Likelihood Asymptotics for Changepoint Problem

K. O. Obisesan
Department of Statistics
University of Ibadan
Nigeria
email:ko.obisesan@ui.edu.ng ;
obidairo@gmail.com

ABSTRACT

Changepoint problems are often encountered when series undergo abrupt changes or discontinuities. Detecting changepoints can signal useful actions towards sustainable developments. However the presence of changepoints have often been known to lead to failure of some regular assumptions. In theory much has not been done on which assumptions fail and to what extent will it affect the score functions of the likelihood asymptotic. In this work we concentrate on simulating the likelihood function using R to establish the failure of regular assumption due to the presence of changepoint. The failure of regular assumption is established using various score functions coded in R thereby making it possible to understand the statistical theory and the consequences of the failure of assumptions as a result changepoints.

CCS Concepts

• **Computing methodologies →Simulation →Simulation evaluation**

Keywords

changepoint, likelihood asymptotic, regular assumption, simulation.

1. INTRODUCTION

Changepoints are referred to as discontinuities that can lead to non-linearity even in complex functions (Chen and Gupta, 2000). The causes of changepoints include; changes in locations of observations, equipment, measurement methods, environmental effects, regulations, standards and so on. Generally we need to investigate the potential presence of possible changes in the data set indicating data quality problems that should be resolved prior to any subsequent analysis. This will therefore signal signs for timely protection and knowing this could be highly advantageous in planning for the future. However Yang.et.al.(2006) noted that changes do occur even in the best regulated systems. They indicated that discrepancies in records, occasional disagreement between documentation and data, abnormal data entry, changed units of measurement and other problems require adequate attention. Most times we need to detect the number of changepoints, or jumps, and their locations whereas it is noted in Mainly(2001) that it is much easier if the point of change is known. This case is referred to as intervention analysis. In contrast when the point is unknown this leads to various complexities and non-linearity.

CoRI'16, Sept 7–9, 2016, Ibadan, Nigeria.

Many applications of changepoints analysis exist. Relevant literature can be found in many fields including; Biology, Physics, Chemistry, Environmental Sciences and Climate Change, Engineering, Econometrics, Medicine, Behavioral Sciences, Political Science, Finance, Image Analysis, Security etc. The earliest works found seem to be those by Page(1954, 1955, 1957) where the cumulative sum(CUSUM) approach was used. Consequently Jandhyala and MacNeil(1986) and Jandhyala et.al(1999) provided detailed reviews of many approaches to changepoint modelling. It is important to state that the large body of literature exists due to the fact that the standard theory breaks down where the time of change is unknown. Much has not been done in showing the breakdown of the standard theory as regards failure of regular assumption. More details in respect of standard theory on changepoint are available in Easterling and Peterson(1995), Chen and Gupta(2000), Lu et.al (2005),Hanesiak and Wang(2005) and Wang(2006).

It is indicated in Obisesan.et.al(2013) that the data analysed were the physico-chemical properties of water samples obtained from two reservoirs in Oyo State Nigeria. The data were seen to contain some abrupt changes in behaviour. In the work various charts and diagrams were engaged in showing the positions and locations of changepoints and the likelihood function was written to show the single changepoint detection. However the theory on changepoint linking failure of assumption was not shown therefore this present work attempts to extend the likelihood theory to show the implications of failure of regular assumptions as a result of the presence of changepoint.

2. STANDARD TECHNIQUES OF LIKELIHOOD ASYMPTOTICS

To study the inference of changepoint problems such as to understand its non-standard nature it is important to review some properties of likelihood functions. The likelihood function for a scalar parameter θ based on data $X = X_1, \cdots\cdots, X_n$ as a collection of independence observations is defined to be

$$L(\theta|X) = f(X{:}\,\theta) = \prod_{i=1}^{n} f(X_i{:}\,\theta)$$

which is simply the joint density of the data, regarded as a function of the parameter (Rice, 2007). For convenience, we study the log-likelihood function $L(\theta|X) = l(\theta)$ and write

$$l(\theta) = lnL(\theta|X)$$
$$= lnf(X_1:\theta) + lnf(X_2:\theta) + \cdots + lnf(X_n:\theta)$$
$$= \sum_{i=1}^{n} lnf(X_i:\theta)$$

The maximum likelihood estimate of $\theta, \hat{\theta}$ which is a value of θ that minimizes the log-likelihood function. If the likelihood function is a differentiable function of θ then $\hat{\theta}$ will be the root $\frac{\partial l(\theta)}{\partial\theta} = 0$ Moreover, for a local maximum we need $\frac{\partial^2 l(\theta)}{\partial\theta} < 0$ at $\hat{\theta}$ The main assumptions here can be stated simply as

Assumption 1 : *The log-likelihood is a twice differentiable function.*

Assumption 2 : *The second derivative* $\frac{\partial^2 l(\theta)}{\partial\theta} < 0 \; at \; \hat{\theta}$

3. THE SCORE FUNCTION: SIMULATION

Under **Assumption 1**, the first derivative is usually called the **score function**: $\frac{\partial lnL(\theta)}{\partial\theta} = \frac{\partial[\sum_{i=1}^{n} lnf(X_i;\theta)]}{\partial\theta}$ and is regarded as a function of θ for fixed **X** This function plays a central role in maximum likelihood theory. We can also define the observed information as

$$J(\theta) = \frac{\partial^2 l(\theta)}{\partial\theta^2} = -\frac{\partial^2}{\partial\theta^2}\sum_{i=1}^{n} l_i(\theta) = \sum_{i=1}^{n}\frac{\partial^2 lnf(x_i;\theta)}{\partial\theta^2}$$

which is a sum of n components. Also the Fisher information is defined as

$$I(\theta) = E\left\{-\frac{\partial^2 l(\theta)}{\partial\theta^2}\right\} = E\,J(\theta)$$

and

$$I(\theta) = E\left\{-\frac{\partial^2}{\partial\theta^2}\sum_{i=1}^{n} ln\,f(x_i;\theta)\right\}$$

Which can be written

$$I(\theta) = \sum_{i=1}^{n} -E\left\{\frac{\partial^2 ln\,f(x_i;\theta)}{\partial\theta^2}\right\}$$
$$= ni_f(\theta)\; say$$

Where $i_f(\theta)$ refers to single observation information.

Now we show some characteristics of the score function when data are assumed generated from $(.\,;\theta_0)$ so that θ (assumed true value of θ is the parameter to be estimated. If we have an independent and identically distributed sample of size n, the log-likelihood is written as

$$l(\theta) = \sum_{i=1}^{n} ln\,f(x_i|\theta) \quad (1)$$

A careful illustration of the behavior of the *score function* is given in Figure 1. This allows the sampling variation of *score function* for different models (Normal, Poisson, Binomial and Cauchy) for samples of size $n = 10$. Figure 1(a) shows 25 *score function*, each based on independent and identically distributed sample of size $n = 10$ from N(4, 1). Each function is exactly linear and the score varies around 10 at the true parameter $\theta = \theta_0 = 4$. Figure 1(b) shows *score function* for 25 independent samples of size 10 from a Poisson distribution with mean 4 (Each function looks approximately linear) and at the true parameter $\theta = 4$ the *score function* also varies around 0. Figure 1(c) shows score function of 25 independent samples of size n=10 from binomial (10, 0.4) where $\theta = 0.4$ In Figure 1(d), the *score function* for Cauchy $\theta = 4$ distributions (also based on 25 independent samples of size 10) are rather irregular and fail to behave as the previous models (although the *score function* also varies around 0 at $\theta = 4$ but there is the potential for multiple roots to the score equation). This case indicates problems with a complicated likelihood.

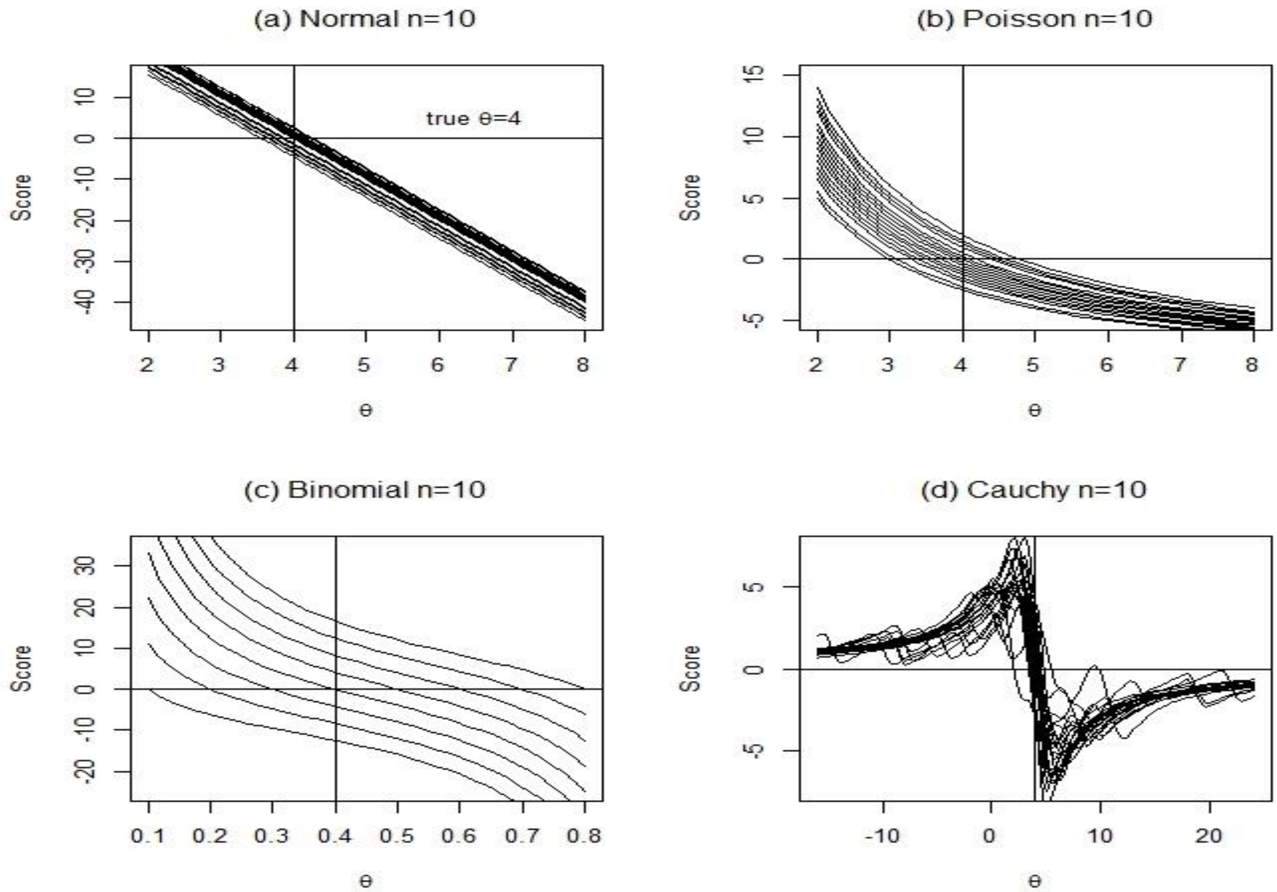

Figure 1: Sampling variation of score functions for different distributions.

In all examples in Figure 1, the score varies around zero at the true parameter value. We can show this is generally the case. Recall from Section 3 that the score function is the first derivative of the log-likelihood function where we set $U(\theta, X) = \frac{\partial l(\theta)}{\partial \theta}$, then at the true value of θ which is θ_0 we have

$$E\big(U(\theta_0; X)\big) = \int_{-\infty}^{\infty} u\,(\theta_0; x)\,.\,f(x|\theta_0)dx$$

$$= \int_{xf(x|\theta_0)>0}^{\cdot} \frac{\partial \partial l(\theta)}{\partial \theta_{|\theta=\theta_0}}\, f(x|\theta)dx$$

$$= \int_{xf(x|\theta_0)>0}^{\cdot} \frac{1}{f(x|\theta_0)} \frac{\partial f(x|\theta_0)}{\partial \theta_{|\theta=\theta_0}}\, f(x|\theta_0)\, dx$$

$$= \frac{\partial}{\partial \theta} f(x|\theta_0)dx = \frac{\partial}{\partial \theta}[1] = 0$$

The major assumption here is needed to justify interchanging the order of differentiation and integration and can be stated in **Assumption 3** as

Assumption 3 : *The range of integration does not depend on* θ

Therefore using the stated assumptions we have $E\big(U(\theta_0; \boldsymbol{X})\big) = 0$ as required. We have also find the variance of the score function} as

$$V\big(U(\theta_0; \boldsymbol{X})\big) = E\big(U^2(\theta_0; X)\big) - \Big(E\big(U(\theta_0; X)\big)\Big)^2$$
$$= E[U^2(\theta_0; X)]$$

since $E\big(U(\theta_0; X)\big) = 0$ as seen above. We can rewrite $U(\theta_0; X)$ as say

$$U(\theta_0; X) = \sum_{i=1}^{n} \left\{ \frac{\partial ln\, f(x_i, \theta)}{\partial \theta_{\theta=\theta_0}} \right\} = \sum_{i=1}^{n} S_i(\theta_0)$$

This implies that

$$E(U(\theta_0; X)) = E\left[\sum_{i=1}^{n} S_i(\theta_0) = nE\big(S_i(\theta_0)\big)\right]$$

Now

$$E\big(S_i(\theta_0)\big) = \int_{-\infty}^{\infty} \frac{\partial ln\, f(x_i, \theta)}{\partial \theta}\, f(x_i, \theta)\, dx \qquad (2)$$

and $E\big(S_i(\theta_0)\big) = 0$. Differentiating Equation 2 with respect to θ we have

$$\frac{\partial}{\partial \theta}\, E\big(S_i(\theta_0)\big) = \int_{-\infty}^{\infty} \frac{\partial^2 ln\, f}{\partial \theta^2}\;\, f(x_i, \theta_0)dx$$

Therefore

$$\frac{\partial}{\partial\theta} E\big(S_i(\theta_0)\big) = \int_{-\infty}^{\infty} \frac{\partial}{\partial\theta}\left[\frac{\partial ln\, f}{\partial\theta}\right] \; f(x_i,\theta_0)dx$$

$$= \int_{-\infty}^{\infty} \frac{\partial}{\partial\theta}\left[\frac{1}{f}\frac{\partial f}{\partial\theta}\right] \; f(x_i,\theta_0)dx$$

$$= \int_{-\infty}^{\infty}\left[\frac{1}{f}\frac{\partial f}{\partial\theta} - \frac{1}{f^2}\left(\frac{\partial f}{\partial\theta}\right)^2\right] f(x_i,\theta_0)dx$$

At θ_0 therefore we have

$$\frac{\partial}{\partial\theta^2{}_{|\theta=\theta_0}} E\big(S_i(\theta_0)\big)$$

$$= \int_{-\infty}^{\infty}\left\{\frac{1}{f(x:\theta_0)}\frac{\partial^2 f}{\partial\theta^2_{|\theta=\theta_0}} - \frac{1}{f^2(x,\theta_0)}\left\{\frac{\partial f}{\partial\theta_{|\theta=\theta_0}}\right\}^2\right\} f(x_i,\theta_0)dx$$

$$= \int_{-\infty}^{\infty} \frac{\partial^2 f}{\partial\theta^2_{|\theta=\theta_0}}\, dx - \int_{-\infty}^{\infty}\left[\frac{1}{f(x:\theta_0)}\frac{\partial f}{\partial\theta_{|\theta=\theta_0}}\right]^2 f(x_i,\theta_0)dx$$

Since the score function is a sum of *n* independent random variables, the last equation above shows that

$$V\big(U(\theta_0;X)\big) = E\big(U^2(\theta_0;X)\big) = nE\left[\frac{\partial^2 ln\, f(x\,;\,\theta)}{\partial\theta^2}\right]_{\theta=\theta_0} \qquad (3)$$

Next we see how $U(\theta_0;X)$ behaves by studying $\frac{1}{n}\,U(\theta_0;X)$ as $n\to\,\infty$ we have

$$E\left[\frac{U(\theta_0;X)}{n}\right] = \frac{1}{n}\,E\big(U(\theta_0;X)\big) = 0$$

and also that (assuming $\partial^4 In\,\frac{f}{\partial\theta^2} < \,\infty$)

$$V\left[\frac{U(\theta_0;X)}{n}\right] - \frac{1}{n}\,nE\left[\frac{\partial^2 ln\, f(x\,;\,\theta)}{\partial\theta^2}\right]_{\theta=\theta_0} \to 0 \qquad (3)$$

as $n\to\,\infty$. Hence $n^{-1}U(\theta_0;X)\to 0$ in probability as $n\to\,\infty$.

The discussion so far has dealt with the behavior of the *score function* at θ_0 the true parameter value. We now consider its behavior at other values of θ. In general (we need to investigate a case that indicate the existence of changepoint) for $\theta = \theta_0$ we find that there may be need for another assumption:

Assumption 4: *For $\theta\neq\,\theta_0$ the density $f(x;\theta)$ differs from $f(x;\theta_0)$ on a set of non zero measure.*

*Note that $E\big(U(\theta;X)\big)\neq 0$ unless $\frac{\partial f}{\partial\theta} = 0$ for all x (which itself contradicts **Assumption 4**).*

Then for an arbitrary value $\theta\neq\,\theta_0$

$$E\Big(U\big((\theta;X)\big)\Big) = \int u(\theta,x)f(x|\theta_0)\,dx = \int\frac{1}{f(x|\theta_0)}\frac{\partial f(\theta,x)}{\partial\theta}\,f(x|\theta_0)\,dx \qquad (4)$$

Now consider $V\big(U(\theta;X)\big) = V\left[\sum_{i=1}^{n}\frac{\partial lnf(x|\theta_0)}{\partial\theta}\right] = nV\left[\frac{\partial lnf(x|\theta_0)}{\partial\theta}\right]\infty\, n\; if\; V\left[\frac{\partial lnf(x|\theta_0)}{\partial\theta}\right]$ is finite for all θ. Therefore for a case where $\theta\neq\,\theta_0$ we have $E\Big(U\big((\theta;X)\big)\Big) = E\left[\frac{\partial lnf(x|\theta_0)}{\partial\theta}\right]\neq 0$ and $V[n^{-1}\,U(\theta,X)] = \frac{1}{n^2}\,V\left[\frac{\partial lnf(x|\theta_0)}{\partial\theta}\right]\to 0$ and so $n^{-1}U(\theta,X)\,\to E\left[\frac{\partial lnf(x|\theta_0)}{\partial\theta}\right]\neq 0$. Therefore as $n\to\infty$ the function $U^*(\theta) = n^{-1}\,U(\theta,X)$ tends to a deterministic function with root θ_0

4. SIMULATION CODE WITH R.

In this section the **R** code used in simulating the likelihood functions for the Normal and Poison distributions are stated as run from the prompt. The Binomial and Cauchy distributions follow similar way. After simulating from the distributions the likelihood functions are plotted to show the distribution of the parameter. It is clear from the code that the expected value of the score function moves around 0.

```
set.seed(3)
n<- 10
#............................... Normal Score Functions:
  t0<- 4
  x<- rnorm(n,t0)
  theta<- seq(t0/2,t0*2,len=40)
  stheta<- n*(mean(x)-theta)

par(mfrow=c(1,2))
plot(theta,stheta,type='n',
     xlab=expression(theta),ylab='Score',cex=.6)
  lines(theta,stheta,lwd=.4)
  title(expression('(a) Normal n=10'))
  text(6.5,5.5,expression(paste('true ',theta,'=4')))
  abline(v=t0,h=0)

for(i in 1:20){
  x<- rnorm(n,t0)
  stheta<- n*(mean(x)-theta)
  lines(theta,stheta,lwd=.1)
}


# ............................... Poisson Score Functions:
t0<- 4
x<- rpois(n,t0)
theta<- seq(t0/2,t0*2,len=40)
stheta<- -n + sum(x)/theta

plot(theta,stheta,type='n',xlab=expression(theta),
     ylab='Score',ylim=c(-5,15),cex=.6)
for(i in 1:20){
  x<- rpois(n,t0)
  stheta<- -n + sum(x)/theta
  lines(theta,stheta,lwd=.1)
}
```

abline(v=t0,h=0)
title(expression('(b) Poisson n=10'))

5. Consistency of Maximum Likelihood Estimators

We now consider whether $\hat{\theta}$ is a constant estimator of θUsing a Taylor expansion for $U(\theta)$ around θ_0, we have

$$U(\theta) = U(\theta_0) + (\theta - \theta_0)\frac{\partial U}{\partial \theta_{|\theta=\theta^{\dagger}}}$$

For some $\theta^{\dagger} \in (\theta_0, \theta)$ and so we can write

$$(\theta - \theta_0) = \frac{U(\theta) - U(\theta_0)}{\frac{\partial(U)}{\partial \theta_{|\theta^{\dagger}}}}$$

In particular, when $\theta = \hat{\theta}$ then we have (nothing that $U(\hat{\theta}) = 0$)

$$\hat{\theta} - \theta_0 = -\frac{U(\theta_0)}{\frac{\partial(U)}{\partial \theta_{|\theta^{\dagger}}}} \quad (6)$$

Which can be rewritten as

$$\hat{\theta} - \theta_0 = -\frac{n^{-1}U(\theta_0)}{n^{-1}\frac{\partial(U)}{\partial \theta_{|\theta^{\dagger}}}} \quad (7)$$

Note that the numerator of Equation 6 approaches 0 as $n \to \infty$ If we assume that the denominator is guaranteed nonzero, then Equation 6 implies that $\theta - \theta_0 \to 0$ and therefore $\hat{\theta} - \theta_0$. This requires the following assumption which can be seen as a strengthened version of **Assumption 2.**

Assumption 5: $\frac{\partial U}{\partial \theta}$ is non- zero in an interval containing θ_0.

6. Limiting Distribution of $\hat{\theta}$

As well as demonstrated the consistency of the maximum likelihood estimator $\hat{\theta}$, Equation 5 allows us to establish its distribution when n is large. Recall again that

$$V(U(\theta_0)) = E(U^2(\theta_0, X)) = -nE\left[\frac{\partial^2 lnf(x;\theta)}{\partial \theta^2}\right]_{\theta=\theta_0}$$

Moreover, $U(\theta_0)$ is a sum of independent and identically distributed contributions. Hence from the central limit theorem we have asymptotically,

$$\frac{U(\theta_0)}{\sqrt{-nE\left[\frac{\partial^2 lnf}{\partial \theta^2_{|\theta_0}}\right]}} \sim N(0,1)$$

Now write $Z = \frac{U(\theta_0)}{\sqrt{-nE\left[\frac{\partial^2 lnf}{\partial \theta^2_{|\theta_0}}\right]}}$. Then from Equation 6, we have

$$\hat{\theta} - \theta_0 = \frac{Z\sqrt{-nE\left[\frac{\partial^2 lnf}{\partial \theta^2_{|\theta_0}}\right]}}{\frac{\partial(U)}{\partial \theta_{|\theta^{\dagger}}}}$$

Hence we write

$$\hat{\theta} - \theta_0\sqrt{-nE\left[\frac{\partial^2 lnf}{\partial \theta^2_{|\theta_0}}\right]} = Z\frac{-nE\left[\frac{\partial^2 lnf}{\partial \theta^2_{|\theta_0}}\right]}{\frac{\partial(U)}{\partial \theta_{|\theta^{\dagger}}}}$$

$$= Z\frac{-E\left[\frac{\partial^2 lnf}{\partial \theta^2_{|\theta_0}}\right]}{n^{-1}\frac{\partial(U)}{\partial \theta_{|\theta^{\dagger}}}} \quad (7)$$

Now as $n \to \infty, n^{-1}\frac{\partial(U)}{\partial \theta_{|\theta^{\dagger}}} \to E\left(\frac{\partial(U)}{\partial \theta_{|\theta^{\dagger}}}\right)$and $\theta^{\dagger} \to \theta_0$ since it lies between $\hat{\theta}$ and θ_0. Also, if $\frac{\partial^2 lnf}{\partial \theta^2}$ ifs a continuous in θ then as $\theta^{\dagger} \to \theta_0$

$$E\left[\frac{\partial^2 lnf}{\partial \theta^2_{|\theta^{\dagger}}}\right] \to E\left[\frac{\partial^2 lnf}{\partial \theta^2_{|\theta_0}}\right]$$

In the case, the final term in Equation 7 $\to$ 1 and we have

$$(\hat{\theta} - \theta_0)\sqrt{-nE\frac{\partial^2 lnf}{\partial \theta^2_{|\theta_0}}} \to Z$$

which has a standard normal distribution asymptotically.

7. Limiting Chi-Square Distributions: Likelihood Ratio Statistic

Now we discuss the basic test statistic used for testing hypothesis using the principles of likelihood functions. Suppose that *l(.)* is the log-likelihood established from the probability density f . then the consistency of $\hat{\theta}$ implies that we can write

$$l(\theta_0) = l(\hat{\theta}) + (\hat{\theta} - \theta_0)l'(\hat{\theta}) + \frac{(\hat{\theta} - \theta_0)^2}{2} l''^{(\theta^{\dagger})}$$

Where $\theta^{\dagger}$ is between $\theta^{\dagger}$ and θ_0.

Then representing the likelihood ratio statistic with $L_r = 2\left(l(\hat{\theta}) - l(\theta_0)\right)$ gives

$$L_r = 2(\hat{\theta} - \theta_0)\, l'(\hat{\theta}) - (\hat{\theta} - \theta_0)^2 l"(\theta^{\dagger})$$

and since $l'^{(\hat{\theta})} = 0$ by definition we can write that

$$L_r = ni_f(\theta_0)(\hat{\theta} - \theta_0)^2 \times \frac{l"(\theta^{\dagger})}{l"(\theta_0)} \times \frac{l"(\theta_0)}{-ni_f(\theta_0)}$$

$$= ni_f(\theta_0)(\hat{\theta} - \theta_0)^2 \times \frac{l"(\theta^{\dagger})}{l"(\theta_0)} \times \frac{\frac{l"(\theta_0)}{n}}{-i_f(\theta_0)} \quad (8)$$

It is clear that the first part of Equation 8 is asymptotically the square of a standard normal random variable and it is therefore a χ^2_1 distribution in addition, the last two ratios $\frac{l"(\theta^{\dagger})}{l"(\theta_0)}$ and $\frac{l"(\theta_0)}{-i_f(\theta_0)}$ tend to 1 using similar arguments to those applied in the previous subsection. In the same direction, we can obtain the χ^2 distribution for a case when θ is vector (without proof) in that as above we write $L_r = 2[l(\hat{\theta}) - l(\theta_0)] = (\hat{\theta} - \theta_0)^{T_{i_f}}(\theta_0)(\hat{\theta} - \theta_0)$. It is therefore noted that $L_r(\theta)$ has an approximate chi-square distribution on p degree(s) of freedom for repeated sampling of data from the model. We can write $L_r(\theta) \to^D \chi^2_p$.

7.1 The two-mean model

In Obisesan et.al(2013), the development of changepoint detection was based on Hinkley(1970) work. Hinkley(1970) considered sequences of random variables and discussed the point at which the probability distribution changes using a normal distribution with changing mean. The asymptotic distribution of the maximum likelihood estimate discussed in this paper is particularly relevant to change-point. The author indicated the simplest model over a whole range of data as $X_t = \theta(t) + \epsilon_t \, for \, t = 1, \cdots T$ as usual where $\theta(t)$is a mean function and ϵ_t refer to error terms. Hinkley (1970) computed the asymptotic distribution in the normal case when θ_0 and θ_1 are unknown. The asymptotic dsitribution is found to be the same when the mean levels are known. The two-mean model to be considered supposes that there exist a mean $\theta_0(t)$ and mean $\theta_1(t) \, for \, t = 1, \cdots, \delta$ and $t = \delta + 1, \cdots, T$ respectively. He also computed the asymptotic distribution of the likelihood estimate of the change-point $\hat{\delta}$ (where θ_0 and θ_1 are known and δ is unknown) is obtained from a sample $x_1, \ldots, x_T$by simply maximizing the likelihood function of the form

$$L(\delta, \theta_0, \theta_1) = \prod_{i=1}^{\delta} f(x_i; \theta_0) \prod_{i=\delta+1}^{T} f(x_i; \theta_1)$$

which can be written in form of log likelihood as

$$L(\delta, \theta_0, \theta_1) = \sum_{i=1}^{\delta} lnf(x_i; \theta_0) \sum_{i=\delta+1}^{T} lnf(x_i; \theta_1) \qquad (9)$$

Moreover, many cases arise when the mean levels are not known. The log-likelihood of the observed sequence $(x_1, \ldots, x_T)$ is

$$L(\delta, \theta_0, \theta_1, \sigma^2 | x_1, \ldots, x_T) = -\frac{T}{2} ln\sigma^2 - \frac{1}{2\sigma^2} \left\{ \sum_{i=1}^{\delta} (x_i - \theta_0)^2 + \sum_{i=\delta+1}^{T} (x_i - \theta_1)^2 \right\} \qquad (10)$$

If we assume that δ is known therefore the maximum likelihood estimators $\theta_0, \theta_1, \, and \, \sigma^2$ respectively are $\hat{\theta} = \frac{\sum_{i=1}^{\delta} x_i}{\delta}$ $\hat{\theta}_1 = \frac{\sum_{i=\delta+1}^{T} x_i}{T+\delta}$ and $\hat{\sigma}^2 = \frac{\sum_{i=1}^{\delta}(x_i-\theta_0)^2 + \sum_{i=\delta+1}^{T}(x_i-\theta_1)^2}{T}$

Particularly for convenience. Hinkley (1970) substituted $\sigma^2 = 1$ as known so that Equation 10 becomes

$$L(\delta, \theta_0, \theta_1, \sigma^2 | x_1, \ldots, x_T) = \frac{1}{2} \left\{ \sum_{i=1}^{\delta} (x_i - \theta_0)^2 + \sum_{i=\delta+1}^{T} (x_i - \theta_1)^2 \right\} \qquad (11)$$

Assuming that σ^2 is unknown and putting the maximum likelihood estimates of θ_0 and θ_1 back into the log-likelihood in Equation 11 and re-arranging the emerging sums of squares conditional on *t* Equation 11 was used to estimate changepoint of water pollution in Eleyele and Asejire reservoirs in Nigeria. This confirms the application of the likelihood theory of changepoint. More on the applications are discussed in Obisesan(2011, 2015).

8. RESULTS

In this work it has been shown that changepoint arises as a result of failure of some regular assumptions specifically in this case **Assumptions 1** and **Assumptions 4** may fail. This work has justified using simulation in the theory of likelihood function for the score functions to show the change in parameter allowing changepoint to occur. The work also justifies the application of changepoint detection as used in Obisesan et.al(2013). The use of **R** has therefore made it possible to show the failure of regular assumption.

9. CONCLUSION

Single changepoint detection has been discussed in the framework of the failure of regular assumptions that have not been commonly noticed. Likelihood function was used to merge the two-mean levels and various score functions were simulated using the successful statistical computing language R. The theoretical implications of failure of regular assumptions were discussed and the failed assumption identified using R . This work has therefore provided a basis for using computational statistics methods in solving a mathematical problem.

10. REFERENCES

[1] Chen, J. amd Gupta, A.L. (2000): *Parametric statistical change Point Analysis*. Birkhauser Boston.

[2] Easterling, D.R. and Peterson, T.C. (1995). *A New Method for Detecting Undocumented Discontinuities in Climatological Time Series*. Int.J. Climatol, 15:369-377.

[3] Hanesiak, J.M. and Wang, X.L. (2005). Adverse *Weather Trends in the Canadian Arctic* J.Climate, 18(2):3140-3156.

[4] Hinkley, D.V. (1970). *Inference about the Change-point in a Sequence of Random Variables*. Biometrika, 57(1):1-17.

[5] Jandhyala, V.K., Fotopoulus, S.B., and Evagelopoulos, N. (1999). *Change-point Methods for Weibull Models with Applications to Detection of Trends in Extreme Temperature.* Environmetrics, 10:547-564.

[6] Jandhyala, V.K., and MacNeil, I.B. (1986) *The Changepoint Problem: A review of Applications*. Elsevier: In statistical Aspects of Water Quality Monitoring (Eds. A.H. ElShaarawi and R.E. Kwiatkowski), pages 381-387.

[7] Lu, Q., Lund, R., and Seymour, L. (2005). *An update of US temperature trends*. J. Climate, 18:4906-4919.

[8] Manly, B.F.L. (2001). *Statistics for Environmental Science and Management*. Chapman Hall/CRC, Boca Raton.

[9] Obisesan K.O, Lawal M, Bamiduro T.A, and Adelakun A.A(2013): *Data Visualization and Changepoints Detection in Environmental Data: The Case of Water Pollution in Oyo State, Nigeria*: Journal of Science Research: Vol.12:181-190.

[10] Obisesan K.O(2015) *Modelling Multiple Changepoints Detection* Ph.D Thesis, Department of Statistics, University of Ibadan.

[11] Obisesan K.O(2011) Changepoint *Detection in Time-Series with Hydrological Applications* M.Phil Thesis, Department of Statistical Science, University College London.

[12] Page, E.S.(1954). *Continuous Inspection Schemes*. .Biometrika, 41:100-114.

[13] Page, E.S. (1955). *A test for a Change in a Parameter Occuring at an Unknown Time Point*. Biometrika, 42:523-526.

[14] Page, E.S. (1957). *On Problems in which a Change in Parameter Occurs at Unknown Point*. Biometrika, 44:248-252.

[15] Rice, J.A. (2007). *Mathematical Statistics and Data Analysis*. Duxbury.

[16] Wang, X.L. (2006). Climatology *and Trends in Some Adverse and Fair Weather Conditions in Canada, 1953-2004*. J. Geophys. Res 111

[17] Yang, C., Chandler, R.E., Isham, V.S., and Wheater, H.S. (2006). *Quality Control for Daily Observational Rainfall Series in UK.*Water and Environment Journal, 20:185-193.