C.$$

Proof. In Appendix.

We have the upper bound (that does not involve n and no conditions are considered on explanatory variables $x_1,\ldots,x_n$) for the significance level of the test (2.4): $\alpha\le 1/C$. Thus, $C=1/\alpha$ determines a test with a level of significance that does not exceed α.

Note that even if we assume that $x_i\equiv c$, then we have i.i.d random variates (under H_0) and use of the known results of the renewal theory regarding overshooting will lead us to the asymptotic ($n\to\infty$) result for α (e.g. Woodroofe (1982)): $\alpha = Z(\beta_0,\beta_1,c)/C + o\left(1/C\right)$, where $0<Z(\beta_0,\beta_1,c)\le 1$ is some function of β_0, β_1 and c, $o\left(1/C\right)C\to 0$ as $C\to\infty$.

Power of the test. We define for all $1 \leq i \leq n$

$$
\begin{aligned}
a_i &\equiv E_\nu \ln \left(\prod_{j=1}^{m_i} \frac{P_\nu\{Y_{ij}|x_i\}}{P_{H_0}\{Y_{ij}|x_i\}} \right) \qquad (2.5) \\
&= m_i \left(\left(x_i^T \beta_1 - x_i^T \beta_0 \right) \frac{\exp(x_i^T \beta_1)}{1 + \exp(x_i^T \beta_1)} + \ln \left(\frac{1 + \exp(x_i^T \beta_0)}{1 + \exp(x_i^T \beta_1)} \right) \right)
\end{aligned}
$$

it is clear that if $i < \nu$ than $a_i \leq 0$ and if $i \geq \nu$: $a_i \geq 0$.

The following proposition presents an expression for the power of the test (2.4).

Proposition 2.2 *Assume: for some* $N > 0$ *and* $(n - \nu) > N$: $\sum_{i=\nu}^n a_i > \ln(C)$ *and*

$$
\frac{\left(\sum_{i=\nu}^n a_i - \ln(C)\right)^2}{\sum_{i=\nu}^n \left(m_i |x_i^T (\beta_1 - \beta_0)|\right)^2} \to \infty, \qquad \text{as} \qquad (n - \nu) \to \infty.
$$

Then for all $0 < \delta < 1/2$

$$
\begin{aligned}
& P_\nu \left\{ \max_{1 \leq k \leq n} \Lambda_k^n \left(\beta_0, \beta_1 | x_k, \ldots, x_n\right) \leq C \right\} \qquad (2.6) \\
\leq\ & \exp \left(- \frac{\left(\sum_{i=\nu}^n a_i - \ln(C)\right)^2}{2 \sum_{i=\nu}^n \left(m_i |x_i^T (\beta_1 - \beta_0)|\right)^2} \right) \qquad \text{and} \\
& \lim_{(n-\nu) \to \infty} \exp \left(\delta \frac{\left(\sum_{i=\nu}^n a_i - \ln(C)\right)^2}{\sum_{i=\nu}^n \left(m_i |x_i^T (\beta_1 - \beta_0)|\right)^2} \right) \\
& \times \left(1 - P_\nu \left\{ \max_{1 \leq k \leq n} \Lambda_k^n \left(\beta_0, \beta_1 | x_k, \ldots, x_n\right) > C \right\} \right) = 0.
\end{aligned}
$$

Proof. In Appendix.

Proposition 2.2 shows that the probability of type II error is bounded by an exponentially vanishing term.

Remark 2.1. Consider the assumption in Proposition 2.2 on a simple example. Let $m_i = 1$ be for all $i \in [\nu, n]$. Hence, by (2.5) we have

$$
a_i = -E_\nu \ln \left(\frac{P_{H_0}\{Y_{i1}|x_i\}}{P_\nu\{Y_{i1}|x_i\}} \right) \geq - \ln \left(E_\nu \frac{P_{H_0}\{Y_{i1}|x_i\}}{P_\nu\{Y_{i1}|x_i\}} \right) = 0. \qquad (2.7)
$$

Suppose that in addition $\epsilon_1 > 0, \epsilon_2 > 0$ exists, such that for all $i \in [\nu, n]$: $a_i \geq \epsilon_1$ and $|x_i^T(\beta_1 - \beta_0)| \leq \epsilon_2$. Obviously, if $(n-\nu)$ is large enough, then for a fixed C: $\sum_{i=\nu}^n a_i > (n-\nu)\epsilon_1 > \ln(C)$ and

$$\frac{\left(\sum_{i=\nu}^n a_i - \ln(C)\right)^2}{\sum_{i=\nu}^n \left(m_i|x_i^T(\beta_1-\beta_0)|\right)^2} \geq \frac{\epsilon_1^2(n-\nu)^2 - 2\ln(C)\sum_{i=\nu}^n a_i}{\epsilon_2^2(n-\nu)} \to \infty, \text{ as } (n-\nu)\to\infty.$$

Therefore, according to Proposition 2.2 constants $c_1 > 0$ and $c_2 > 0$ are such that for the large N

$$P_\nu\left\{\max_{1\leq k\leq n} \Lambda_k^n\left(\beta_0, \beta_1|x_k, \ldots, x_n\right) \leq C\right\} \leq c_1 \exp\left(-c_2(n-\nu)\right).$$

Maximum Likelihood Estimation of a change point. If H_0 is rejected, then, following Borovkov (1998), the following estimator of ν is considered

$$\widehat{\nu}_n = \arg\max_{1\leq k\leq n} \Lambda_k^n\left(\beta_0, \beta_1|x_k, \ldots, x_n\right). \tag{2.8}$$

The following proposition delineates several properties of this estimator.

Proposition 2.3 *For all $t_1 > \nu$ and $t_2 < \nu$*

$$P_\nu\left\{\widehat{\nu}_n \geq t_1\right\} \quad \leq \quad \sum_{k\geq t_1}^n \exp\left(-\frac{\left(\sum_{i=\nu}^{k-1} a_i\right)^2}{2\sum_{i=\nu}^{k-1}\left(m_i|x_i^T(\beta_1-\beta_0)|\right)^2}\right) \tag{2.9}$$

and

$$P_\nu\left\{\widehat{\nu}_n \leq t_2\right\} \quad \leq \quad \sum_{k\leq t_2} \exp\left(-\frac{\left(\sum_{i=k}^{\nu-1} -a_i\right)^2}{2\sum_{i=k}^{\nu-1}\left(m_i|X_i^T(\beta_1-\beta_0)|\right)^2}\right). \tag{2.10}$$

If, additionally, we assume that for some $N > 0$, $c > 0$, $p > 2$ and $(n-\nu) > N$:

$$\begin{aligned}
\sum_{k\geq t_1}^n \exp\left(-\frac{\left(\sum_{i=\nu}^{k-1} a_i\right)^2}{2\sum_{i=\nu}^{k-1}\left(m_i|x_i^T(\beta_1-\beta_0)|\right)^2}\right) \quad &\leq \quad \frac{c}{|t_1-\nu|^p}, \\
\sum_{k<t_2} \exp\left(-\frac{\left(\sum_{i=k}^{\nu-1} -a_i\right)^2}{2\sum_{i=k}^{\nu-1}\left(m_i|x_i^T(\beta_1-\beta_0)|\right)^2}\right) \quad &\leq \quad \frac{c}{|t_2-\nu|^p},
\end{aligned} \tag{2.11}$$

then $E_\nu \widehat{\nu}_n = \nu + O(1)$ and the estimator $\widehat{\nu}_n$ is asymptotically homogeneous (see Borovkov (1998)) as $(n-\nu) \to \infty$, where $O(1)$ is bounded and a_i by (2.5).

Proof. In Appendix.

Remark 2.2. Consider the conditions (2.11) for the simple example in Remark 2.1. Note that for $i \in [1, \nu - 1]$

$$a_i \quad = \quad E_{H_0} \ln \left(\frac{P_\nu\{Y_{i1}|x_i\}}{P_{H_0}\{Y_{i1}|x_i\}} \right) \le \ln \left(E_{H_0} \frac{P_\nu\{Y_{i1}|x_i\}}{P_{H_0}\{Y_{i1}|x_i\}} \right) = 0.$$

Suppose additionally that for all $i \in [1, \nu - 1]$: $-a_i \geq \epsilon_1$ and $|x_i^T(\beta_1 - \beta_0)| \le \epsilon_2$. It is obvious that in this case there are constants $c_3 > 0$, $c_4 > 0$, $c_5 > 0$ and $c_6 > 0$ so that

$$\begin{aligned} t > \nu :\ & \sum_{k \geq t}^{n} \exp \left(- \frac{\left(\sum_{i=\nu}^{k-1} a_i \right)^2}{2 \sum_{i=\nu}^{k-1} \left(m_i |x_i^T(\beta_1 - \beta_0)| \right)^2} \right) \\ \le & \sum_{k=[t]-1}^{n} c_3 e^{-c_4(k-\nu)} = c_3 \frac{e^{-c_4([t]-\nu-1)} - e^{-c_4(n-\nu)}}{1 - e^{-c_4}}, \\ t < \nu :\ & \sum_{k<t} \exp \left(- \frac{\left(\sum_{i=k}^{\nu-1} -a_i \right)^2}{2 \sum_{i=k}^{\nu-1} \left(m_i |x_i^T(\beta_1 - \beta_0)| \right)^2} \right) \\ \le & \sum_{k=1}^{[t]} c_5 e^{-c_6(\nu-k)} = c_5 \frac{e^{-c_6(\nu-[t])} - e^{-c_6(\nu-1)}}{e^{c_6} - 1} \end{aligned} \tag{2.12}$$

where $[t]$ is an integer value of t. Here, it is clear that the condition (2.11) is fulfilled.

3 Procedure for known initial parameters and unknown final parameters

This Section considers the same problem as in Section 2, but β_1 is unknown. This case has been dealt with extensively in the literature (e.g. Yakir, Krieger, Pollak (1999) and Lorden, Pollak (2004)) and corresponds to a control problem where the baseline in-control distribution is known and the post-change out-of-control distribution is not. One application of this problem is found in epidemiological studies where it is often assumed that explanatory variables have no effect on a response prior to a certain unknown change point.

We denote

$$\Lambda_k^n\left(\beta_0, \widehat{\beta}|x_k, \ldots, x_n\right) = \prod_{i=k}^{n}\prod_{j=1}^{m_i}\left(\frac{\exp(x_i^T\beta_0 Y_{ij})}{1+\exp(x_i^T\beta_0)}\right)^{-1} \times \frac{\exp(x_i^T\widehat{\beta}^{(i+1,n)}Y_{ij})}{1+\exp(x_i^T\widehat{\beta}^{(i+1,n)})}, \tag{3.13}$$

where $\widehat{\beta}^{(i+1,n)} \in \Im\left(\{(Y_r, x_r)\}_{r=i+1}^n\right)$, $1 \leq i \leq n$ is any estimator of β_1 in the σ-algebra generated by $\{(Y_r, x_r)\}_{r=i+1}^n$ ($\widehat{\beta}^{(i+1,n)}$ based upon $\{(Y_r, x_r)\}_{r=i+1}^n$), and where $\widehat{\beta}^{(n+1,n)} = \beta_0$. Therefore, $\Lambda_k^n\left(\beta_0, \widehat{\beta}|x_k, \ldots, x_n\right)$ is the estimator of the likelihood ratio $\Lambda_k^n\left(\beta_0, \beta_1|x_k, \ldots, x_n\right)$ from (2.3).

We propose the following test: reject H_0 iff

$$\max_{1\leq k\leq n} \Lambda_k^n\left(\beta_0, \widehat{\beta}|x_k, \ldots, x_n\right) > C. \tag{3.14}$$

Significance level of the test.

Proposition 3.1 *The significance level α of the test satisfies:*

$$\alpha = P_{H_0}\left\{\max_{1\leq k\leq n} \Lambda_k^n\left(\beta_0, \widehat{\beta}|x_k, \ldots, x_n\right) > C\right\} \leq 1/C.$$

Proof. In Appendix.

Power of the test. The next result gives an expression for the power of the test (3.14).

Proposition 3.2 *Assume: for some* $0 < \rho < 1$, $N > 0$ *and* $(n-\nu) > N$: $(1-\rho)\sum_{i=\nu}^{n} a_i > \ln(C)$ *and*

$$\frac{\left((1-\rho)\sum_{i=\nu}^{n} a_i - \ln(C)\right)^2}{\sum_{i=\nu}^{n}\left(m_i|x_i^T(\beta_1-\beta_0)|\right)^2} \to \infty, \qquad as \qquad (n-\nu)\to\infty,$$

where $a_i > 0$ *is defined in (2.5). Then*

$$\begin{aligned} 0 \ &\leq\ 1 - P_\nu\left\{\max_{1\leq k\leq n} \Lambda_k^n\left(\beta_0, \widehat{\beta}|x_k,\ldots,x_n\right) > C\right\} \\ &\leq\ \exp\left(-\frac{\left((1-\rho)\sum_{i=\nu}^{n} a_i - \ln(C)\right)^2}{2\sum_{i=\nu}^{n}\left(m_i|x_i^T(\beta_1-\beta_0)|\right)^2}\right) + \tau_{\nu,n}, \end{aligned} \tag{3.15}$$

where

$$\tau_{\nu,n} \ \equiv\ P_\nu\left\{\sum_{i=\nu}^{n} 2m_i|x_i^T(\widehat{\beta}^{(i+1,n)}-\beta_1)| \geq \rho\sum_{i=\nu}^{n} a_i\right\}. \tag{3.16}$$

Proof. In Appendix.

From this proposition, we ascertain that the probability of type II error is bounded by an exponentially vanishing term plus a term $\tau_{\nu,n}$ that depends on the estimates used.

Remark 3.1. In the case that the explanatory variables x are such that $\sum_{i=\nu}^{n} a_i \to +\infty$ (see for example, Remark 2.1). At this rate, if we suppose consistency of the estimator $\widehat{\beta}^{(i+1,n)}$ ($\widehat{\beta}^{(i+1,n)} \to \beta_1$, since $(n-i)$ is arbitrarily large, $i \geq \nu$), it can thus reasonably assumed that $\sum_{i=\nu}^{n} 2m_i|x_i^T(\widehat{\beta}^{(i+1,n)}-\beta_1)| \Big/ \sum_{i=\nu}^{n} a_i \to 0$. Hence, for (3.15), (3.16) $\tau_{\nu,n} \to 0$, as $(n-\nu)\to\infty$.

Maximum Likelihood Estimation of a change point. If H_0 is rejected, then following Borovkov (1998), the following estimator of ν is considered

$$\widehat{\nu}_n = \arg\max_{1\leq k\leq n} \prod_{i=k}^{n}\prod_{j=1}^{m_i}\left(\frac{\exp(x_i^T\beta_0 Y_{ij})}{1+\exp(x_i^T\beta_0)}\right)^{-1} \frac{\exp(x_i^T\widehat{\beta}^{(k,n)}Y_{ij})}{1+\exp(x_i^T\widehat{\beta}^{(k,n)})}, \tag{3.17}$$

where $\widehat{\beta}^{(k,n)}$ is the MLE of β_1 in the σ-algebra generated by $(Y_k,x_k),\ldots,(Y_n,x_n)$.

Proposition 3.3 *Assume: for some $N > 0,\ 0 < \rho_1 < 1,\ 0 < \rho_2 < 1,\ c > 0,\ t_1 > \nu,\ t_2 < \nu,\ p > 2$ and $n - \nu > N$:*

$$\sum_{t_1 \le k \le n} P_\nu \left\{ \sum_{i=\nu}^{k-1} 2m_i |x_i^T(\widehat{\beta}^{(k,n)} - \beta_1)| \ge \rho_1 \sum_{i=\nu}^{k-1} a_i \right\} \le \frac{c}{|t_1-\nu|^p}, \tag{3.18}$$

$$\sum_{1 \le k \le t_2} P_\nu \left\{ \sum_{i=k}^{\nu-1} 2m_i |x_i^T(\widehat{\beta}^{(k,n)} - \beta_1)| \ge \rho_2 \sum_{i=k}^{\nu-1} (-a_i) \right\} \le \frac{c}{|t_2-\nu|^p}, \tag{3.19}$$

$$\sum_{t_1 \le k \le n} \exp\left(- \frac{(1-\rho_1)^2 \left(\sum_{i=\nu}^{k-1} a_i \right)^2}{2 \sum_{i=\nu}^{k-1} \left(m_i |x_i^T(\beta_1 - \beta_0)| \right)^2} \right) \le \frac{c}{|t_1-\nu|^p}, \tag{3.20}$$

$$\sum_{1 \le k \le t_2} \exp\left(- \frac{(1-\rho_2)^2 \left(\sum_{i=k}^{\nu-1} -a_i \right)^2}{2 \sum_{i=k}^{\nu-1} \left(m_i |x_i^T(\beta_1 - \beta_0)| \right)^2} \right) \le \frac{c}{|t_2-\nu|^p},$$

where a_i by (2.5).

Then $E_\nu \widehat{\nu}_n = \nu + O(1)$ and the estimator $\widehat{\nu}_n$ is asymptotically homogeneous (see Borovkov (1998)) as $(n-\nu) \to \infty$, where $O(1)$ is bounded.

Proof. In Appendix.

Remark 3.2. An analysis of the conditions (3.18), (3.19) may be applied, as in Remark 3.1. The investigation of a condition similar to (3.20) is presented in Remark 2.2.

4 Procedure for both initial and final parameters unknown

This section considers situations in which both β_0 and β_1 are unknown, and proposes a test based on the Bayesian approach.

Denote for some $1 \le L \ll n$ and $L \le k \le n$

$$\Lambda_k^n \left(\widehat{\beta} | x_k, \ldots, x_n \right) = \prod_{i=L}^{k-1} \prod_{j=1}^{m_i} \frac{f\{Y_{ij} | x_i, \widehat{\beta}^{(1,i-1)}\}}{f\{Y_{ij} | x_i, \widehat{\beta}^{(L,n)}\}} \tag{4.21}$$

$$\times \prod_{i=k}^{n} \prod_{j=1}^{m_i} \frac{f\{Y_{ij}|x_i, \widehat{\beta}^{(k,i-1)}\}}{f\{Y_{ij}|x_i, \widehat{\beta}^{(L,n)}\}},$$

where L may be arbitrarily chosen depending on the dimension of parameter β (on the a priori assumption that the first $L-1$ observations come before the change). Such observations are usually called a learning sample (e.g. Yakir (1998)). The condition probability function

$$f\{Y|x,\beta\} \equiv \frac{\exp\left(x^T\beta Y\right)}{1+\exp\left(x^T\beta\right)}, \quad \prod_{m=1}^{0} = 1,$$

$\widehat{\beta}^{(L,n)}$ is the MLE of β_0 (under the hypothesis H_0) based upon $((Y_L, x_L), \ldots, (Y_n, x_n))$:

$$\widehat{\beta}^{(L,n)} \equiv \arg\max_{\beta^*} \prod_{r=L}^{n} \prod_{h=1}^{m_r} f\{Y_{rh}|x_r, \beta^*\},$$

$\widehat{\beta}^{(l_1,l_2)} \in \Im\left((Y_{l_1}, x_{l_1}), \ldots, (Y_{l_2}, x_{l_2})\right)$, $1 \le l_1 \le l_2 \le n-1$ is an estimator of β based upon $(Y_{l_1}, x_{l_1}), \ldots, (Y_{l_2}, x_{l_2})$ (for example, the MLE based upon $(Y_{l_1}, x_{l_1}), \ldots, (Y_{l_2}, x_{l_2})$ observations) and for $l_1 > l_2$: $\widehat{\beta}^{(l_1,l_2)} = \widehat{\beta}^{(1,l_2)}$, $\widehat{\beta}^{(1,0)} = [0, \ldots, 0]^T$. Therefore, $\Lambda_k^n\left(\widehat{\beta}|x_k, \ldots, x_n\right)$ is the estimator of the likelihood ratio
$\Lambda_k^n\left(\beta_0, \beta_1|x_k, \ldots, x_n\right)$ from (2.3).

We propose the following test: reject H_0 iff

$$\sum_{k=L}^{n} \omega_k \Lambda_k^n\left(\widehat{\beta}|x_k, \ldots, x_n\right) > C, \tag{4.22}$$

where $\{\omega_k,\ k = L, \ldots, n\}$ is a sequence of positive numbers such that $\sum_{k=L}^{n} \omega_k = 1$, e.g., we can choose $\omega_k = 1/(n-L+1)$. In other words, we consider $\omega_k,\ L \le k \le n$ as a prior distribution of the change point: $P(\nu = k) = \omega_k$.

Proposition 4.1 *The significance level α of the test satisfies:*

$$\alpha = P_{H_0}\left\{\sum_{k=L}^{n} \omega_k \Lambda_k^n\left(\widehat{\beta}|x_k, \ldots, x_n\right) > C\right\} \le 1/C.$$

Proof. In Appendix.

If H_0 is rejected, then the estimator of ν is considered

$$\widehat{\nu}_n = \arg\max_{2\le k\le n} \prod_{i=k}^{n}\prod_{j=1}^{m_i} \left(\frac{\exp(x_i^T \widehat{\beta}^{(1,k-1)} Y_{ij})}{1+\exp(x_i^T \widehat{\beta}^{(1,k-1)})} \right)^{-1} \frac{\exp(x_i^T \widehat{\beta}^{(k,n)} Y_{ij})}{1+\exp(x_i^T \widehat{\beta}^{(k,n)})}, \tag{4.23}$$

where $\widehat{\beta}^{(j,m)} \in \Im\left((Y_j, x_j), \ldots, (Y_m, x_m)\right), \quad 1 \le j \le m \le n$ is some esimator of β (e.g. MLE based upon $((Y_j, x_j), \ldots, (Y_m, x_m))$).

5 Monte Carlo Simulation Study

The results reported in this section focus on the performance of the model (1.1), where $m_i = 1$. The samples were generated from the following scheme for all of the results presented in Tables 1-4: We generated $x_i = [1, \zeta_i]^T$, where $\zeta_i \sim$ Uniform$[0,1]$ and applied $\beta_0 = [-0.1, 1]^T,\ \beta_1 = [-0.1, -3.7]^T$ to model (1.1). Then 3000 simulations of the random sequence $\{Y_{i1}, 1 \le i \le n\}$ were generated, which satisfy (1.1). In our application of the tests (3.14), (4.22) and the estimators (3.17), (4.23) of ν (the change point), we used MLEs of β. At this rate, $\widehat{\beta}^{(l,n)} = \beta_0,\ n - l \le 5$ was applied in test (3.14), and $L = 5,\ \omega_k = 1/(n-4)$ were used in the formulation of (4.22).

Table 1 displays the simulated significance levels for the proposed tests (2.4), (3.14) and (4.22) for different sample sizes n and thresholds C, where $\nu > n$ in (1.1). From these results, we can experimentally conclude that the upper bounds by Propositions 2.1, 3.1 and 4.1 are reasonable in the considered case.

-Table 1 here-

Table 2 presents the simulated power of the test (2.4), the boundary by Proposition 2.2, the Monte Carlo mean and the standard error (SE) of estimator $\widehat{\nu}_n$ (2.8). The values of the boundaries B in Table 2 asymptotically approach the simulated power.

-Table 2 here-

Tables 3-4 display the simulated powers of the tests (3.14), (4.22), the Monte Carlo means and the standard errors of the estimators (3.17), (4.23).

-Tables 3 and 4 here-

Additionally to Table 3, we point out that for the situation where $n = 75$ and $\nu = 60$, if we use the correct critical value $C = 16.5$ (obtained by simulations) corresponding to significance level $\alpha = 0.05$, then the simulated power of the test is $\widehat{P} = 0.457$. This result confirms existence of practical meaning of the upper bounds for the significance level of the proposed test.

6 An example: Change in fertility

The Israel Central Bureau of Statistics conducted a research aimed at determining whether there was a change in the distribution of the number of children per Israeli woman over the past few decades, and if so, in what year it occurred. The survey, based on a random sample from the 1995 population census of Israel, assumed that the possible change in the distribution may be a result of immigration to Israel. Therefore, all of the women who arrived in Israel after the year in which the change in the distribution occurred would be designated as New Immigrants (*NI*). This designation is relevant for certain sociological applications. Clearly, if there is no change in the distribution then there is no need to define the status *NI*.

Consider the following simplified model. Let

$$x_i = \left[1, i, \frac{1}{m_i}\sum_{j=1}^{m_i} a_{ij}\right]^T,$$

and define the random variables:

$$Y_{ij} = \begin{cases} 1, & \text{if the number of children per woman is } \leq 2 \text{ (by 1995)} \\ 0, & \text{if the number of children per woman is } > 2 \text{ (by 1995)}, \end{cases}$$

where $i = 80, \ldots, 95$ is the year of arrival in to Israel, a_{ij} the age of the woman ij in the year i, $1 \leq j \leq m_i$, m_i represents the number of women who arrived in Israel during year i and appear in the data. It is known that the average number of children per woman in Israel is about two. Table 5 lists the characteristics of observations Y and the explanatory variables x.

-Table 5 here-

Now we consider the model (1.1), where $d = 3$, $j = 1, \ldots, m_i$, $i = 80 \ldots, 95$. Then we test the hypothesis

$$\mathbf{H_0} \; : \; \nu \notin [80, 95], \qquad \text{versus} \quad \mathbf{H_1} \; : \; 80 \leq \nu \leq 95; \; \nu > 79 \text{ is unknown.}$$

There was a prior decision that the definition of the specific status *NI* for women who arrived in Israel up 1980 is irrelevant. Therefore we use the logistic regression model that is commonly accepted at Central Bureau of Statistics in Jerusalem, based on observations related to the years prior to 1980. The model assumes $\beta_0 = [-0.8302, -0.0081, 0.0501]^T$.

By applying (3.13), we have

$$\ln\left(\max_{80 \leq k \leq 95} \Lambda_k^{95}\left(\beta_0, \widehat{\beta}|x_k, \ldots, x_{95}\right)\right) = \ln\left(\Lambda_{89}^{95}\left(\beta_0, \widehat{\beta}|x_{89}, \ldots, x_{95}\right)\right) = 3631.2,$$

where $\widehat{\beta}$ is the MLE of β_1 ($\widehat{\beta}^{(l,n)} = \beta_0,\ n - l \leq 7$). Therefore, based on the test (3.14) and Proposition 3.1 we reject H_0 with a $p-value \leq \exp(-3631)$.

By applying (3.17), we obtain $\widehat{\nu} = 90$ as the estimator of the change point. Thus, we can define the status *NI* for women that came to Israel starting from 1990.

7 Appendix

7.1 Proofs of the Propositions

Proof of Proposition 2.1. In order to find the relation between C and the significance level of the test (2.4) note:

Lemma 7.1 *Under H_0, the sequence $\{\Lambda_k^n(\beta_0, \beta_1 | x_k, \ldots, x_n)\}_{k=1}^n$ is a non-negative reverse martingale with respect to $\{(Y_k, x_k)\}_{k=1}^n$ and*
$E_{H_0}\Lambda_k^n(\beta_0, \beta_1 | x_k, \ldots, x_n) = 1$ for all $k = 1, 2, \ldots, n$.

Proof. Noting that, for all $k = 1, \ldots, n-1$ we have by (2.3)

$$\begin{aligned}
& E_{H_0}\left(\Lambda_k^n(\beta_0, \beta_1 | x_k, \ldots, x_n) \,|(Y_n, x_n), \ldots, (Y_{k+1}, x_{k+1})\right) \\
= \; & \prod_{i=k+1}^{n} \prod_{j=1}^{m_i} \frac{P_{\nu=k}\{Y_{ij}|x_i\}}{P_{H_0}\{Y_{ij}|x_i\}} E_{H_0}\left(\prod_{j=1}^{m_k} \frac{P_{\nu=k}\{Y_{kj}|x_k\}}{P_{H_0}\{Y_{kj}|x_k\}} \Big| \{(Y_r, x_r)\}_{r=k+1}^n \right) \\
= \; & \Lambda_{k+1}^n(\beta_0, \beta_1 | x_{k+1}, \ldots, x_n) \prod_{j=1}^{m_k} E_{H_0} \frac{P_{\nu=k}\{Y_{kj}|x_k\}}{P_{H_0}\{Y_{kj}|x_k\}} \\
= \; & \Lambda_{k+1}^n(\beta_0, \beta_1 | x_{k+1}, \ldots, x_n).
\end{aligned}$$

Therefore

$$E_{H_0}\Lambda_1^n(\beta_0, \beta_1 | x_1, \ldots, x_n) = E_{H_0}\Lambda_2^n(\beta_0, \beta_1 | x_2, \ldots, x_n)$$

$$= \ldots = E_{H_0}\Lambda_n^n(\beta_0, \beta_1 | x_n) = 1.$$

This completes the proof of Lemma 7.1.

Denote

$$M_n \quad = \quad \max_{1\le k\le n} \Lambda_k^n\left(\beta_0, \beta_1 | x_k, \ldots, x_n\right) \tag{7.24}$$

and define the stopping rule $N_n = \min\left\{\tau_n(C), n\right\}$, where for $C > 0$,

$$\begin{aligned}\tau_n(C) \quad &= \quad \inf\left\{l > 1 : \quad \frac{\Lambda_1^n\left(\beta_0, \beta_1 | x_1, \ldots, x_n\right)}{\Lambda_1^{l-1}\left(\beta_0, \beta_1 | x_{l-1}, \ldots, x_n\right)} > C\right\} \\ &= \quad \infty \quad \text{ if no such } l \text{ exists.}\end{aligned}$$

Now, by Lemma 7.1 and Doob's theorem for a nonnegative submartingale we have

$$\begin{aligned}&\alpha = P_{H_0}\left\{M_n > C\right\} = P_{H_0}\left\{\Lambda_{N_n}^n\left(\beta_0, \beta_1 | x_{N_n}, \ldots, x_n\right) > C\right\} \\ &\le \frac{E_{H_0}\Lambda_{N_n}^n\left(\beta_0, \beta_1 | x_{N_n}, \ldots, x_n\right)}{C} = \frac{E_{H_0}\Lambda_n^n\left(\beta_0, \beta_1 | x_n\right)}{C} = \frac{1}{C}.\end{aligned} \tag{7.25}$$

Proof of Proposition 2.2. We define for all $1 \le i \le n$

$$\begin{aligned}\lambda_i \quad &\equiv \quad \ln\left(\prod_{j=1}^{m_i} \frac{P_\nu\{Y_{ij}|x_i\}}{P_{H_0}\{Y_{ij}|x_i\}}\right) - a_i \\ &= \quad \left(x_i^T\beta_1 - x_i^T\beta_0\right) \sum_{j=1}^{m_i}\left(Y_{ij} - \frac{\exp(x_i^T\beta_1)}{1+\exp(x_i^T\beta_1)}\right),\end{aligned} \tag{7.26}$$

where a_i by (2.5). Noting that, by (7.24) and (2.3)

$$\begin{aligned}0 \le 1 - P_\nu\left\{M_n > C\right\} \quad &\le \quad P_\nu\left\{\sum_{i=\nu}^{n}(\lambda_i + a_i) \le \ln(C)\right\} \\ &= \quad P_\nu\left\{\sum_{i=\nu}^{n}(-\lambda_i) \ge \sum_{i=\nu}^{n} a_i - \ln(C)\right\},\end{aligned}$$

where, by (2.5) and (7.26) $E_\nu\lambda_i = 0$ and $|\lambda_i| \le m_i\left|x_i^T\beta_1 - x_i^T\beta_0\right|,\ i \ge \nu$. Therefore, by Hoeffding (1963) we have

$$0 \le 1 - P_\nu\left\{M_n > C\right\} \le \exp\left(-\frac{\left(\sum_{i=\nu}^{n} a_i - \ln(C)\right)^2}{2\sum_{i=\nu}^{n}\left(m_i|x_i^T(\beta_1 - \beta_0)|\right)^2}\right). \tag{7.27}$$

The proof of Proposition 2.2 is now complete.

Proof of Proposition 2.3. Note that

$$
\begin{aligned}
& P_\nu\left\{\widehat{\nu}_n \geq t_1\right\} \\
= & \sum_{k \geq t_1}^n P_\nu\left\{\max_{1 \leq j \leq n} \Lambda_j^n\left(\beta_0, \beta_1 | x_j, \ldots, x_n\right)=\Lambda_k^n\left(\beta_0, \beta_1 | x_k, \ldots, x_n\right)\right\} \\
\leq & \sum_{k \geq t_1}^n P_\nu\left\{\Lambda_\nu^n\left(\beta_0, \beta_1 | x_\nu, \ldots, x_n\right) \leq \Lambda_k^n\left(\beta_0, \beta_1 | x_k, \ldots, x_n\right)\right\} .
\end{aligned}
$$

Hence, by definitions (2.3) of $\Lambda_k^n\left(\beta_0, \beta_1 | x_k, \ldots, x_n\right)$ and (2.5), (7.26) we have

$$
P_\nu\left\{\widehat{\nu}_n \geq t_1\right\} \leq \sum_{k \geq t_1}^n P_\nu\left\{\sum_{i=\nu}^{k-1}\left(-\lambda_i\right) \geq \sum_{i=\nu}^{k-1} a_i\right\} .
$$

Therefore, by applying the same principle as in the proof of Proposition 2.2, we obtain (2.9). In a similar manner for $t_2<\nu$

$$
\begin{aligned}
& P_\nu\left\{\widehat{\nu}_n \leq t_2\right\} \\
= & \sum_{k \leq t_2} P_\nu\left\{\max_{1 \leq j \leq n} \Lambda_j^n\left(\beta_0, \beta_1 | x_j, \ldots, x_n\right)=\Lambda_k^n\left(\beta_0, \beta_1 | x_k, \ldots, x_n\right)\right\} \\
\leq & \sum_{k \leq t_2} P_\nu\left\{\Lambda_\nu^n\left(\beta_0, \beta_1 | x_\nu, \ldots, x_n\right) \leq \Lambda_k^n\left(\beta_0, \beta_1 | x_k, \ldots, x_n\right)\right\} \\
\leq & \sum_{k \leq t_2} P_\nu\left\{\sum_{i=k}^{\nu-1} \lambda_i \geq \sum_{i=k}^{\nu-1}\left(-a_i\right)\right\} \leq \sum_{k \leq t_2} \exp \left(-\frac{\left(\sum_{i=k}^{\nu-1} a_i\right)^2}{2 \sum_{i=k}^{\nu-1}\left(m_i | X_i^T\left(\beta_1-\beta_0\right)|\right)^2}\right),
\end{aligned}
$$

Now, additionally, if condition (2.11) is satisfied then by (2.9), (2.10) and elementary inequalities

$$
\begin{aligned}
E_\nu \widehat{\nu}_n & =\int_0^n P_\nu\{\widehat{\nu}_n>t\} d t \qquad (7.28)\\
& \leq \nu+\int_\nu^n P_\nu\{\widehat{\nu}_n>t\} d t \\
& \geq \nu-\int_0^\nu P_\nu\{\widehat{\nu}_n \leq t\} d t, \\
E_\nu(\widehat{\nu}_n-\nu)^2 & =\int_0^{\max \left(\nu^2,(n-\nu)^2\right)} P_\nu\{(\widehat{\nu}_n-\nu)^2>z\} d z \\
& =\int_0^{\nu^2} P_\nu\{\widehat{\nu}_n \geq(z)^{1 / 2}+\nu\}+P_\nu\{\widehat{\nu}_n \leq-(z)^{1 / 2}+\nu\} d z \\
& \quad +\int_{\nu^2}^{\max \left(\nu^2,(n-\nu)^2\right)} P_\nu\{\widehat{\nu}_n \geq(z)^{1 / 2}+\nu\} d z,
\end{aligned}
$$

it follows that the proof of Proposition 2.3 is complete.

Proof of Proposition 3.1. Inasmuch as for all $k = 1, \ldots, n$ by (3.13)

$$
\begin{aligned}
& E_{H_0}\left(\Lambda_k^n\left(\beta_0, \widehat{\beta}|x_k, \ldots, x_n\right)|(Y_n, x_n), \ldots, (Y_{k+1}, x_{k+1})\right) \\
& = \prod_{i=k+1}^{n}\prod_{j=1}^{m_i}\frac{\frac{\exp(x_i^T\widehat{\beta}^{(i+1,n)}Y_{ij})}{1+\exp(x_i^T\widehat{\beta}^{(i+1,n)})}}{\frac{\exp(x_i^T\beta_0 Y_{ij})}{1+\exp(x_i^T\beta_0)}} E_{H_0}\prod_{j=1}^{m_k}\frac{\frac{\exp(x_k^T\widehat{\beta}^{(k+1,n)}Y_{kj})}{1+\exp(x_k^T\widehat{\beta}^{(k+1,n)})}}{\frac{\exp(x_k^T\beta_0 Y_{kj})}{1+\exp(x_k^T\beta_0)}}\Big|\{(Y_r, x_r)\}_{r=k+1}^n \\
= \quad & \Lambda_{k+1}^n\left(\beta_0, \widehat{\beta}|x_{k+1}, \ldots, x_n\right) \text{ and } E_{H_0}\Lambda_k^n\left(\beta_0, \widehat{\beta}|x_k, \ldots, x_n\right) \\
& = E_{H_0}E_{H_0}\left(\Lambda_k^n\left(\beta_0, \widehat{\beta}|x_k, \ldots, x_n\right)\Big|\{(Y_r, x_r)\}_{r=k+1}^n\right) \\
& = E_{H_0}\Lambda_{k+1}^n\left(\beta_0, \widehat{\beta}|x_k, \ldots, x_n\right) = \\
& E_{H_0}E_{H_0}\left(\Lambda_{k+1}^n\left(\beta_0, \widehat{\beta}|x_k, \ldots, x_n\right)\Big|\{(Y_r, x_r)\}_{r=k+2}^n\right) = \ldots = 1,
\end{aligned}
$$

hence under H_0 the sequence $\left\{\Lambda_k^n\left(\beta_0, \widehat{\beta}|x_k, \ldots, x_n\right)\right\}_{k=1}^n$ is a non-negative reverse martingale with respect to $\{(Y_k, x_k)\}_{k=1}^n$.

Combining these results with considerations similar to (7.25) completes the proof of Proposition 3.1.

Proof of Proposition 3.2. By definitions (3.13), (2.5), (7.26) as well as a Taylor series expansion and some rearranging of terms, the probability of type II error for test (3.14) is bounded

$$
\begin{aligned}
& P_\nu\left\{\max_{1\le k\le n}\Lambda_k^n\left(\beta_0, \widehat{\beta}|x_k, \ldots, x_n\right) \le C\right\} \qquad (7.29)\\
& \le P_\nu\left\{\ln\left(\Lambda_\nu^n\left(\beta_0, \widehat{\beta}|x_i, \ldots, x_n\right)\right) \le \ln(C)\right\} \\
& = P_\nu\left\{\sum_{i=\nu}^{n}\left(\sum_{j=1}^{m_i}x_i^T(\beta_1-\beta_0)Y_{ij} + m_i\ln\left(\frac{1+\exp(x_i^T\beta_0)}{1+\exp(x_i^T\beta_1)}\right)\right)\right. \\
& + \sum_{i=\nu}^{n}\left(\sum_{j=1}^{m_i}x_i^T(\widehat{\beta}^{(i+1,n)}-\beta_1)Y_{ij} + m_i\ln\left(\frac{1+\exp(x_i^T\beta_1)}{1+\exp(x_i^T\widehat{\beta}^{(i+1,n)})}\right)\right)
\end{aligned}
$$

$$\leq \ \ln(C)\Big\} = P_\nu \left\{ \sum_{i=\nu}^{n} \lambda_i + \sum_{i=\nu}^{n} a_i \right.$$

$$+ \sum_{i=\nu}^{n} x_i^T (\widehat{\beta}^{(i+1,n)} - \beta_1) \left(\sum_{j=1}^{m_i} Y_{ij} - m_i \frac{\exp(\theta_i)}{1+\exp(\theta_i)} \right) \leq \ln(C) \Bigg\}$$

$$\leq P_\nu \left\{ \sum_{i=\nu}^{n} \lambda_i + (1-\rho) \sum_{i=\nu}^{n} a_i \leq \ln(C) \right\} + \tau_{\nu,n},$$

where $\theta_i \in (x_i^T \widehat{\beta}^{(i+1,n)}, x_i^T \beta_1)$.

Applying to (7.29), the same principle as in the proof of Proposition 2.2 yields

$$\begin{aligned} 0 \ &\leq \ 1 - P_\nu \left\{ \max_{2 \leq k \leq n} \Lambda_k^n \left(\beta_0, \widehat{\beta} | x_k, \ldots, x_n \right) > C \right\} \\ &\leq \ \exp\left(- \frac{\left((1-\rho) \sum_{i=\nu}^{n} a_i - \ln(C) \right)^2}{2 \sum_{i=\nu}^{n} \left(m_i | x_i^T (\beta_1 - \beta_0) | \right)^2} \right) + \tau_{\nu,n}. \end{aligned}$$

where $\tau_{\nu,n}$ by (3.16), thereby completing the proof.

Proof of Proposition 3.3. For $t_1 > \nu$ by (3.17) we have

$$P_\nu\{\widehat{\nu}_n \geq t_1\} \leq \sum_{t_1 \leq k \leq n} P_\nu \left\{ \prod_{i=k}^{n} \prod_{j=1}^{m_i} \frac{\frac{\exp(x_i^T \widehat{\beta}^{(k,n)} Y_{ij})}{1+\exp(x_i^T \widehat{\beta}^{(k,n)})}}{\frac{\exp(x_i^T \beta_0 Y_{ij})}{1+\exp(x_i^T \beta_0)}} \geq \prod_{i=\nu}^{n} \prod_{j=1}^{m_i} \frac{\frac{\exp(x_i^T \widehat{\beta}^{(\nu,n)} Y_{ij})}{1+\exp(x_i^T \widehat{\beta}^{(\nu,n)})}}{\frac{\exp(x_i^T \beta_0 Y_{ij})}{1+\exp(x_i^T \beta_0)}} \right\}.$$

Therefore, by definition of $\widehat{\beta}^{(\nu,n)}$ in (3.17), a Taylor series expansion and some rearranging of terms, we have

$$\begin{aligned} P_\nu\{\widehat{\nu}_n \geq t_1\} &\leq \sum_{t_1 \leq k \leq n} P_\nu \left\{ \prod_{i=k}^{n} \prod_{j=1}^{m_i} \frac{\frac{\exp(x_i^T \widehat{\beta}^{(k,n)} Y_{ij})}{1+\exp(x_i^T \widehat{\beta}^{(k,n)})}}{\frac{\exp(x_i^T \beta_0 Y_{ij})}{1+\exp(x_i^T \beta_0)}} \geq \prod_{i=\nu}^{n} \prod_{j=1}^{m_i} \frac{\frac{\exp(x_i^T \beta_1 Y_{ij})}{1+\exp(x_i^T \beta_1)}}{\frac{\exp(x_i^T \beta_0 Y_{ij})}{1+\exp(x_i^T \beta_0)}} \right\} \\ &\leq \sum_{t_1 \leq k \leq n} P_\nu \left\{ \sum_{i=\nu}^{k-1} (-\lambda_i) \geq \sum_{i=\nu}^{k-1} a_i - 2 \sum_{i=\nu}^{k-1} m_i | x_i^T (\widehat{\beta}^{(k,n)} - \beta_1) | \right\}, \end{aligned}$$

where a_i by (2.5), λ_i by (7.26).

By applying the same principle as in the proofs of Propositions 2.2, 2.3 and (3.18)-(3.20) we obtain

$$P_\nu\{\widehat{\nu}_n \geq t_1\} \leq \frac{2c}{|t_1 - \nu|^p}. \tag{7.30}$$

In a similar manner

$$P_\nu\{\widehat{\nu}_n \le t_2\} \le \frac{2c}{|t_2-\nu|^p}. \tag{7.31}$$

By (7.30), (7.31) and elementary inequalities (7.28), it follows that the proof of Proposition 3.3 is complete.

Proof of Proposition 4.1. Without loss of generality, we assume that $L=1$ and for all $1 \le i \le n$: $m_i = 1$. Noting that by definition of $\widehat{\beta}^{(1,n)}$ in (4.21) we have

$$\Lambda_k^n\left(\widehat{\beta}|X_k,\ldots,X_n\right) \le \prod_{i=1}^{k-1} \frac{f\{Y_{i1}|x_i,\widehat{\beta}^{(1,i-1)}\}}{f\{Y_{i1}|x_i,\beta_0\}} \prod_{i=k}^{n} \frac{f\{Y_{i1}|x_i,\widehat{\beta}^{(k,i-1)}\}}{f\{Y_{i1}|x_i,\beta_0\}}. \tag{7.32}$$

Now, for all $k=1,\ldots,n$

$$\begin{aligned}
&E_{H_0}\prod_{i=1}^{k-1} \frac{f\{Y_{i1}|x_i,\widehat{\beta}^{(1,i-1)}\}}{f\{Y_{i1}|x_i,\beta_0\}} \prod_{i=k}^{n} \frac{f\{Y_{i1}|x_i,\widehat{\beta}^{(k,i-1)}\}}{f\{Y_{i1}|x_i,\beta_0\}}\\
&= E_{H_0}E_{H_0}\left(\prod_{i=1}^{k-1} \frac{f\{Y_{i1}|x_i,\widehat{\beta}^{(1,i-1)}\}}{f\{Y_{i1}|x_i,\beta_0\}} \prod_{i=k}^{n} \frac{f\{Y_{i1}|x_i,\widehat{\beta}^{(k,i-1)}\}}{f\{Y_{i1}|x_i,\beta_0\}}\Big| \{(Y_r,x_r)\}_{r=1}^{n-1}\right)\\
&= E_{H_0}\prod_{i=1}^{k-1} \frac{f\{Y_{i1}|x_i,\widehat{\beta}^{(1,i-1)}\}}{f\{Y_{i1}|x_i,\beta_0\}} \prod_{i=k}^{n-1} \frac{f\{Y_{i1}|x_i,\widehat{\beta}^{(k,i-1)}\}}{f\{Y_{i1}|x_i,\beta_0\}}\\
&\quad \times E_{H_0}\left(\frac{f\{Y_{n1}|x_n,\widehat{\beta}^{(k,n-1)}\}}{f\{Y_{n1}|x_n,\beta_0\}}\Big| \{(Y_r,x_r)\}_{r=1}^{n-1}\right)\\
&= E_{H_0}\prod_{i=1}^{k-1} \frac{f\{Y_{i1}|x_i,\widehat{\beta}^{(1,i-1)}\}}{f\{Y_{i1}|x_i,\beta_0\}} \prod_{i=k}^{n-1} \frac{f\{Y_{i1}|x_i,\widehat{\beta}^{(k,i-1)}\}}{f\{Y_{i1}|x_i,\beta_0\}} = \ldots = 1.
\end{aligned}$$

Therefore, by (7.32) $E_{H_0}\Lambda_k^n\left(\widehat{\beta}|x_k,\ldots,x_n\right) \le 1$ and by applying Markov's inequality to α, we complete the proof of Proposition 4.1.

Acknowledgments. The authors are indebted to Professor Moshe Pollak for many helpful discussions and comments. We thank the two referees and Professor Marie Hušková for some helpful comments.

References

Borovkov, A. A. (1998). Asymptotically optimal solutions in the change-point problem. *Theory Probab. Appl.* **43** 539-561.

Brown, R. L., Durbin, J. and Evans, J. M. (1975). Techniques for testing the constancy of regression relationships over time (with discussion). *J. R. Statist. Soc. B* **37** 149-192.

Gombay, E. and Horvath, L. (1994). An application of the maximum likelihood test to the change-point problem. *Stochastic Processes and their Applications.* **50** 161-171.

Hoeffding, W. (1963). Probability inequalities for sums of bounded random variables. *J. Amer. Statist. Assoc.* **58** 13-30.

Gossl, C. and Kuchenhoff, H. (2001). Bayesian analysis of logistic regression with an unknown change point and covariate measurement error. *Statistics in Medicine.* **20** 3109-3121.

Kim, Hyune-Ju and Siegmund, D. (1989). The likelihood ratio test for a change-point in simple linear regression. *Biometrika.* **76** 409-423.

Lorden, G. and Pollak, M. (2004). Non-Anticipating Estimation Applied to Sequential Analysis and Changepoint Detection. *Under revision for The Annals of Statistics.*

Page, E. S. (1954). Continuous inspectation schemes. *Biometrika.* **41** 100-114.

Pastor, R. and Guallar, E. (1998). Use of two-segmented logistic regression to estimate change-point in epidemiologic studies. *Am. J. Epidemiol.* **148** 631-642.

Pettitt, A. N. (1980). A simple cumulative sum type statistic for the change-point with zero-one observations. *Biometrika.* **67** 79-84.

Quandt, R. E. (1960). Tests of the hypothesis that a linear regression system obeying two separate regimes. *J. Amer. Statist. Assoc.* **55** 324-330.

Robbins, H. and Siegmund, D. (1973). *A class of stopping rules for testing parametric hypotheses.* Proc. Sixth Berkeley Symp., Math. Statist. Prob. 4, Univ. of Calif. Press.

Sen, A. and Srivastava, M. S. (1975). On tests for detecting change in mean. *Ann. Statist.* **1** 98-108.

Stefanski, L. A. and Carroll, R. J. (1985). Covariance measurement error in logistic regression. *Ann. Statist.* **13** 1335-1351.

Ulm, K. (1991). A statistical method for assessing a threshold in epidemiological studies. *Statistics in Medicine.* **10** 341-349.

Woodroofe, M. (1982).*Nonlinear Renewal Theory in Sequential Analysis.* SIAM, Philadelphia.

Worsley, K. J. (1983). The power of likelihood ratio and cumulative sum tests for a change in a binomial probability. *Biometrika .* **70** 455-464.

Yakir, B. (1998). On the average run length to false alarm in surveillance problems which possess an invariance structure. *Ann. Statist.* **26** 1198-1214.

Yakir, B., Krieger, A. M. and Pollak, M. (1999). Detecting a change in regression: first-order optimality. *Ann. Statist.* **27** 1896-1913.

Dr. Gregory Gurevich
Faculty of Industrial Engineering
and Management
Technion-Israel Institute of Technology
Haifa 32000, Israel
Phone: +972-4-8292855
E-mail: greggure@tx.technion.ac.il

Dr. Albert Vexler
Central Bureau of Statistics
66, Kanfey Nesharim
Jerusalem 91130, Israel
Phone: +972-2-6592745
Fax: +972-2-6522319
E-mail: msvex@hotmail.com

Table 1: The simulated significance levels $\widehat{\alpha_1}$, $\widehat{\alpha_2}$, $\widehat{\alpha_3}$ of the tests (2.4), (3.14) and ((4.22), where $L = 5$) respectively

$\nu > n$		$n = 75$		$n = 120$		
C	Upper Bound: $1/C$	$\widehat{\alpha_1}$	$\widehat{\alpha_2}$	$\widehat{\alpha_1}$	$\widehat{\alpha_2}$	$\widehat{\alpha_3}$
10	0.1000	0.0783	0.0737	0.0870	0.0862	0.0531
20	0.0500	0.0435	0.0398	0.0460	0.0450	0.0272
30	0.0333	0.0280	0.0265	0.0290	0.0313	0.0171
40	0.0250	0.0199	0.0221	0.0201	0.0196	0.0124
100	0.0100	0.0095	0.0105	0.0097	0.0088	0.0063
150	0.0066	0.0039	0.0051	0.0060	0.0058	0.0035
200	0.0050	0.0048	0.0045	0.0047	0.0046	0.0032

Table 2: The simulations of the power $\widehat{P}$ (test (2.4)), the boundary B= $\exp\left(-\frac{\left(\sum_{i=\nu}^{n} a_i - \ln(C)\right)^2}{2\sum_{i=\nu}^{n}\left(m_i|x_i^T(\beta_1-\beta_0)|\right)^2}\right)$ by (2.6) and the MLE $\widehat{\nu}_n$ by (2.8)

$C = 20$	$\nu = 10$				$\nu = 50$			
n	$\widehat{P}$	1-B	$\overline{\widehat{\nu}_n}$	SE($\widehat{\nu}_n$)	$\widehat{P}$	1-B	$\overline{\widehat{\nu}_n}$	SE($\widehat{\nu}_n$)
75	0.991	0.657	8.71	2.14	0.987	0.275	51.51	2.05
120	0.997	0.803	9.07	0.99	0.989	0.645	50.97	0.75
200	0.998	0.946	10.30	0.73	0.998	0.892	50.30	0.55

Table 3: The simulations of the power $\widehat{P}$ (test (3.14)) and the MLE $\widehat{\nu}_n$ by (3.17)

$C = 20$	$\nu = 10$			$\nu = 50$			$\nu = n - 15$		
n	$\widehat{P}$	$\overline{\widehat{\nu}_n}$	SE($\widehat{\nu}_n$)	$\widehat{P}$	$\overline{\widehat{\nu}_n}$	SE($\widehat{\nu}_n$)	$\widehat{P}$	$\overline{\widehat{\nu}_n}$	SE($\widehat{\nu}_n$)
75	0.920	8.41	3.10	0.913	49.15	2.50	0.450	61.90	5.12
120	0.996	8.83	1.40	0.995	49.50	1.04	0.531	104.59	5.07
200	0.997	10.90	0.95	0.996	50.80	0.87	0.540	182.99	5.21

Table 4: The simulations of the power $\widehat{P}$ (test (4.22), where $L = 5$) and the MLE $\overline{\widehat{\nu}_n}$ by (4.23)

$C = 20$	$\nu = 50$		
n	$\widehat{P}$	$\overline{\widehat{\nu}_n}$	SE($\widehat{\nu}_n$)
120	0.55	48.03	2.01
200	0.87	51.8	1.08

Table 5: Characteristics of the observations used in the example

i	m_i	$\frac{1}{m_i}\sum_{j=1}^{m_i} a_{ij}$	$\frac{1}{m_i}\sum_{j=1}^{m_i} Y_{ij}$	i	m_i	$\frac{1}{m_i}\sum_{j=1}^{m_i} a_{ij}$	$\frac{1}{m_i}\sum_{j=1}^{m_i} Y_{ij}$
80	1357	38.44	0.59	88	834	39.85	0.71
81	783	38.51	0.58	89	1678	39.92	0.80
82	776	38.61	0.59	90	14235	39.98	0.82
83	856	38.95	0.59	91	11956	39.99	0.81
84	1168	39.01	0.62	92	5578	40.01	0.82
85	768	39.06	0.61	93	5300	40.00	0.81
86	598	39.73	0.65	94	5642	40.08	0.83
87	829	39.80	0.72	95	4133	40.43	0.82