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

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

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

[2]Department of Statistics & Actuarial, Jomo Kenyatta University of Agriculture and Technology

E-mail:

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 among others. In this paper, the power of the likelihood ratio tests for a change point in binomial observations whose mean is dependent on explanatory variables is investigated. Artificial neural network technique is used to estimate the conditional means. These estimates are compared with ones obtained using the generalized link functions.

It is shown through simulation that the power of the test increases as the size of sample. The test is found to have less power when the change point is near the edges than when the change point is at the centre. The test is also more likely to detect a change if the magnitude of the change is large. In all the instances, the neural network method is found to perform better than the parametric method.

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

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: - $m_1,\dots,m_b$ are independent binomial random variables, such that, for a value *k*, $1 < k < b$, m_i are distributed as

$$m_i = \begin{cases} B(n_i, p_i) & 1 \le i \le k \\ B(n_i, p_i') & k+1 \le i \le b \end{cases} \tag{1}$$

---

where $p_i(x)$ and $p_i'(x)$ are the success probabilities that depend on the explanatory variables $x = (x_1,\dots,x_D) \in \Re^D$. 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 $p_i(x)$ and $p_i'(x)$ 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 (1983) 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 (2008) investigated power of likelihood test for change in the bernoulli model and 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 and parametric methods are used to estimate the conditional probabilities and the power of likelihood ratio

test of a change point is investigated. The results of the two methods are then compared.

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 discussed. In Section 4 we show that the test is consistent. In Section 5, simulated data is used to investigate the power of the
test for the change point at various locations within the data and the effect of the size of change on the power of the test is also investigated. In section 6 we give an application to real data where we compare
the estimates of conditional means obtained through the use of the generalized link function and those obtained through the neural network.

2 The Model

The observations mi are independently distributed binomial random variables whose probability distributions may be denoted as

$$f(m_i, p_i(x)) = {}_{n_i}C_{m_i} \left[p_i(x)\right]^{m_i} \left[1 - p_i(x)\right]^{n_i - m_i} \qquad (2)$$

As the functional form of $p_i(x)$ is not known one may use a parametric method and the logistic regression to estimate $p_i(x)$. Here we obtain as in Chao-Ying and Gary (2002)

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

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 $p_i(x)$. The output of the network may be presented as

$$\varphi(x,\theta) = \psi(\zeta(x,\theta))$$
$$\zeta(x,\theta) = \alpha_0 + \sum_{h=o}^{H} \alpha_h \left\{ wh_0 + \sum_{d=1}^{D} w_{hd} x_d \right\} \qquad (4)$$

where $\theta \in \Omega = (w_{hj}, \alpha_h \quad h = 0,1,\cdots,H \quad j = 0,1,\cdots,D)$ 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 it is bounded and closed.

The network is trained so that the *error function*

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

is minimized. The average of this error function is

$$l_0(\theta) = -E\left\{ \frac{1}{b}\sum_{i=1}^{b}\left\{ \ln\binom{n_i}{m_i} + m_i \ln\varphi(x,\theta) + (n_i - m_i)(1 - \ln\varphi(x,\theta)) \right\} \right\}$$

$$= -E\left\{ \sum_{i=1}^{b}\left\{ \ln\binom{n_1}{m_1} + m_1 \ln\varphi(x_1,\theta) + (n_1 - m_1)(1 - \ln\varphi(x_1,\theta)) \right\} \right\}$$

$$= -E\left\{ \sum_{i=1}^{b}\left\{ \ln\binom{n_1}{n_1 p(x_1)} + n_1 p(x_1) \ln\varphi(x_1,\theta) + (n_1 - n_1 p(x_1))(1 - \ln\varphi(x_1,\theta)) \right\} \right\}$$

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

$$\nabla l_0(\theta) = -n_1 E\left\{ \frac{p(x_1)}{\varphi(x_1,\theta)} - \frac{1 - p(x_1)}{1 - \varphi(x_1,\theta)} \right\} \nabla\varphi(x_1,\theta) = 0 \tag{6}$$

where $\nabla l_0(\theta) = \frac{\partial}{\partial\theta} l_0(\theta)$

Here the fact that the neural network output function is continuous in x and θ and is 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 is θ' defined as

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

An estimator $\hat{\theta}$ of θ is the value θ that minimizes 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 be expressed as

$$m_i = n_i p(x_i) - \varepsilon_i \tag{8}$$

As the observations (m_i, x_i) are independent and $P(m_i \mid x_i) = \frac{1}{n_i} E(m_i \mid x_i)$ then we have that $E(\varepsilon_i) = 0$ and $\mathrm{var}(\varepsilon_i) = \sigma_\varepsilon^2 < \infty$

The consistency and asymptotic normality of network parameter estimates may be inferred from Franke and Neumann (2000).

3.0 Testing for the change point

The change point hypothesis problem will be stated as

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

Against

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

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

The general likelihood function is of the form

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

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

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

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 Horvath (1996). These values are presented in the Appendix.

4.0 Power of the test

The likelihood ratio statistic is $Q_b = \max_{1<k\le b-1} -2\log\Lambda_k$ where $\Lambda_k = \frac{L(\hat{\Omega}_0)}{L(\hat{\Omega}_1)}$. $\hat{\Omega}_0$ contains, the $\hat{\theta}_0$ maximum likelihood estimate of θ under the null hypothesis while

$\hat{\Omega}_1$ 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 $\max_{1<k\le b-1} \frac{1}{\Lambda_k}$ and therefore the null hypothesis is rejected if Q_b is large, i.e. reject H_0 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 $P(m_i \mid x_i)$ is the conditional probability of $m_i = m$ given that $x_i = x$ provided that θ is the true parameter then,

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

Where $\hat{\theta}_0 \in \hat{\Omega}_0$ and $\hat{\theta}_k, \hat{\theta}_{k+1} \in \hat{\Omega}_a$

This conditional probability may be estimated using a parametric method or a non-parametric method.

A commonly used parametric method is the generalized link function model. We propose the use of the neural network, a non-parametric method and compare the results of the two methods.

From Theorem 2.1 of Gombay and Horvath (1996) 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. Thus for a change point k, $2 \le k \le b-1$ and as $b \to \infty$, then we have that $k, b-k \to \infty$, $\frac{k}{b} = \iota \in (0,1)$

Let $\theta_\iota, \theta_\iota^*$ 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 $b \to \infty$

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

So that asymptotically by the law of large numbers

$$\frac{1}{b}\log \Lambda_k \approx \iota E_{\theta_\iota} \frac{P_{\hat{\theta}_o}(m_i \mid x_i)}{P_{\hat{\theta}_k}(m_i \mid x_i)} + (1-\iota) E_{\theta^*_\iota} \frac{P_{\hat{\theta}_o}(m_i \mid x_i)}{P_{\hat{\theta}^*_k}(m_i \mid x_i)} \tag{13}$$

Under the alternative hypothesis then $\theta_l \neq \theta^*_\iota$ 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 (1997) for assumptions in identifiability) then, $P_{\theta_0} \neq P_{\theta_\iota}$ .and $P_{\theta_0} \neq P_{\theta^*_\iota}$

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

$$E_{\theta_\iota} \frac{P_{\hat{\theta}_o}(m_i \mid x_i)}{P_{\hat{\theta}_k}(m_i \mid x_i)} < \log E_{\theta_\iota} \frac{P_{\hat{\theta}_o}(m_i \mid x_i)}{P_{\hat{\theta}_k}(m_i \mid x_i)}$$

$$= \log \iint \frac{P_{\hat{\theta}_o}(m_i \mid x_i)}{P_{\hat{\theta}_k}(m_i \mid x_i)} P_{\hat{\theta}_k}(m_i \mid x_i) d\nu(x) d\mu(x)$$

$$= \log \iint P_{\hat{\theta}_0}(m_i \mid x_i) d\nu(x) d\mu(x)$$

$$= 0$$

Similar results are obtained for the last term of equation (13). Hence for some constant $\gamma > 0$,

$\frac{1}{b}\log \Lambda_k \approx -\gamma$. Thus $\log \Lambda_k \approx -b\gamma$. The size of type II error which depends on the power of the test under the alternative vanishes since

$$P(\max_{1<k\leq b-1} (\Lambda_k)^{-1} \leq C \mid H_a) \leq P((\Lambda_k)^{-1} \leq C \mid H_a) \to 0 \quad \text{as} \quad b \to \infty$$

(14)

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

5.0 Simulation Studies

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

The null hypothesis was rejected if the test statistic was 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 (1996).

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

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

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

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

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

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

Our model was assumed to be of the form $p_i(m_i \mid X_i = x) = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i}$ and using the logistic regression as in equation (3) we have that

$$P(m_i \mid X = x_i) = \frac{1}{1 + \exp-(\beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i})}$$

(17)

For simulation purposes, we assumed that H_a is true and used the following model

$$P(m_i \mid X = x_i) = \begin{cases} (1 + \exp(-(-1.5 + x_{1i} + x_{2i})))^{-1} & 1 \le i \le k \\ (1 + \exp(-(-1.5 + 2x_{1i} + 1.8x_{2i})))^{-1} & k + 1 \le i < b \end{cases}$$

(18)

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

For a sample of size $b = 200$, x_{1i} and x_{2i} were generated as *Uniform* [0, 1]. n_i the size of the 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 was generated in line with equation (18). 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 was computed first using estimates of parameters from a generalized link function and then using a neural network. Using the critical values $C1$ and $C2$ which were generated using Theorem 2.1 or 3.1 in Gombay and Horvath (1996) the power of the test was estimated using equation (16). The results are presented in Tables 1 and 2 respectively. 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 was evaluated. 500 simulations were done to determine each estimate and critical values $C1$ were used. 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 further 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 Δ where,

$$\Delta^2 = \| \theta - \theta^* \| \tag{19}$$

and change point location. To compute the power of the test we used the critical values, $C1$. The results are presented 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.

6.0 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. For each value of $\alpha = 0.01$ the upper row shows the power when the parameters were estimated using a parametric method while in the lower row the parameters were estimated using a neural network. The differences in the power as indicated in Figure 1 could be due to the fact that the critical values $C1$ are in a squared Gumbel distribution, an extreme value distribution with a slow rate of convergence as noted in Gombay and Horvath (1996). The values in the Figure 1(b) were estimated using a parametric method while the values in the Figure 1(a) were estimated using a neural network.

When the change point is located 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 (1997). Table3 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. This is of importance since it would be desirable to detect a change once it occurs. The values in Figure 2(a) were evaluated when the conditional probabilities were estimated using a parametric method while those in Figure 2(b) the probabilities were estimated using a neural network.

Figure 3 shows that as the size of the change increases the more the chance of detecting it. The values in

Figure 3(a) were evaluated when the conditional probabilities were estimated using a parametric method while those in Figure 3(b) the probabilities were estimated using a neural. It is noted that in all the instances the neural network performs better than the parametric method.

Table 1: Power of the likelihood ratio test from a sample size $b = 200$ = 200 using critical values $C1$

				$\hat{\kappa}(\alpha)$			
				Change point location			
	20	40	50	100	150	160	180
α							
0.01	0.003992	0.4411	0.9901	1	0.9901	0.8705	0.03237
	0.0099	0.5941	1	1	1	1	0.9661
0.05	0.8323	1	1	1	1	1	0.9661
0.10	1	1	1	1	1	1	1

Table 2: Power of the likelihood ratio test from a sample size $b = 200$ *= 200 using critical values* $C2$

	$\hat{\kappa}(\alpha)$						
	Change point location						
	20	40	50	100	150	160	180
α							
0.01	0.05389	0.9980	1	1	1	1	0.3094
	0.0791	1	1	1	1	1	0.5049
0.05	0.8762	1	1	1	1	1	0.9741
0.10	1	1	1	1	1	1	1

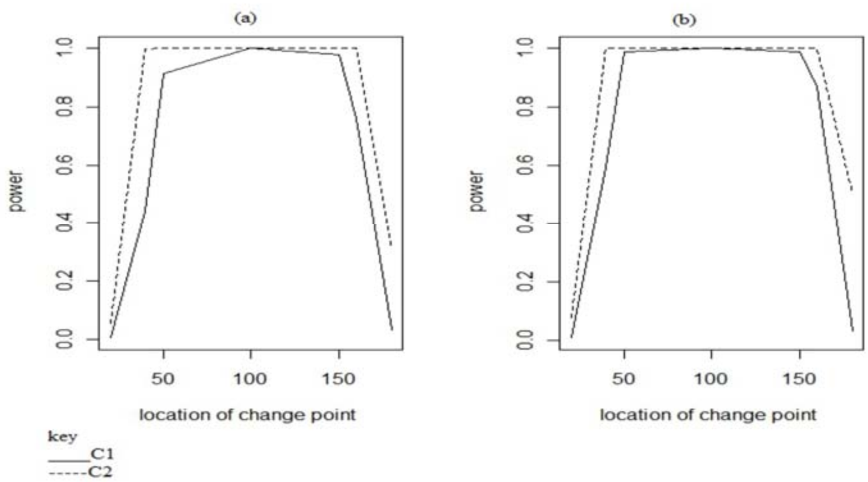

Figure 1: A plot of the power of the test against the location of change point.

Table 3: Power of the likelihood ratio test when the change point is at $\frac{b}{4}$

	$\hat{\kappa}(\alpha)$				
α	Sample size				
	50	100	150	200	500
0.01	0.005988024	0.001996008	0.005988024	0.9121756	1
	0.008594	0.009102	0.015620	1	1
0.05	0.001996008	0.02794411	0.998004	1	1
0.10	0.01596806	0.7325349	1	1	1

Table 4: Power of the likelihood ratio test when the change point is at $\frac{b}{2}$

	$\hat{\kappa}(\alpha)$				
α	Sample size				
	50	100	150	200	500
0.01	0.003992016	0.001996008	0.0259481	0.9780439	1
	0.0023297	0.026902	0.039186	1	1
0.05	0.003992016	0.0998004	1	1	1
0.10	0.02794411	0.8742515	1		1

Table 5: Power of the likelihood ratio test when the change point is at $\frac{3b}{4}$

	$\hat{\kappa}(\alpha)$				
α	Sample size				
	50	100	150	200	500
0.01	0.001996008 1	0.001996008	0.003992016	0.1836327	1
	0.002583	0.0027153	0.039142	1	1
0.05	0.001996008	0.3812375	1	1	1
0.10	0.06586826	0.998004			

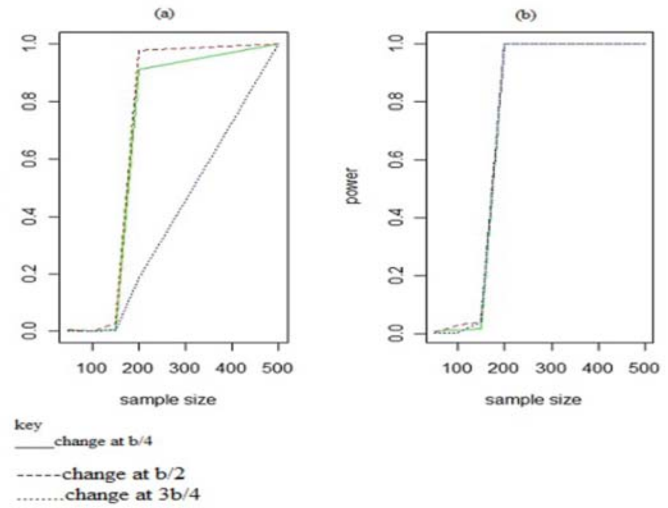

Figure 2: A plot of the power of the test against the size of the sample at $\alpha = 0.01$

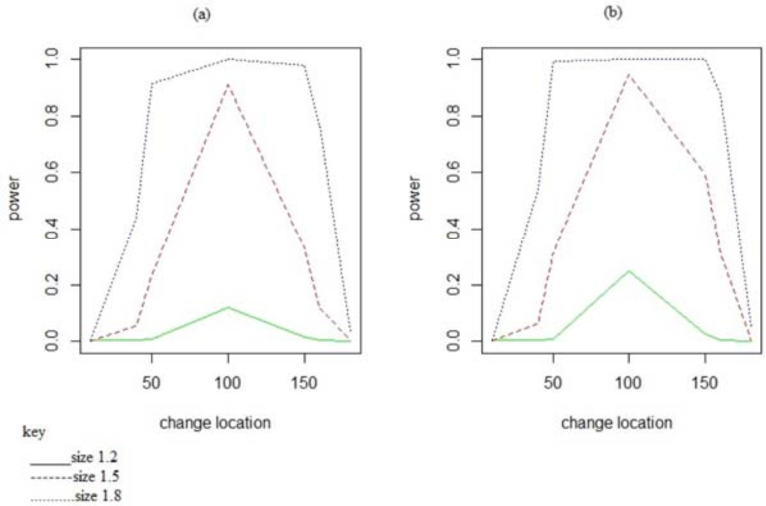

Figure 3: *A plot of the power of the test against the location of the change point at* $\alpha = 0.01$ *for the changes of size 1.2, 1.5 and 1.8.*

Table 6: Power of the likelihood ratio test for different sizes of change and change point locations k.

	$\hat{\kappa}(\alpha)$ under $C1$			
	Size of change			
k	α	$\Delta = 1.2$	$\Delta = 1.5$	$\Delta = 1.8$
20	0.01	0.003992016	0.001996008	0.003992
		0.00487432	0.00219754	0.00538710
	0.05	0.06387226	0.4530938	0.8323
	0.10	0.8023952	1	1
	0.01	0.003992016	0.05588822	0.4411
40		0.00473981	0.06429013	0.53961
	0.05	0.8163673	1	1
	0.10	1	1	1
50	0.01	0.007984032	0.2315369	0.9142
		0.00842108	0.3154287	0.99412764
	0.05	0.9520958	1	1
	0.10	1	1	1
100	0.01	0.1197605	0.9121756	1
		0.251964	0.9458210	1
	0.05	1	1	1
	0.10	1	1	1
	0.01	0.01796407	0.3313373	0.9800
150		0.027210945	0.59430631	1
	0.05	0.9820359	1	1
	0.10	1	1	1
160	0.01	0.003992016	0.1157685	0.7625
		0.00492373	0.3154287	0.8764
	0.05	0.8582834	1	1
	0.10	1	1	1
	0.01	0.001996008	0.003992016	0.02994
180		0.00284714	0.00572859	0.0529173
	0.05	0.1077844	0.499002	0.9661
	0.10	0.8622754	1	1

7.0 Application to real data

To demonstrate the use of artificial neural networks in the estimation of the conditional means we used the Bliss (1935) beetles data, where batches of adult beetles were exposed to gaseous carbon disulphide for five hours. This data has been extensively used by statisticians in studies of generalized link functions e.g., Prentice and Ross (1976), Stukel (1988) and is used by Spiegelhalter et al. (1996) to demonstrate how BUGS handles generalized linear models for binomial data. The data is given below.

Table 7: Beetles Data

Dosage (CS2mg/litre)	No. of beetles	No. of beetles killed
49.057	59	6
52.991	60	13
56.911	62	18
60.842	56	28
64.759	63	52
68.691	59	53
72.611	62	61
76.542	60	60

Here we assumed that $p_i(m_i | X_i = x) = \beta_0 + \beta_1 X_{1i}$ where mi is the number of deaths due to the i^{th} dose and X_{1i} is the respective dose. Then as in equations (3) and (17), the values of $P(m_i | X_1)$ may be estimated. The dosage at which 50% of the beetles are killed is called the LD50. One may be interested in the determination of this dosage since it indicates a significant change in the structure of the probability of death. From the data the fourth dosage of 60.842 $CS_2\ mg\ /\ litre$ kills 50% of the beetles. This shows that there might be a change in the functional structure of probability at the fourth dosage.

We also compared the estimates of conditional means obtained through the parametric method using a generalized linear fit and those obtained using the neural network. The results are presented in Table 8.

A graph of the estimated probabilities against the dose is given in Figure 4. It is evident from this graph that the estimates obtained using neural networks are nearer the actual values than those obtained through the generalized link function. The probability of death LD50 is approximately 0.5. A horizontal line through this point indicates that the fourth dosage is the LD50 and that the neural network method estimate is nearer the actual dosage than the generalized link function method estimate.

Taking the estimated probabilities from the data as the actual probabilities we computed the mean square.

Table 8: Estimated probabilities of death

Actual probabilities= $\frac{m_i}{n_i}$	Estimates fitted using glm	Estimates fitted using nnet
0.1016949	0.7011985	0.1189710
	0.16732799	0.1801028
0.216667	0.34796279	0.3027226
	0.58696004	0.5230217
0.2903226	0.79040075	0.7834121
	0.90945597	0.9394865
0.500000	0.96386496	0.987661
	0.98611696	0.9977424
0.8253968		
0.8983051		
0.9838710		
1.00000		

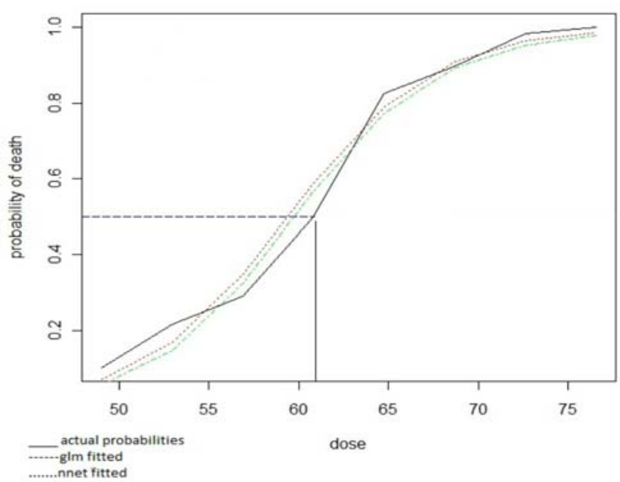

Figure 4: A plot of the estimated probabilities of death against dosage

error. The generalized link function method had an m.s.e. of 0.002032222 while the neural network estimates had an m.s.e. of 0.0007246531. Thus in terms of m.s.e. the neural network estimates are better than the generalized link function method estimates.

We wanted to determine whether our test would be able to detect the dosage at which 50% of the beetles are killed. In line with Gombay and Horvath (1996) we generated the critical values $C1$ and $C2$, using a sample size as 481 which are presented in the Appendix. The graph in Figure 5 gives us a value test statistic of 15.44, which is the maximum and this leads to the acceptance of the hypothesis of change.

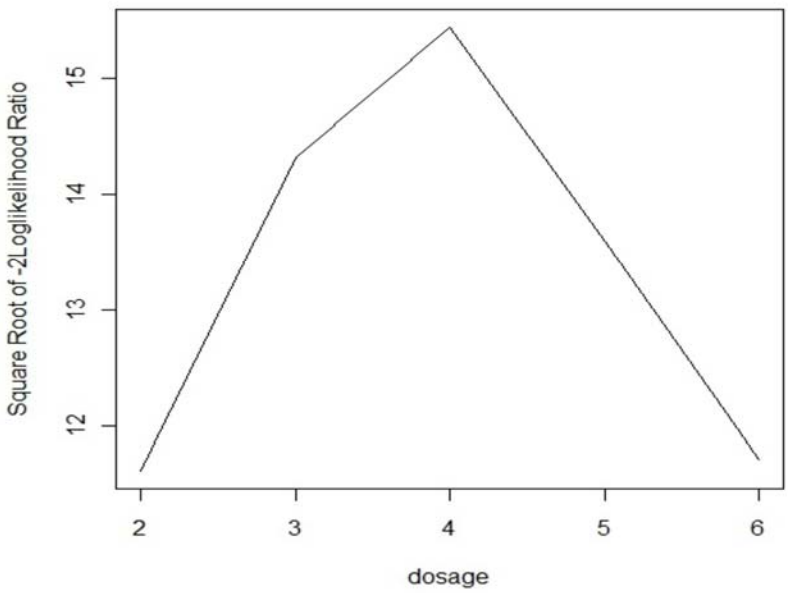

Figure 5: Plot of the values of the test statistic for the Bliss data

8.0 Conclusion

This paper proposes the use of artificial neural network in the estimation of the conditional binomial probabilities and then uses 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. It is also evident that the neural network method gives better estimates than the parametric method.

In the analysis of the beetles' data we were able to estimate the conditional means using the neural network which we compare with the parametric method using a generalized link function model. The values of the m.s.e. indicate that the neural network performs better than the generalized link function method in estimating the conditional probabilities. Our test was able to correctly identify change at the fourth dosage which corresponds to LD50.

The simulations 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

Table 9: The critical values $C1$ and $C2$, generated using Theorem 2.1 and 3.1 respectively in Gombay and Horvath (1996)

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
481	0.01	5.310178	4.73092
	0.05	4.456003	4.274104
	0.1	4.078778	4.049254
500	0.01	5.319912	4.966611
	0.05	4.167178	4.517474
	0.1	4.09062	4.296645

References

Bliss, C. I. (1935). The calculation of the dosage-mortality curve. Annals of Applied Biology, 22:134-167.

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

Franke, J. and Neumann, M. (2000). Bootstrapping neural networks. Neural Computation, 12(12):1929-1949.

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

Hwang, J. and Ding, A. (1997). Prediction intervals for artificial neural network. Journal of the American Statistical Association, 92(438):748-757.

Jaruskova, D. (1997). Some problems with application of change point detection methods to environmental data. Envirometrics, 8:469-483.

Prentice, A. and Ross, L. (1976). A generalization of the probit and logit methods for dose response curves. Biometrics, 32:134-167.

Spiegelhalter, D. J., Thomas, A., Best, N., and Gilks, W. R. (1996). Bugs: Bayesian inference using gibbs sampling, version 0.5 (version ii). Cambridge, UK:Biostatistics Unit.

Stukel, T. A. (1988). Generalized logistic models. Journal of the American Statistical Association, 83:426-431.

Waititu, A. (2008). Nonparametric change point analysis for Bernoulli random variables based on neuralnetworks. PhD thesis.

Worsley, K. (1983). The power of the likelihood ratio and cumulative sum tests for change in binomialprobability. Biometrika, (70):455-464.