Open Science Repository Mathematics

Maina, M. S., Gichuhi, W. A., & Mwaniniki, K. J. (2013). The Power of Likelihood Ratio Test for a Change Point in Binomial Distribution. Open Science Repository Mathematics, Online(open-access), e70081979. doi:10.7392/openaccess.70081979

THE POWER OF LIKELIHOOD RATIO TEST FOR A CHANGE POINT IN BINOMIAL DISTRIBUTION.

June 23, 2013

Mundia S.M.[1],Gichuhi A.W.[2], Kihoro J.M.[2]

Abstract

Statistically, change point is the location or the time point such that observations follow one distribution up to the point and then another afterwards. Change point problems are encountered in our daily life and in disciplines such as economics, finance, medicine, geology, Literature among others. In this paper, the power of the likelihood ratio tests for a change point in binomial observations whose mean is dependent on the explanatory variables was investigated. Artificial neural network technique was used to estimate the conditional means. It was shown through simulation that the power of the test increases as the size of sample. The test was found to have less power when the change point was near the edges than when the change point is at the centre. The test was more likely to detect a change if the magnitude of the change was large.

Keywords change point, likelihood ratio test, binomial distribution, power of a test, artificial neural-network

1 Introduction

We consider a situation where a sequence of independent binomial variables is subject to a change in distribution after an unknown point. Formally, we can describe this situation as follows. $m_1, m_2, \ldots, m_b$ are independent binomial random variables, such that, for a value k, $1 < k \leq b$, m_i are distributed as

$$m_i \sim \begin{cases} B(n_i, p_i(\mathbf{x})), & 1 \leq i \leq k \\ B(n_i, p_i'(\mathbf{x})), & k+1 \leq i \leq b \end{cases} \tag{1}$$

[1]Corresponding author, Department of Actuarial & Statistics, DeKUT

[2]Department of Actuarial & Statistics, JKUAT

where $p_i(\mathbf{x})$ and $p_i'(\mathbf{x})$ are the success probabilities that depend on the explanatory variables $\mathbf{x} = (x_1, \ldots, x_D) \in \Re^D$. Here the assumption made is that there is a single change point at the point k. Previous work, with this type of model, has been directed towards

(i) Estimating the change-point, k

(ii) Testing the hypothesis that no change in distribution has occurred.

Most analytical approaches, developed for dealing with binomial change-point data, assume the parameters $p_i(\mathbf{x})$ and $p_i'(\mathbf{x})$, like k, to be unknown. The assumption in most approaches is that the conditional probabilities do not depend on explanatory variables. Particular attention has also been devoted to the case of the m_i being zero-one variables i.e. with $n_i = 1$ for all i.

Worsley [7] studied the power of the likelihood ratio and cumulative sum tests for the binomial model. He found the exact null and alternative distributions of likelihood ratio, cumulative sum and related statistics for testing for a change in probability of a sequence of independent binomial random variables. Waititu[6] investigated power of likelihood test for change in the bernoulli model where he used the artificial neural networks to estimate the conditional means. In his work the conditional means of the bernoulli random variables are assumed to depend on explanatory variables. In this work artificial neural networks are used to estimate the conditional probabilities and the power of likelihood ratio test of a change point is investigated.

The paper is presented as follows. We define the model in section 2. Artificial neural networks are used to estimate the binomial probabilities. In Section 3 the hypothesis testing problem is given. In Section 4 we show that the test is consistent. Simulated data is used to investigate the power of the test when the change point is put at various locations and how the size of change affects the power of the test in Section 5. In section 6 we have an application to real data.

2 The Model

The observations m_i are independently distributed binomial random variables whose probability distribution may be denoted as

$$f(m_i, p_i(\mathbf{x})) = \begin{pmatrix} n_i \\ m_i \end{pmatrix} [p_i(\mathbf{x})]^{m_i} [1 - p_i(\mathbf{x})]^{n_i - m_i} \tag{2}$$

As the functional form of $p_i(\mathbf{x})$ is not known one may use a parametric method and the logistic regression to estimate $p_i(\mathbf{x})$. Here we obtain as in in Chao *et al*[1]

$$p_i(\mathbf{x}) = \frac{1}{1 + \exp\{-(\beta_0 + \sum_{i=1}^{d} \beta_i x_i\}} \tag{3}$$

/ An alternative would be the use of non-parametric method where the output of a single hidden-layer feedforward neural network with $H \geq 1$ hidden nodes and a single output node is used to approximate $p_i(\mathbf{x})$. The output of the network may be presented as

$$\begin{aligned} \varphi(x;\theta) &= \psi(\zeta(x;\theta)) \\ \zeta(x;\theta) &= \alpha_0 + \sum_{h=1}^{H} \alpha_h \{w_{h0} + \sum_{d=1}^{D} w_{hd} x_d\} \end{aligned} \tag{4}$$

where $\theta \in \Omega = (w_{hj} \quad h = 0,1,\ldots,H \quad j = 0,1,\ldots,D \quad \alpha_0,\alpha_1,\ldots,\alpha_H)$ is the vector of network weights and ψ is the activation function of the network. The unipolar function is used as its output is in the range [0,1] making it appropriate in estimating probabilities. The set Ω is compact to ensure that the set is bounded and closed.

The network is trained so that the *error function*

$$l(\theta) = -\frac{1}{b}\sum_{i=1}^{b}\left\{\ln\begin{pmatrix} n_i \\ m_i \end{pmatrix} + m_i \ln\varphi(x;\theta) + (n_i - m_i)(1 - \ln\varphi(x;\theta))\right\} \tag{5}$$

is minimized. The average of this *error function* is

$$\begin{aligned} l_0(\theta) &= -E\left\{\frac{1}{b}\sum_{i=1}^{b}\left\{\ln\begin{pmatrix} n_i \\ m_i \end{pmatrix} + m_i \ln\varphi(x;\theta) + (n_i - m_i)(1 - \ln\varphi(x;\theta))\right\}\right\} \\ &= -E\left\{\ln\begin{pmatrix} n_1 \\ m_1 \end{pmatrix} + m_1 \ln\varphi(x;\theta) + (n_1 - m_1)(1 - \ln\varphi(x;\theta))\right\} \\ &= -E\left\{\ln\begin{pmatrix} n_1 \\ n_1 p(x_1) \end{pmatrix} + n_1 p(x_1) \ln\varphi(x_1;\theta) + (n_1 - n_1 p(x_1)(1 - \ln\varphi(x_1;\theta))\right\} \end{aligned}$$

Assuming that $l_0(\theta)$ has a unique minimum in $\theta \in \Omega$, then this minimum is characterized by

$$\begin{aligned} \nabla l_0(\theta) &= -n_1 E\left\{\frac{p(x_1)}{\varphi(x_1;\theta)} - \frac{1 - p(x_1)}{1 - \varphi(x_1;\theta)}\right\} \nabla\varphi(x_1;\theta) \\ &= 0 \end{aligned} \tag{6}$$

Here the fact that the neural network output function is continuous in x and θ and continuously differentiable with respect to θ makes it possible to interchange expectation and differentiation.

If the model is correctly specified then $p(x) = \varphi(x;\theta^{'})$ for some $\theta^{'} \in \Omega$ then equation (6) is solved but in a general situation $\theta^{'}$ is defined as

$$\theta^{'} = \arg\min_{\theta\in\Omega} l_0(\theta) \tag{7}$$

An estimator $\hat{\theta}$ of θ is the value θ that minimises the *error function* in equation (5). $\hat{\theta}$ is consistent if $\hat{\theta} \to \theta^{'}$ as $b \to \infty$.

In the context of classical regression models, our model may expressed as

$$m_i = n_i p(x_i) + \epsilon_i \quad i = 1, \ldots, b \tag{8}$$

As the observations (m_i, x_i) are independent and $P(m_i|x_i) = \frac{1}{n_i} E(m_i|x_i)$ then we have that $E(\epsilon_i) = 0$ and $var(\epsilon_i) = \sigma^2_\epsilon < \infty$

The consistency and asymptotic normality may be inferred from Franke and Neumann [2]

3 Testing for the change point

The change point hypothesis problem will be stated as

$H_0 : p_i(x) = p_0(x) \qquad 1 \leq i \leq b$

against

$H_a : p_i(x) = p_0(x) \qquad$ for some $i \leq k$, and for some $i > k$, $p_i(x) = p'(x)$

where $2 \leq k \leq b-1$ is the unknown change point location and $p_0(x) \neq p'(x)$.

The general likelihood function is of the form

$$L(m, x, p) = \prod_{i=1}^{b} \begin{pmatrix} n_i \\ m_i \end{pmatrix} [p_i(x)]^{m_i} [1 - p_i(x)]^{n_i - m_i} \tag{9}$$

Thus if k is not fixed and its location is unknown then H_0 is rejected if and only if

$$Q_b = \max_{1<k\leq b-1} -2 \log \Lambda_k \geq C \tag{10}$$

where Λ_k is the ratio of the likelihoods of the sample after and before the change. The critical values, C for the corresponding sample size b and level of significance of the test are computed using Theorems 2.1 and 3.1 in Gombay and Horvarth[3]. These values are presented in the Appendix.

4 Power of the test

The likelihood ratio statistic is $Q_b = \max_{1<k\leq b-1} -2 \log \Lambda_k$ where $\Lambda_k = \frac{L(\hat{\Omega}_o)}{L(\hat{\Omega}_a)}$. $\hat{\Omega}_o$ contains $\hat{\theta}_o$, the maximum likelihood estimate of θ under the null hypothesis while $\hat{\Omega}_a$ contains $\hat{\theta}_k, \hat{\theta}_{k+1}$ the maximum likelihood estimate of θ under the alternative hypothesis before and after the change point respectively. Q_b is also an increasing function of $\max_{1<k\leq b-1} \dfrac{1}{\Lambda_k}$ and therefore the null hypothesis is rejected if Q_b is large, i.e. reject H_o if $Q_b > C$ where C is some bound that depends on the size α of the test and the size b of the sample. If $P_\theta(m_i|x_i)$ is the conditional probability of $m_i = m$ given that $x_i = x$ provided that θ is the true parameter then,

$$\Lambda_k = \prod_{i=1}^{k} \frac{P_{\hat{\theta}_0}(m_i|x_i)}{P_{\hat{\theta}_k}(m_i|x_i)} \prod_{i=k+1}^{b} \frac{P_{\hat{\theta}_0}(m_i|x_i)}{P_{\hat{\theta}^*_k}(m_i|x_i)} \tag{11}$$

where $\hat{\theta}_0 \in \hat{\Omega}_o$ and $\hat{\theta}_k, \hat{\theta}_k^* \in \hat{\Omega}_a$

From Theorem 2.1 of Gombay and Horvath [3] it is observed that C grows asymptotically as b and for a given x depending on the size of the test so that,

$$\begin{aligned} Q_b &= \frac{(x + f(\log b))^2}{a^2(\log b)} \\ &\approx 2\log b \end{aligned} \tag{12}$$

To argue that this test is consistent, we show that for a given size α its power converges to 1.

If there is change, then it occurs at a certain point in the data. That is there is a change point k, $2 \leq k \leq b-1$ and as $b \to \infty$,then we have $k, b-k \to \infty$, $\frac{k}{b} = \iota \in (0,1)$.

Let θ_ι, θ_ι^* be the parameter values before and after the change point respectively and θ_0 denote the parameter value under the null hypothesis.

Since the estimator $\hat{\theta}$ is consistent then as $b \to \infty$

$$\hat{\theta}_0 \to \theta_0 \qquad \hat{\theta}_k \to \theta_\iota \qquad \hat{\theta}_k^* \to \theta_\iota^*$$

So that asymptotically by the law of large numbers

$$\frac{1}{b}\log\Lambda_k \sim \iota E_{\theta_\iota} \log \frac{P_{\hat{\theta}_0}(m_i|x_i)}{P_{\hat{\theta}_\iota}(m_i|x_i)} + (1-\iota)E_{\theta_\iota^*} \log \frac{P_{\hat{\theta}_0}(m_i|x_i)}{P_{\hat{\theta}_\iota^*}(m_i|x_i)} \tag{13}$$

Under the alternative hypothesis then $\theta_\iota \neq \theta_\iota^*$ and $\theta_0 \neq \theta_\iota^*, \theta_0 \neq \theta_\iota$ by the definition of θ_0. In a correctly specified model and assuming that θ is identifiable (see Hwang and Ding [4] for assumptions in identifiability) then, $P_{\theta_0} \neq P_{\theta_\iota}$, $P_{\theta_0} \neq P_{\theta_\iota}{}^*$.

From Jeans's inequality and the fact the logarithm is a strictly concave function we have that

$$\begin{aligned} E_{\theta_\iota} \log \frac{P_{\hat{\theta}_0}(m_i|x_i)}{P_{\hat{\theta}_\iota}(m_i|x_i)} &< \log E_{\theta_\iota} \frac{P_{\hat{\theta}_0}(m_i|x_i)}{P_{\hat{\theta}_\iota}(m_i|x_i)} \\ &= \log \int\int \frac{P_{\hat{\theta}_0}(m|x)}{P_{\hat{\theta}_\iota}(m|x)} P_{\hat{\theta}_\iota}(m|x) d\nu(x) d\mu(x) \\ &= \log \int\int P_{\hat{\theta}_0}(m|x) d\nu(x) d\mu(x) \\ &= 0 \end{aligned}$$

similar results are obtained for the last term of equation (13). Hence for some constant $\gamma > 0$, $\frac{1}{b}\log\Lambda_k \sim -\gamma$. Thus $\log(\Lambda_k)^{-1} \sim b\gamma$. The size of type II error which depends on the power of the test under the alternative vanishes since

$$P(\max_{2\leq k\leq b-1}(\Lambda_k)^{-1} \leq C|H_a) \leq P((\Lambda_k)^{-1} \leq C|H_a) \to 0 \quad \text{as} \quad b \to \infty \tag{14}$$

as $(\Lambda_k)^{-1}$ changes as $e^{b\gamma}$ and C changes only as b. Thus the asymptotic power of the test is unity.

5 Simulation Studies

The power of a change point test for finite sample size for specific alternatives of one change point is investigated .

The null hypothesis is rejected if the test statistic is large i.e. $Q_b^{0.5} > C$ where C is the asymptotic critical value which depends on the size of the test α and the size b of the sample is obtained using either Theorem 2.1 or 3.1 in Gombay and Horvath [3]

For a given level α the power of the test for a specific alternative is the probability of accepting this alternative correctly which is presented as

$$\kappa(\alpha) = P(Q_b^{0.5} > C|H_a) \tag{15}$$

Since the distribution of $Q_b^{0.5}$ under H_a is not known simulations are used to estimate the power of the test as follows:-

For a sample size b, B replicates are made and in each replicate $Q_b^{0.5}$ is estimated. Then the power at α is estimated as

$$\hat{\kappa}(\alpha) = \frac{1 + n\underline{o}(Q_b^{0.5} > C_b(\alpha))}{1+B} \tag{16}$$

where $n\underline{o}(Q_b^{0.5} > C_b(\alpha))$ is the number of times $Q_b^{0.5} > C_b(\alpha)$.

For simulation purposes, we will assume the $P(m_i|X_i = x) = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i}$. Under H_a, following model was used

$$P(m_i|X_i = x) = \begin{cases} (1 + exp(-(-1.5 + x_{1i} + x_{2i})))^{-1} & 1 \leq i \leq k \\ (1 + exp(-(-1.5 + 2x_{1i} + 1.8x_{2i})))^{-1} & k+1 \leq i \leq b \end{cases} \tag{17}$$

where we arbitrarily pick the values of β_0, β_1 and β_2 as -1.5, 1 and 1 $1 \leq i \leq k$. Similarly β_0, β_1 and β_2 as -1.5, 2 and 1.8 $k+1 \leq i \leq b$.

For a sample of size b=200, x_{1i} and x_{2i} were generated as $uniform[0, 1]$. n_i, the size i^{th} group was generated as the whole part of *uniform*[2, b]. The location of the change point k was placed at 20,40,50,100,150,160 and 180. Then the binomial random variable m_i is generated in line with equation (17). 500 simulations were done at each of the change point location. The value of the test statistic $Q_b^{0.5}$ in each of the 500 simulations and using the critical values $C1$ and $C2$, which are generated using Theorem 2.1 or 3.1 in Gombay and Horvath [3] the power of the test is estimated using equation (16). The results are presented in Tables 1 and 2. A plot of the power of the test against the location of change point at $\alpha = 0.01$ is presented in Figure 1.

The change point k was then put at $\frac{b}{4}$, $\frac{b}{2}$ and $\frac{3b}{4}$ for the samples sizes 50,100,150,200 and 500. For each sample, the power of the test at each change point location evaluated in 500 simulations. The results are presented in Tables 3, 4 and 5 A plot of the power of the test against the size of the sample at $\alpha = 0.01$ is presented in Figure 2.

500 simulations were carried out to investigate the power of the test for a sample size of 200 in relation to the size of the change, denoted as $\triangle$ where,

$$\triangle^2 = \parallel \theta - \theta^* \parallel^2 \tag{18}$$

and change point location. To compute the power of the test we used the critical values, C1. The results represented in the Table 6. A plot of the power of the test against the location of the change point at $\alpha = 0.01$ for the changes of size 1.2, 1.5 and 1.8 is presented in Figure 3

	$\hat{\kappa}(\alpha)$						
α	Change points location						
	20	40	50	100	150	160	180
0.01	0.003992	0.4411	0.9142	1	0.9800	0.7625	0.02994
0.05	0.8323	1	1	1	1	1	0.9661
0.10	1	1	1	1	1	1	1

Table 1: *Power of the likelihood ratio test of a sample size b= 200. 500 simulations were done to determine each estimate and critical values C1 were used.*

	$\hat{\kappa}(\alpha)$						
α	Change points location						
	20	40	50	100	150	160	180
0.01	0.05389	0.9980	1	1	1	1	0.3094
0.05	0.8762	1	1	1	1	1	0.9741
0.10	1	1	1	1	1	1	1

Table 2: *Power of the likelihood ratio test of a sample of 200. 500 simulations were done to determine each estimate and critical values C2 were used.*

	$\hat{\kappa}(\alpha)$				
α	Sample size				
	50	100	150	200	500
0.01	0.005988024	0.001996008	0.005988024	0.9121756	1
0.05	0.001996008	0.02794411	0.998004	1	1
0.10	0.01596806	0.7325349	1	1	1

Table 3: *Power of the likelihood ratio test when the change point is at* $\frac{b}{4}$*. 500 simulations were done to determine each estimate and critical values C1 were used.*

	$\hat{\kappa}(\alpha)$				
α	Sample size				
	50	100	150	200	500
0.01	0.003992016	0.001996008	0.0259481	0.9780439	1
0.05	0.003992016	0.0998004	1	1	1
0.10	0.02794411	0.8742515	1	1	1

Table 4: *Power of the likelihood ratio test when the change point is at* $\frac{b}{2}$*. 500 simulations were done to determine each estimate and critical values C1 were used.*

	$\hat{\kappa}(\alpha)$				
α	Sample size				
	50	100	150	200	500
0.01	0.001996008	0.001996008	0.003992016	0.1836327	1
0.05	0.001996008	0.3812375	1	1	1
0.10	0.06586826	0.998004	1	1	1

Table 5: *Power of the likelihood ratio test when the change point is at* $\frac{3b}{4}$. *500 simulations were done to determine each estimate and critical values C1 were used.*

	$\hat{\kappa}(\alpha)$ under C1			
		size of change		
k	α	$\triangle = 1.2$	$\triangle = 1.5$	$\triangle = 1.8$
20	0.01	0.003992016	0.001996008	0.003992
	0.05	0.06387226	0.4530938	0.8323
	0.1	0.8023952	1	1
40	0.01	0.003992016	0.05588822	0.4411
	0.05	0.8163673	1	1
	0.1	1	1	1
50	0.01	0.007984032	0.2315369	0.9142
	0.05	0.9520958	1	1
	0.1	1	1	1
100	0.01	0.1197605	0.9121756	1
	0.05	1	1	1
	0.1	1	1	1
150	0.01	0.01796407	0.3313373	0.9800
	0.05	0.9820359	1	1
	0.1	1	1	1
160	0.01	0.003992016	0.1157685	0.7625
	0.05	0.8582834	1	1
	0.1	1	1	1
180	0.01	0.001996008	0.003992016	0.02994
	0.05	0.1077844	0.499002	0.9661
	0.1	0.8622754	0.998004	1

Table 6: *Power of the likelihood ratio test for different sizes of change and change point locations k. A sample size $b = 200$ and 500 simulations were done for each corresponding case.*

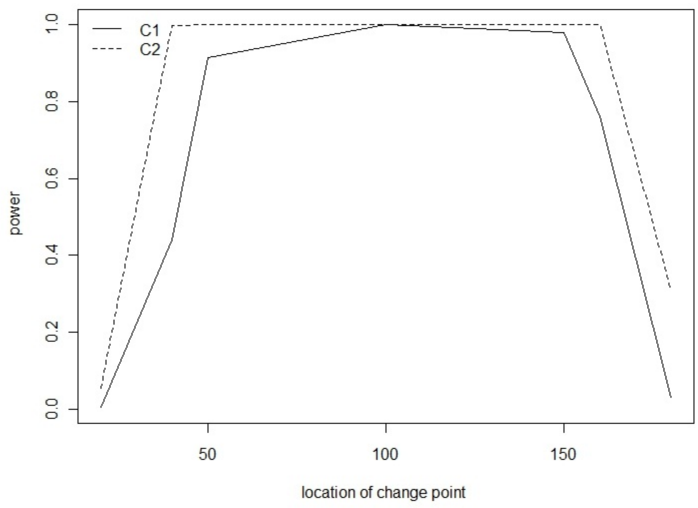

Figure 1: *A plot of the power of the test against the location of change point*

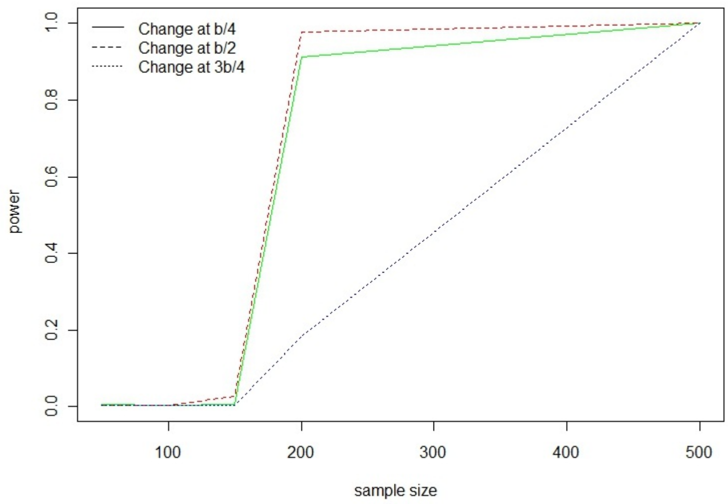

Figure 2: *A plot of the power of the test against the size of the sample at* $\alpha = 0.01$

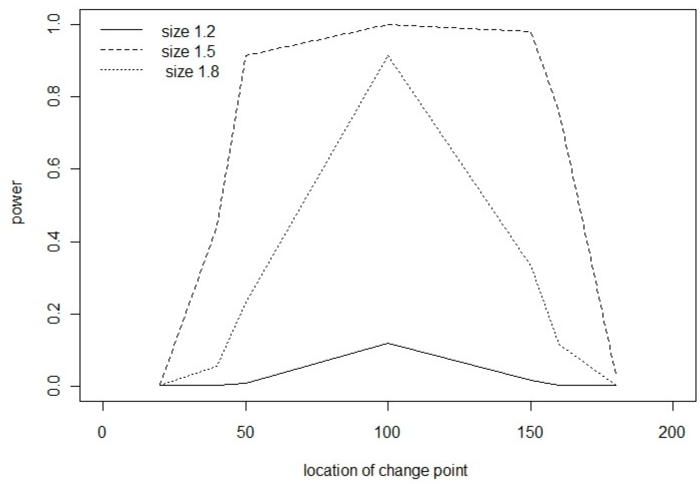

Figure 3: *A plot of the power of the test against the location of the change point at* $\alpha = 0.01$ *for the changes of size 1.2, 1.5 and 1.8*

6 Results and Discussions

Results in Table 1 and Table 2 show that the power of the test is less when the change point is located near the edges of the data. The differences in the power as indicated in Figure 1 could be because the critical regions used in 1 are in a square of Gumbel distribution an extreme value distribution which has a slow rate of convergence as noted in Gombay and Horvath [3]. When the location is in the upper edges the test has more power compared with the power at the lower edges. The test has more power when the change location is at the centre of the data i.e. the test will most probably detect a change when the change point is at the centre. This is due to the comparison of an estimate calculated using a relatively small number of observation, the first k and an estimate calculated in a large number of observations, the last b-k observations. This is as noted by Jaruskova [5]. Table 3 and Table 4 indicate that an increase in the sample size increases the power of the test, as expected. As Figure 2 shows the loss of power is more due to the size of the sample rather than the location of the change point. Table 5 This is of importance since it would be desirable to detect a change once it occurs.

7 Conclusion

This paper proposes the use of artificial neural network in the estimation of the conditional binomial probabilities and then use the likelihood ratio test to check for change point. Simulations studies show

that the power of the test depends on the size(a large change has a high chance of being detected), the location of the change point(a change near the center of the data is more likely to be detected than a change near the edges of the data) and the size of the sample(as the sample size increases a change is more likely to be detected)under consideration.
The simulation and data analysis programs in R are available from the first author.

Acknowledgments: The authors wish to thank the Editor and referees for their valuable comments and suggestions that helped to improve the presentation of the paper.

Appendix

Sample size	α	C1	C2
50	0.01	5.154013	4.787015
	0.05	4.167178	4.306045
	0.1	3.3731367	4.063449
100	0.01	5.219244	4.854494
	0.05	4.286601	4.385838
	0.1	3.874723	4.151836
150	0.01	5.249661	4.887406
	0.05	4.341763	4.42462
	0.1	3.940813	4.194628
200	0.01	5.268792	4.908558
	0.05	4.467199	4.449472
	0.1	3.982043	4.22199
500	0.01	5.319912	4.966611
	0.05	4.167178	4.517474
	0.1	4.09062	4.296645

Table 7: *The critical values $C1$ and $C2$, generated using Theorem 2.1 and 3.1 respectively in Gombay and Horvath [3]*

References

[1] Chao-Ying, J., and Gary, M. An introduction to logistic regression analysis and reporting. *The Journal of Educational Research 96*, 1 (2002), 3–5.

[2] Franke, J. and Neumann, M. Bootstrapping neural networks. *Neural Computation 12*, 12 (2000), 1929–1949.

[3] Gombay, E. and Horvath, L. On the rate of approximation for maximum likelihood tests in change-points models. *Journal of Multivariate Analysis*, 56 (1996), 120–152.

[4] Hwang, J.T. and Ding, A. Prediction intervals for artificial neural network. *Journal of the American Statistical Association 92*, 438 (1997), 748–757.

[5] Jaruskova, D. Some problems with application of change point detection methods to enviromental data. *Envirometrics 8* (1997), 469–483.

[6] Waititu, A. *Nonparametric change point analysis for Bernoulli random variables based on neural networks.* PhD thesis, 2008.

[7] Worsley, K. The power of the likelihood ratio and cumulative sum tests for change in binomial probability. *Biometrika*, 70 (1983), 455–464.