

Limit Distribution of the ϕ -Divergence Based Change Point Estimator

Mwelu Susan¹, Anthony G. Waititu², Peter N. Mwita³, Charity Wamwea²

¹Pan African University Institute for Basic Sciences Technology and Innovation, Nairobi, Kenya

²Department of Statistics and Actuarial Sciences, JKUAT, Nairobi, Kenya

³Department of Mathematics, Statistics and Actuarial Sciences, Machakos University, Machakos, Kenya

Email: suemwelu@gmail.com

How to cite this paper: Susan, M., Waititu, A.G., Mwita, P.N. and Wamwea, C. (2021) Limit Distribution of the ϕ -Divergence Based Change Point Estimator. *Open Journal of Statistics*, 11, 337-350.
<https://doi.org/10.4236/ojs.2021.113020>

Received: February 10, 2021

Accepted: May 7, 2021

Published: May 10, 2021

Copyright © 2021 by author(s) and Scientific Research Publishing Inc. This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

Abstract

The assumption of stationarity is too restrictive especially for long time series. This paper studies the change point problem through a change point estimator based on the ϕ -divergence which provides a rich set of distance like measures between pairs of distributions. The change point problem is considered in the following sub-fields: the problem of divergence estimation, testing for the homogeneity between two samples as well as estimating the time of change. The asymptotic distribution of the change point estimator is estimated by the limiting distribution of a stochastic process within given bounds through asymptotic theory surrounding the likelihood theory. The distribution is found to converge to that of a standardized Brownian bridge process.

Keywords

Change Point, ϕ -Divergence, Estimator, Test Statistic, Distribution, Wald

1. Introduction

Many real world data are made up of consecutive regimes that are separated by abrupt changes [1]. Statistical research works have shown that with time, the underlying data generating processes undergo occasional sudden changes [2]. As a result the assumption of stationarity is often too strong and more often violated. Stationarity in the strict sense, implies time-invariance of the distribution underlying the process. The overall behavior of observations can change over time due to internal systemic changes in distribution dynamics or due to external factors. Modeling time series processes using stationary methods to capture their time-evolving dependence aspects will most likely result in a crude approximation as abrupt changes fail to be accounted for [3]. Reviewed literature

reveals that the use of one model may not be appropriate to model a non-stationary series and as such various parametric and non-parametric change-point estimation methods have been proposed [4] [5] [6] [7] [8]. However, they are limited in different ways and their suitability depend on the underlying assumptions. A point of interest in all aspects of life would be to detect and estimate these changes as their implication is crucial. A time series data containing change points is assumed to be piece-wise stationary implying that some characteristics of the process change abruptly at unknown points in time. Parametric tests for change point are mainly based on the likelihood ratio statistics and estimation based on the maximum likelihood method whose general results can be found in [4]. In its simplest form, change-point detection is the name given to the problem of estimating the point at which the statistical properties of a sequence of observations change [9]. Changing point problems can be classified as off-line which deals with only a fixed sample or on-line which considers new information as it is observed. Off-line change point problems deal with fixed sample sizes which are first observed and then detection and estimation of change points are done. [10] introduced the change point problem within the off-line setting. Since this pioneering work, methodologies used for change point detection have been widely researched on with methods extending to techniques for higher order moments within time series data. Change point analysis methods are applicable in a wide range of fields including but not limited to climate (climate change), quality management, medicine, finance, and genetics.

For a given set of data x_1, \dots, x_n a change point is said to occur when there exists a time $\tau \in 1, \dots, n-1$ such that the statistical properties of x_1, \dots, x_τ and $x_{(\tau+1)}, \dots, x_n$ are different. If τ is known then the two samples only need to be compared. However, if τ is unknown then it has to be analyzed through change point analysis that entails both detection and estimation of the change point/change time. The null hypothesis of no change against the alternative that there exists a time when the distribution characteristics of the series changed is then tested. Considering a change in model parameters the problem would be stated as

$$\begin{aligned} H_0 : \theta_1 = \theta_2 = \dots = \theta_n = \theta_0 \text{ (unknown)} & \text{ versus} \\ H_1 : \theta_1 = \dots = \theta_\tau \neq \theta_{\tau+1} = \dots = \theta_n & \end{aligned} \quad (1)$$

where τ is unknown and needs to be estimated.

If $\tau < n$ then the process distribution has changed and τ is referred to as the change point. Assume that there exists $\lambda \in (0,1)$ such that τ satisfies

$$\tau = \lambda n \quad (2)$$

i.e. λ is a fraction that divides the data process at the change point and n is the number of observations in a given data set. Then hypothesis 1 can be restated as

$$\begin{aligned} H_0 : \tau = n, \quad (\lambda = 1) \\ H_1 : \tau < n, \quad (0 < \lambda < 1) \end{aligned} \quad (3)$$

At a given level of significance, if the null hypothesis is rejected, then the

process X is said to be locally piecewise-stationary and can be approximated by a sequence of stationary processes that may share certain features such as the general functional form of the distribution. With the assumption that change time is unknown, [4] gives eight limiting conditions that yields the null distribution of the likelihood ratio test statistic as the supremum of a standardized Brownian bridge. [11] applied these results within a non-parametric framework and obtained similar results. [12] apply the likelihood ratio test within a parametric framework on assumption that data are drawn from extreme value distributions. Through assumption of the Von Misses condition, their test statistic weakly converges in distribution to the supremum of a squared standardized Brownian bridge.

The rest of this paper is organized as follows: Section 2 gives an overview of the change point estimator based on a the ϕ -divergence. Section 3 provides key results for the limit distribution of the divergence based change point estimator, Section 4 gives some simulation results and finally Section 5 gives the conclusion.

2. Single Change Point Detection and Estimation

The change point problem is addressed by using a ‘distance’ function between distributions. Given a distance function, a test statistic is constructed to guarantee a distance (≥ 0) between any two distributions based on a sample size n . Consider a given parametric model $f_\theta : \theta \in \Theta$ where Θ is the parameter space defined on a data set of size n . Let X_1, \dots, X_n be random variables and have probability densities $f(x; \theta_1), \dots, f(x; \theta_n)$ with respect to σ -finite measure μ with $F(x; \theta)$ generating distinct measures if $\theta \in \Theta$

Definition 2.1 (ϕ -divergence). Let F_{θ_1} and F_{θ_2} be two probability distributions. Define the ϕ -divergence between the two distributions as

$$D_\phi(F_{\theta_1}, F_{\theta_2}) = D_\phi(\theta_1, \theta_2)$$

The broader family of f-divergences (ϕ -divergences) take the general form

$$\begin{aligned} D_\phi(F_{\theta_1}, F_{\theta_2}) &= D_\phi(\theta_1, \theta_2) = \int \phi(dF_{\theta_1}/dF_{\theta_2}) dF_{\theta_2} \\ &= \int_x f_{\theta_2}(x) \phi\left(\frac{f_{\theta_1}(x)}{f_{\theta_2}(x)}\right) d\mu(x) \\ &= \mathbb{E}_{\theta_2} \left[\phi\left(\frac{f_{\theta_1}(X)}{f_{\theta_2}(X)}\right) \right], \quad \phi \in \Phi \end{aligned} \quad (4)$$

where Φ is the class of all convex functions $\phi(t)$, $t > 0$ satisfying $\phi(1) = 0$. To avoid indeterminate expressions at any point $t = 0$, the following assumptions in relation to the functions ϕ involved in the general definition of ϕ -divergence statistics are given in [13].

$$\begin{aligned} 0\phi\left(\frac{0}{0}\right) &= 0 \\ 0\phi\left(\frac{p}{0}\right) &= \lim_{u \rightarrow \infty} \frac{\phi(u)}{u} \end{aligned} \quad (5)$$

Assumption 1. The function $\phi \in \Phi : [0, \infty) \rightarrow (-\infty, +\infty)$ is convex and continuous. The restriction on $[0, \infty)$ is finite, twice continuously differentiable with $\phi(1) = \phi'(1) = 0, \phi''(1) = 1$.

Different choices of ϕ result in many divergences that play important roles in statistics. $D_\phi(\theta_1, \theta_2) \neq D_\phi(\theta_2, \theta_1)$ hence divergence measures are not distance measures but give some difference between two probability measures hence the term “pseudo-distance”. More generally a divergence measure is a function of two probability density (or distribution) functions, which has non-negative values and takes the value zero only when the two arguments (distributions) are the same. A divergence measure grows larger as two distributions are further apart. Hence, a large divergence implies departure from the null hypothesis.

Based on the divergence 4 a change point estimator can be constructed as;

$$D(\lambda) = \arg \max_{a < \tau < b} (\lambda(1-\lambda)) \frac{2}{\phi''(1)} D_\phi(\hat{\theta}_1, \hat{\theta}_2) \tag{6}$$

To test for the possibility of having a change in distribution of X it is natural to compare the distribution function of the first τ observations to that of the last $(n-\tau)$ since the location of the change time is unknown. When τ is near the boundary points, an estimation calculated on a correct large number of observations $(n-\tau)$ is compared to an estimation from a small number of observations τ . This may result to an erratic behavior of the test statistic due to instability of the estimators of the parameters [6]. [14] provides the following result.

Theorem 1. Suppose that λ maximizes the test statistic over $[0,1]$ then under the null hypothesis,

$$\begin{aligned} \sup_{\lambda \in [\varepsilon, 1-\varepsilon]} D(\lambda) &= O_p(1) \quad \forall \varepsilon \\ \sup_{\lambda \in [0,1]} D(\lambda) &\xrightarrow{p} \infty \end{aligned} \tag{7}$$

for proof of theorem 1 see [14].

If λ is not bounded away from zero and one the test statistic does not converge in distribution. However, fixed critical values can be obtained for increasing sample sizes when λ is bounded away from zero and one and yields significant power gains if the change point is in Λ . Let $\varepsilon > 0$ be small enough such that $\lambda \in (\varepsilon, 1-\varepsilon)$.

Suppose there exists constants a, b such that the unique maximum likelihood estimates $\hat{\theta}_1, \hat{\theta}_2$ exist for all $a \leq \tau \leq b$. Then the test statistic is maximized over λ such that

$$D_{nr} = \arg \max_{\varepsilon n < \tau < (1-\varepsilon)n} \left(\frac{\tau}{n} \left(1 - \frac{\tau}{n} \right) \right) \frac{2}{\phi''(1)} D_\phi(\hat{\theta}_1, \hat{\theta}_2) \tag{8}$$

where $a = \varepsilon n$ and $b = (1-\varepsilon)n$

The trimming parameter is usually taken to satisfy $0 < \varepsilon < 0.5$ [15].

Let $N(\varepsilon) = \{n\varepsilon, n\varepsilon + 1, \dots, (1 - \varepsilon)n\}$ be the set of all values over which the test statistic S is maximized. A change time τ is estimated by the least value of τ that maximizes the test statistic S .

$$\hat{\tau} = \min \left\{ \tau : D_{n\tau} = \arg \max_{N(\varepsilon)} \left(\frac{\tau}{n} \left(1 - \frac{\tau}{n} \right) \right) \frac{2}{\phi''(1)} D_\phi(\hat{\theta}_1, \hat{\theta}_2) \right\} \quad (9)$$

with $\hat{\theta}_1, \hat{\theta}_2$ being parameter estimates of θ_1 and θ_2 respectively and that they are dependent on the change point τ . $\hat{\theta}_1$ represents the parameter estimates before the change point and $\hat{\theta}_2$ gives the parameter estimates after the change point. The difference between the two estimators $\hat{\theta}_1, \hat{\theta}_2$ give an idea of the difference between the two samples hence departure from the null hypothesis.

3. Main Result

Consider a second order Taylor expansion of $D(\hat{\theta}_1, \hat{\theta}_2)$ about the true parameter values θ_1, θ_2 .

For $i = 1, \dots, d$

$$\begin{aligned} D(\hat{\theta}_1, \hat{\theta}_2) &= D(\theta_1, \theta_2) + \sum_{i=1}^d \frac{\partial D_\phi(\theta_1, \theta_2)}{\partial \theta_{1i}} (\hat{\theta}_{1i} - \theta_{1i}) + \sum_{i=1}^d \frac{\partial D_\phi(\theta_1, \theta_2)}{\partial \theta_{2i}} (\hat{\theta}_{2i} - \theta_{2i}) \\ &+ \frac{1}{2} \sum_{i=1}^d \sum_{j=1}^d \frac{\partial^2 D_\phi(\theta_1, \theta_2)}{\partial \theta_{1i} \partial \theta_{1j}} (\hat{\theta}_{1i} - \theta_{1i})' (\hat{\theta}_{1j} - \theta_{1j}) \\ &+ \frac{1}{2} \sum_{i=1}^d \sum_{j=1}^d \frac{\partial^2 D_\phi(\theta_1, \theta_2)}{\partial \theta_{2i} \partial \theta_{2j}} (\hat{\theta}_{2i} - \theta_{2i})' (\hat{\theta}_{2j} - \theta_{2j}) \\ &+ \sum_{i=1}^d \sum_{j=1}^d \frac{\partial^2 D_\phi(\theta_1, \theta_2)}{\partial \theta_{1i} \partial \theta_{2j}} (\hat{\theta}_{1i} - \theta_{1i})' (\hat{\theta}_{2j} - \theta_{2j}) \\ &+ o\left(\|\hat{\theta}_1 - \theta_1\|^2\right) + o\left(\|\hat{\theta}_2 - \theta_2\|^2\right) \end{aligned} \quad (10)$$

$$D(\theta_1, \theta_2) = 0$$

$$\begin{aligned} \frac{\partial D_\phi(\theta_1, \theta_2)}{\partial \theta_{1i}} &= \int \phi' \left(\frac{f_{\theta_1}(x)}{f_{\theta_2}(x)} \right) \frac{\partial f_{\theta_1}(x)}{\partial \theta_{1i}} d\mu(x) \\ &= \int \phi'(1) \frac{\partial f_{\theta_1}(x)}{\partial \theta_{1i}} d\mu(x) \\ &= \phi'(1) \int \frac{\partial f_{\theta_1}(x)}{\partial \theta_{1i}} d\mu(x) \\ &= \phi'(1) \frac{\partial}{\partial \theta_{1i}} \int f_{\theta_1}(x) d\mu(x) = 0 \end{aligned} \quad (11)$$

This is by assumption 1 and that $f_{\theta_1}(x)$ is a pdf.

$$\begin{aligned} \frac{\partial D_\phi(\theta_1, \theta_2)}{\partial \theta_{2i}} &= \int \phi \left(\frac{f_{\theta_1}(x)}{f_{\theta_2}(x)} \right) \frac{\partial f_{\theta_2}(x)}{\partial \theta_{2i}} - f_{\theta_2}(x) \phi' \left(\frac{f_{\theta_1}(x)}{f_{\theta_2}(x)} \right) \left(\frac{f_{\theta_1}(x)}{f_{\theta_2}(x)^2} \right) \\ &= \int \phi(1) \frac{\partial f_{\theta_2}(x)}{\partial \theta_{2i}} - \phi'(1) = 0 \end{aligned} \quad (12)$$

$$\begin{aligned} \frac{\partial^2 D_\phi(\theta_1, \theta_2)}{\partial \theta_{1i} \partial \theta_{1j}} &= \int \frac{\partial f_{\theta_1}(x)}{\partial \theta_{1i}} \phi'' \left(\frac{f_{\theta_1}(x)}{f_{\theta_2}(x)} \right) \frac{\partial f_{\theta_1}(x)}{\partial \theta_{1j}} \frac{1}{f_{\theta_2}(x)} d\mu(x) \\ \text{since } \theta_1 &= \theta_2 \\ &= \phi''(1) \int \frac{\partial f_{\theta_1}(x)}{\partial \theta_{1i}} \frac{\partial f_{\theta_1}(x)}{\partial \theta_{1j}} \frac{1}{f_{\theta_1}(x)} d\mu(x), \text{ by assumption 1} \\ &= \int \frac{\partial f_{\theta_1}(x)}{\partial \theta_{1i}} \frac{\partial f_{\theta_1}(x)}{\partial \theta_{1j}} \frac{1}{f_{\theta_1}(x)} d\mu(x) \\ \frac{\partial^2 D_\phi(\theta_1, \theta_2)}{\partial \theta_{2i} \partial \theta_{2j}} &= \int \phi'' \left(\frac{f_{\theta_2}(x)}{f_{\theta_1}(x)} \right) \frac{\partial f_{\theta_2}(x)}{\partial \theta_{2i}} \frac{\partial f_{\theta_2}(x)}{\partial \theta_{2j}} \frac{1}{f_{\theta_1}(x)} d\mu(x) \\ &= \phi''(1) \int \frac{\partial f_{\theta_2}(x)}{\partial \theta_{2i}} \frac{\partial f_{\theta_2}(x)}{\partial \theta_{2j}} \frac{1}{f_{\theta_1}(x)} d\mu(x) \tag{13} \\ &= \int \frac{\partial f_{\theta_2}(x)}{\partial \theta_{2i}} \frac{\partial f_{\theta_2}(x)}{\partial \theta_{2j}} \frac{1}{f_{\theta_1}(x)} d\mu(x), \text{ by assumption 1} \end{aligned}$$

$$\begin{aligned} \frac{\partial^2 D_\phi(\theta_1, \theta_2)}{\partial \theta_{1i} \partial \theta_{2j}} &= \int \phi'' \left(\frac{f_{\theta_1}(x)}{f_{\theta_2}(x)} \right) \frac{\partial f_{\theta_1}(x)}{\partial \theta_{1j}} \frac{\partial f_{\theta_2}(x)}{\partial \theta_{2j}} - \frac{1}{(f_{\theta_2}(x))^2} f_{\theta_1}(x) d\mu(x) \\ &= -\phi''(1) \int \frac{\partial f_{\theta_1}(x)}{\partial \theta_{1j}} \frac{\partial f_{\theta_2}(x)}{\partial \theta_{2j}} \frac{1}{f_{\theta_1}(x)} d\mu(x) \tag{14} \\ &= - \left\{ \frac{\partial^2 D_\phi(\theta_1, \theta_2)}{\partial \theta_{1i} \partial \theta_{1j}} \right\} \end{aligned}$$

By definition of the Fisher information matrix,

$$\begin{aligned} \frac{\partial^2 D_\phi(\theta_1, \theta_2)}{\partial \theta_{1i} \partial \theta_{1j}} &= \frac{\partial^2 D_\phi(\theta_1, \theta_2)}{\partial \theta_{2i} \partial \theta_{2j}} = I(\theta) \\ \frac{\partial^2 D_\phi(\theta_1, \theta_2)}{\partial \theta_{1i} \partial \theta_{2j}} &= -I(\theta) \end{aligned} \tag{15}$$

Equation (10) reduces to

$$\begin{aligned} &\frac{1}{2}(\hat{\theta}_1 - \theta_1)' I(\theta_0)(\hat{\theta}_1 - \theta_1) + \frac{1}{2}(\hat{\theta}_2 - \theta_2)' I(\theta_0)(\hat{\theta}_2 - \theta_2) \\ &- (\hat{\theta}_1 - \theta_1)' I(\theta_0)(\hat{\theta}_2 - \theta_2) + o(\|\hat{\theta}_1 - \theta_1\|^2) + o(\|\hat{\theta}_2 - \theta_2\|^2) \end{aligned} \tag{16}$$

Further,

$$\begin{aligned} \frac{2}{\phi''(1)} D_\phi(\hat{\theta}_1, \hat{\theta}_2) &= (\hat{\theta}_1 - \theta_1)' I(\theta_0)(\hat{\theta}_1 - \theta_1) + (\hat{\theta}_2 - \theta_2)' I(\theta_0)(\hat{\theta}_2 - \theta_2) \\ &- 2(\hat{\theta}_1 - \theta_1)' I(\theta_0)(\hat{\theta}_2 - \theta_2) + o(\|\hat{\theta}_1 - \theta_1\|^2) + o(\|\hat{\theta}_2 - \theta_2\|^2) \tag{17} \\ &= (\hat{\theta}_1 - \hat{\theta}_2)' I(\theta_0)(\hat{\theta}_1 - \hat{\theta}_2) + o(\|\hat{\theta}_1 - \theta_1\|^2) + o(\|\hat{\theta}_2 - \theta_2\|^2) \end{aligned}$$

From Equation (17) we obtain

$$\frac{\tau(n-\tau)}{n} \left\{ (\hat{\theta}_1 - \hat{\theta}_2)' \widehat{I}(\theta_0)(\hat{\theta}_1 - \hat{\theta}_2) + o(\|\hat{\theta}_1 - \theta_1\|^2) + o(\|\hat{\theta}_2 - \theta_2\|^2) \right\} \tag{18}$$

From 8 and Equations (10)-(18) then the test statistic can be expressed as

$$D_{nr} = \max_{\tau \in N(\varepsilon)} \frac{\tau(n-\tau)}{n} \left\{ (\hat{\theta}_1 - \hat{\theta}_2)' \widehat{I}(\hat{\theta}_0) (\hat{\theta}_1 - \hat{\theta}_2) + o\left(\|\hat{\theta}_1 - \theta_1\|^2\right) + o\left(\|\hat{\theta}_2 - \theta_2\|^2\right) \right\} \quad (19)$$

Let,

$$\max_{\tau \in N(\varepsilon)} \frac{\tau(n-\tau)}{n} (\hat{\theta}_1 - \hat{\theta}_2)' \widehat{I}(\hat{\theta}_0) (\hat{\theta}_1 - \hat{\theta}_2) = W_{nr}$$

Then,

$$\max_{\tau \in N(\varepsilon)} D_{nr} = \max_{\tau \in N(\varepsilon)} W_{nr} + o\left(\|\hat{\theta}_1 - \theta_1\|^2\right) + o\left(\|\hat{\theta}_2 - \theta_2\|^2\right) \quad (20)$$

Consider the second and third terms on the RHS.

$$\begin{aligned} \|\hat{\theta}_1 - \theta_1\|^2 &= (\hat{\theta}_1 - \theta_1)' (\hat{\theta}_1 - \theta_1) \\ \|\hat{\theta}_2 - \theta_2\|^2 &= (\hat{\theta}_2 - \theta_2)' (\hat{\theta}_2 - \theta_2) \end{aligned} \quad (21)$$

for $n \rightarrow \infty$

$$o\left(\|\hat{\theta}_1 - \theta_1\|^2\right) = o_p(1), \quad o\left(\|\hat{\theta}_2 - \theta_2\|^2\right) = o_p(1) \quad (22)$$

The distribution of D_{nr} is similar to that of W_{nr} since the second and third terms of 20 are $o_p(1)$.

The change point estimator is reduced to a trimmed maximal Wald type test statistic. Consider the following conditions [16] [17].

(C1) Regularity: Interchanges of derivative and integral operations be valid so that,

$$\mathbb{E} \left\{ \frac{\partial}{\partial \theta} \log f(x; \theta) \right\} = 0 \quad (23)$$

$$\mathbb{E} \left\{ \frac{\partial^2}{\partial \theta^2} \log f(x; \theta) \right\} = -I(\theta) \quad (24)$$

(C2) for $i, j = 1, \dots, d$

$$\frac{\partial}{\partial \theta_i} f(x; \theta), \quad \frac{\partial^2}{\partial \theta_i \partial \theta_j} f(x; \theta)$$

exist almost everywhere such that

$$\left| \frac{\partial}{\partial \theta_i} f(x; \theta) \right| \leq H_i(x), \quad \left| \frac{\partial^2}{\partial \theta_i \partial \theta_j} f(x; \theta) \right| \leq G_{ij}(x)$$

where

$$\int_{\mathbb{R}^d} H_i(x) < \infty, \quad \int_{\mathbb{R}^d} G_{ij}(x) < \infty$$

i.e. the first and second partial derivatives of $f(x; \theta)$ with respect to θ are bounded by functions with finite integrals.

(C3) for $i, j = 1, \dots, d$

$$\frac{\partial}{\partial \theta_i} \log f(x; \theta), \quad \frac{\partial^2}{\partial \theta_i \partial \theta_j} \log f(x; \theta)$$

exist almost everywhere and are such that the information matrix $I(\theta)$ exists

and is positive definite throughout Θ and is continuous in θ . Hence

$$I^{-1}(\theta) = I^{-1/2}(\theta)' I^{-1/2}(\theta)$$

for some non-singular matrix $I^{-1/2}(\theta)$

A positive definite matrix is always non-singular (determinant $\neq 0$) and the determinant is always positive implying that the matrix is invertible *i.e.*

$I^{-1}(\theta) = \Sigma$ (variance-covariance matrix).

(C4) There are constants a, b such that we can find unique $\hat{\theta}_1, \hat{\theta}_2$ for each $a \leq \tau \leq b$.

(C5) There is an open subset $\Theta_0 \in \Theta \subseteq \mathbb{R}$ containing θ_0 such that $g_i(x; \theta), g_{ij}(x; \theta), 1 \leq i, j, k \leq d$ exist and are continuous in θ for all $x \in \mathbb{R}^d$ and $\theta \in \Theta_0$.

(C6) as $\delta \rightarrow 0$

$$\mathbb{E} \left\{ \sup_{|h| \leq \delta} \left| \frac{\partial^2}{\partial \theta^2} \log f(x; \theta + h) - \frac{\partial^2}{\partial \theta^2} \log f(x; \theta) \right| \right\} = v_\delta \rightarrow 0 \tag{25}$$

Theorem 2. Under the null hypothesis and that conditions C1-C6 hold then the asymptotic distribution of the test statistic is given by,

$$W_{nr} \xrightarrow{D} S_{p,\varepsilon} \tag{26}$$

as $n \rightarrow \infty, p = \dim \theta$. where

$$S_{p,\varepsilon} = \sup_{t \in [\varepsilon, 1-\varepsilon]} B_p^2(t)$$

for $B_p(t) = \frac{1}{t^{(1/2)}(1-t)^{(1/2)}} \left\{ (W_p(t))' (W_p(t)) \right\}^{(1/2)}$ such that $W_p(t)$ is a p -dimensional Brownian bridge process.

The following results hold for approximation of the distribution function using the inverted Laplace transformation,

$$\begin{aligned} & P \left(\sup_{\varepsilon < t < 1-\varepsilon} B_d^2(t) \geq u \right) \\ &= \frac{1}{\Gamma(d/2)} \left(\frac{u}{2} \right)^{d/2} e^{-u/2} \left(\log \left(\frac{(1-\varepsilon)^2}{\varepsilon^2} \right) \left(1 - \frac{d}{u} \right) + \frac{2}{u} + O(u^{-2}) \right) \end{aligned} \tag{27}$$

for proof of this see [18]. Rather than considering a fixed trimming value (ε) for all sample sizes, the approach of [4] [11] is followed such that the trimming parameter is a function of the sample size n .

Critical Values

At any given level of significance, the asymptotic critical values of the test can then be estimated. Depending on the dimension of the parameter space (d), the critical values can be estimated. For a bi-variate parameter space *i.e.* $d = 2$, $\alpha \in (10\%, 5\%, 1\%)$, the asymptotic critical values are presented in **Table 1** for different sample sizes such that $n \in \{50, 200, 500, 1000\}$.

Table 1. Asymptotic critical values.

Sample size	$1 - \alpha$	CV
50	0.90	9.954
	0.95	11.715
	0.99	15.569
200	0.90	11.115
	0.95	12.835
	0.99	16.638
500	0.90	11.676
	0.95	13.377
	0.99	17.157
1000	0.90	12.034
	0.95	13.725
	0.99	17.490

4. Simulation Study

In this section change point estimation for the generalized Pareto distribution is considered. For any given finite set of data, at least one of the following is likely at any given change point τ : ξ changes by a non-zero quantity; σ changes by a non-zero quantity; both ξ and σ change by non-zero quantities. A simple change point problem can be formulated in the following way;

$$\begin{aligned} H_0 : X_t &\sim GP(\xi_1, \sigma_1) : \text{for } t = 1, \dots, n \\ H_1 : X_t &\sim GP(\xi_1, \sigma_1) : \text{for } t \leq \tau \\ &X_t \sim GP(\xi_2, \sigma_2) : \text{for } t > \tau \end{aligned} \quad (28)$$

where $\xi_1 \neq \xi_2, \sigma_1 \neq \sigma_2$

Assumption: This work assumes that both parameters of the GP distribution (shape and scale) change at the same time.

Let $\phi(t) = -\log(t)$. The resulting divergence is the Kullback Leibler (KL) divergence. The KL divergence between two GP distributions is given as

$$D_{\{KL\}}(\hat{\theta}_2; \hat{\theta}_1) = \log\left(\frac{\hat{\sigma}_1}{\hat{\sigma}_2}\right) - (1 + \hat{\xi}_2) + \left(\frac{1}{\hat{\sigma}_1} + \frac{\hat{\xi}_1}{\hat{\sigma}_1}\right) \int \left(1 + \frac{\hat{\xi}_2}{\hat{\sigma}_2} x\right)^{-\frac{1}{\hat{\xi}_2}} \left(1 + \frac{\hat{\xi}_1}{\hat{\sigma}_1} x\right)^{-1} \quad (29)$$

The KL divergence is a function of the parameters of two densities (before and after the change point).

The change point estimator thus becomes

$$D_{\tau, n} = \max_{\epsilon < \tau < (1-\epsilon)n} \left(\frac{\tau}{n} \left(1 - \frac{\tau}{n}\right)\right) \frac{2}{\phi''(1)} D_{KL}(\hat{\theta}_1; \hat{\theta}_2) \quad (30)$$

For the simulation study, the following model is considered under the alternative hypothesis;

$$\begin{aligned} H_1 : X_t &\sim GP(1, 0.1) \quad \text{for } t \leq \tau \\ &X_t \sim GP(3, 0.35) \quad \text{for } t > \tau \end{aligned} \quad (31)$$

The change point τ is fixed at $n/2$ for $n = 100, 200, 500$ and 1000. For a 5%

level of significance the change point is estimated and the results are presented in **Figures 1-3**. The change point process $D_{n,\tau}$ takes a hill shape with the peak being observed at the point where the detected change point lies. To estimate the change time, the estimated critical value is superimposed on the graph and the change points taken as the maximum value exceeding the critical bound (critical value at α level of significance). The estimated time of change is also superimposed on the respective plots of the time series data.

Figure 1 shows the evolution of the change point test statistic on the left panel and the time series data on the right panel. The graph of hypothesis testing gives $D_{\{n\tau\}} = 24.86981$. $D_{\{n\tau\}} > C_{0.05} = 12.835$. H_0 is therefore rejected and it is concluded that a change point exists in the time series data. The largest divergence is estimated at $\hat{\tau} = 88$ against the actual true change point $\hat{\tau} = 100$. **Figure 2** of hypothesis testing has $D_{\{n\tau\}} = 18.525111$. $D_{\{n\tau\}} > C_{0.05} = 13.3777$. H_0 is therefore rejected and a change point is declared in the time series. The largest divergence is estimated at $\hat{\tau} = 245$ against the actual true change point of $\tau = 250$. **Figure 3** has the evolution of the change point process on the top panel and the time series data on the bottom panel. $D_{n\tau} = 19.42708$ which rejects H_0 at $\alpha = 5\%$ level of significance. The estimated change point is at $\hat{\tau} = 494$ against the true change point $\tau = 500$.

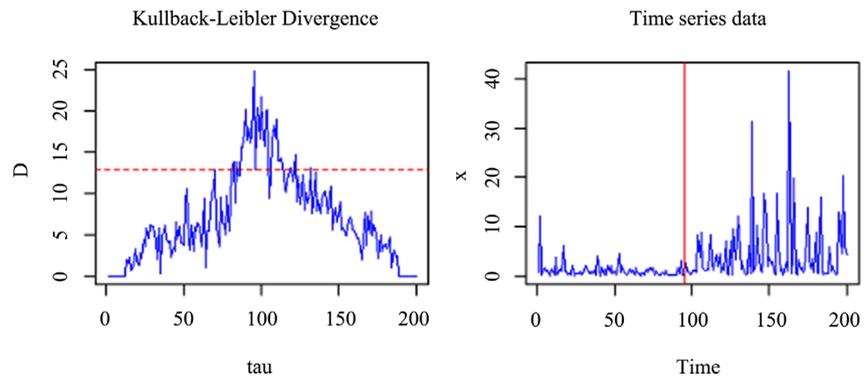


Figure 1. Change point test process (left panel) and time series data (right panel) for $n = 200$ and $\tau = n/2$.

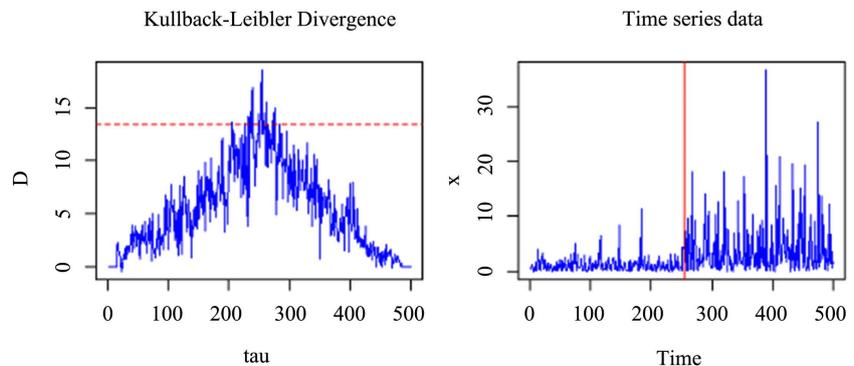


Figure 2. Change point test process (left panel) and time series data (right panel) for $n = 500$ and $\tau = n/2$.

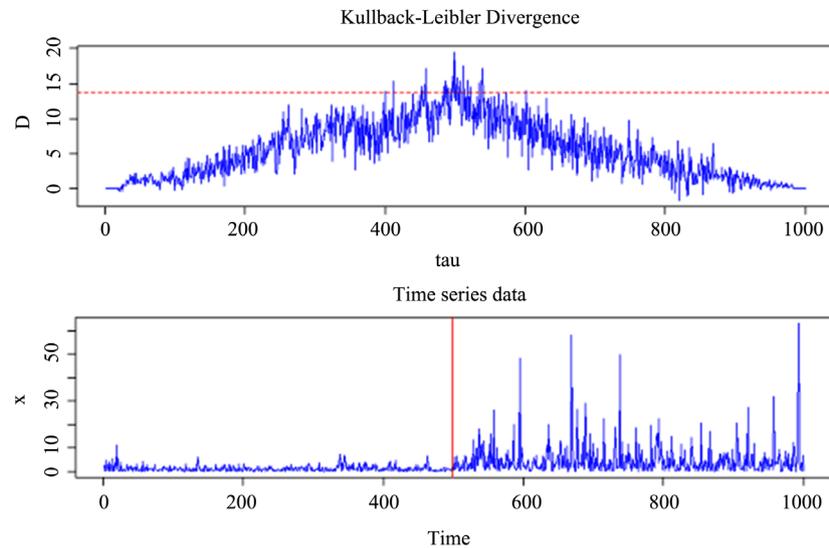


Figure 3. Change point test process (left panel) and time series data (right panel) for $n = 1000$ and $\tau = n/2$.

The change point estimator is examined when the true change point is no longer located in the middle of the sample size but is fixed towards the boundary points at $n/3$ for $n = 200, 500, 1000$. The results are shown in **Figures 4-6**.

Figure 4 shows the evolution of the change point test statistic on the left panel and the time series data on the right panel. The graph of hypothesis testing gives $D_{nr} = 13.9438 > C_{0.05} = 12.835$. H_0 is therefore rejected and it is concluded that a change point exists at the respective level of significance. The largest divergence is estimated at $\hat{\tau} = 70$ against the actual true change point $\tau = 66$. **Figure 5** of hypothesis testing has $D_{nr} = 16.7829 > C_{0.05}$. H_0 is therefore rejected and a change point is declared. The largest divergence is estimated at $\hat{\tau} = 164$ against the actual true change point of $\tau = 166$. **Figure 6** has $D_{nr} = 14.6752$ which rejects H_0 at 5% level of significance. The estimated change point is at $\hat{\tau} = 330$ against the true change point $\tau = 333$.

The asymptotic power of the change point test is examined. The most commonly used criteria for checking the optimality of a statistical test involves fixing the false alarm probability (type I error) and maximizing the detection probability (minimizing the type II error). The power of the test at a given level α against a particular alternative is defined as the probability of rejecting the null hypothesis when the alternative is actually true.

$$p(\alpha) = P(D_{nr} > C_\alpha | H_1) \quad (32)$$

For power or sample-size computation, not only the distribution of the test statistic under the null hypothesis needs to be obtained but also its distribution under the alternative hypothesis. This is beyond the scope of this work hence reliance on simulation results. D_{nr} is estimated for 1000 replicates of simulated data using defined sample size $n = 500$. The behavior of the test as the change point approaches the data boundary points is analyzed using the power function such that.

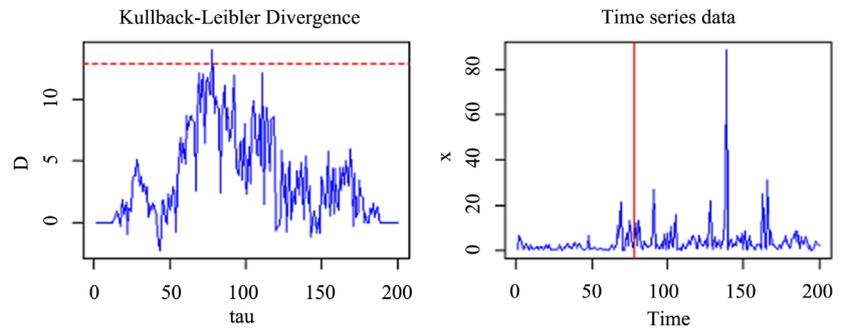


Figure 4. Change point test process (left panel) and time series data (right panel) for $n = 200$ and $\tau = n/3$.

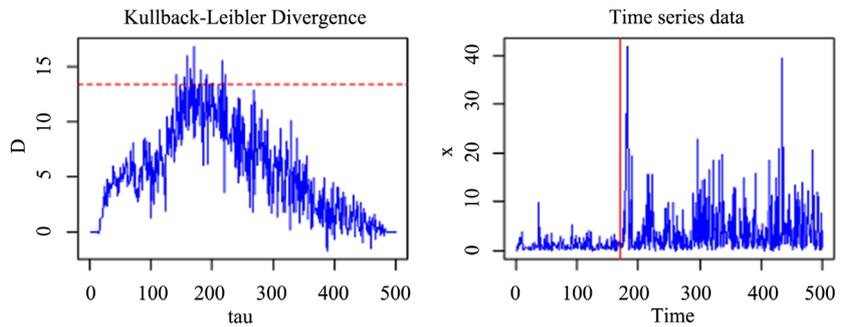


Figure 5. Change point test process (left panel) and time series data (right panel) for $n = 500$ and $\tau = n/3$.

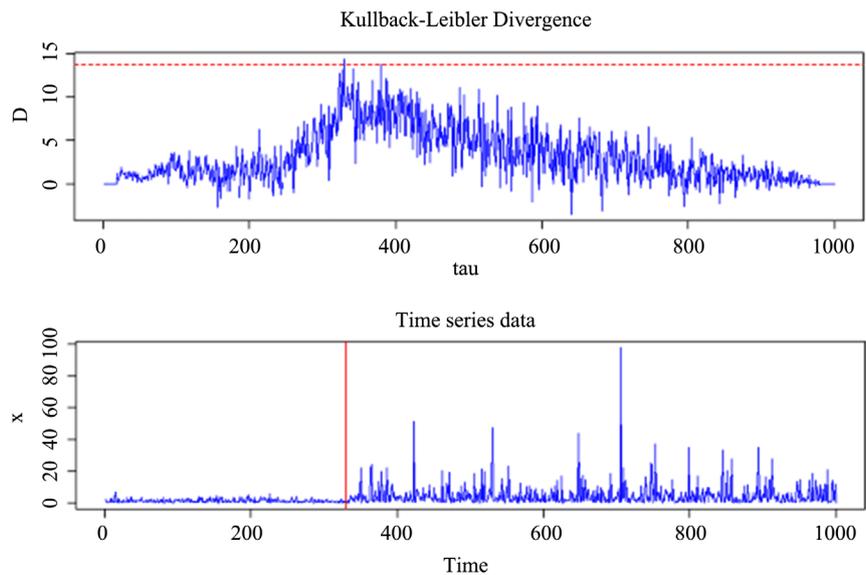


Figure 6. Change point test process (left panel) and time series data (right panel) for $n = 1000$ and $\tau = n/3$.

$$\hat{p}(\alpha) = (1 + \#(D_{n\tau} > C_\alpha)) / (1 + N) \tag{33}$$

The results in **Table 2** indicate that the change point test is most powerful when the change point is located at the middle of the data set and less powerful

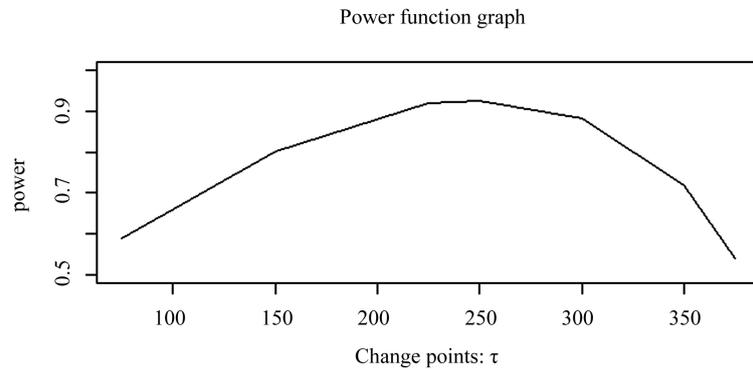


Figure 7. The 95% power function for $n = 500$.

Table 2. Change point power estimates of the test with $n = 500$.

τ	75	150	225	250	300	350	375
$\hat{p}(\alpha)$	0.5894	0.8022	0.9191	0.9250	0.8821	0.7183	0.5395

when the change point is located towards the boundary points. This behavior is attributed to comparing an estimate computed from a small sample (say the first τ observations) to one computed from a larger sample (say the last $n - \tau$ observations). Small sample sizes result to erratic behavior of the test statistic due to instability of the parameter estimates. This implies that the test is more likely to incorrectly reject a change point when it is located towards the boundary points of any given data set (This is shown in **Figure 7** that is the power function graph).

5. Conclusion

In this work, a divergence based (pseudo-distance) estimator has been used for estimating change in the parameters of any given parametric distribution. The change-point estimator is the first point at which there is maximal sample evidence of a change characterized by maximum divergence exceeding a critical bound. By application of the likelihood standard regularity conditions the distribution of the pseudo-metric based estimator is found to converge to that of a Brownian bridge process on a given interval. The distribution of the pseudo-distance based change point process is found to be similar to that of a maximal trimmed Wald-type test statistic under the null hypothesis of no change point. The distribution does not depend on the choice of the function ϕ and this is therefore applicable within a parametric framework when using other choices of the function ϕ for other statistical divergences. Further work can be done on the theoretical power properties of this particular change point estimator.

Acknowledgements

The first author thanks the Pan-African University Institute of Basic Sciences, Technology and Innovation (PAUSTI) for funding this research.

Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

References

- [1] Truong, C., Oudre, L., & Vayatis, N. (2018). A Review of Change Point Detection Methods. arXiv preprint arXiv:1801.00718.
- [2] Brodsky, E. and Darkhovsky, B.S. (2013) Nonparametric Methods in Change Point Problems. Vol. 243, Springer Science & Business Media, Berlin.
- [3] Korkas, K.K. and Fryzlewicz, P. (2017) Multiple Change Point Detection for Non-Stationary Time Series Using Wild Binary Segmentation. *Statistica Sinica*, **27**, 287-311. <https://doi.org/10.5705/ss.202015.0262>
- [4] Csorgo, M. and Horváth, L. (1997) Limit Theorems in Change-Point Analysis. Vol. 18, John Wiley & Sons, Hoboken.
- [5] Cheng, L., AghaKouchak, A., Gilleland, E. and Katz, R.W. (2014) Non-Stationary Extreme Value Analysis in a Changing Climate. *Climate Change*, **127**, 353-369. <https://doi.org/10.1007/s10584-014-1254-5>
- [6] Jarusková, D. and Rencová, M. (2008) Analysis of Annual Maximal and Minimal Temperatures for Some European Cities by Change Point Methods. *Environmetrics*, **19**, 221-223. <https://doi.org/10.1002/env.865>
- [7] Dupuis, D., Sun, Y. and Wang, H.J. (2015) Detecting Change Point in Extremes. *Statistics and Its Interface*, **8**, 19-31. <https://doi.org/10.4310/SII.2015.v8.n1.a3>
- [8] Dette, H. and Wu, W. (2018) Change Point Analysis in Non-Stationary Processes—A Mass Excess Approach. arXiv: 1801.09874.
- [9] Killick, R. and Ekcley, I (2014) An R Package for Change Point Analysis. *Journal of Statistical Software*, **58**, 1-19. <https://doi.org/10.18637/jss.v058.i03>
- [10] Page, E. (1955) A Test for a Change in a Parameter Occurring at an Unknown Point. *Biometrika*, **42**, 523-527. <https://doi.org/10.1093/biomet/42.3-4.523>
- [11] Gichuh, A.W. (2008) Nonparametric Change Point Analysis for Bernoulli Random Variables Based on Neural Networks.
- [12] Dierckx, G. and Teugels, J.L. (2010) Change Point Analysis of Extremes. *Environmetrics*, **21**, 661-686. <https://doi.org/10.1002/env.1041>
- [13] Pardo, L. (2018) Statistical Inference Based on Divergence Measures. Chapman and Hall/CRC, New York. <https://doi.org/10.1201/9781420034813>
- [14] Andrews, D.W. (1993) Tests for Parameter Instability and Structural Change with Unknown Change Point. *Econometrica*, **61**, 821-856. <https://doi.org/10.2307/2951764>
- [15] Horváth, L., Miller, C. and Rice, G. (2019) A New Class of Change Point Test Statistics of Rényi Type. *Journal of Business & Economic Statistics*, **38**, 570-579. <https://doi.org/10.1080/07350015.2018.1537923>
- [16] Hawkins Jr., D.L. (1983) Sequential Detection Procedures for Autoregressive Processes. Technical Report, North Carolina State University, Raleigh.
- [17] Batsidis, A., Martin, N., Pardo, L. and Zogfaros, K. (2016) ϕ -Divergence Based Procedure for Parametric Change Point Problems. *Methodology and Computing in Applied Probability*, **18**, 21-35. <https://doi.org/10.1007/s11009-014-9398-3>
- [18] Estrella, A. (2003) Critical Values and P Values of Bessel Processes Distributions: Computation and Application to Structural Break Tests. *Econometric Theory*, **19**, 1128-1143. <https://doi.org/10.1017/S0266466603196107>