**International Journal of Applied Mathematics**
**Volume 38 No. 11s, 2025**
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

# CHANGE POINT DETECTION USING SKEWED DISTRIBUTION IN THE FRAMEWORK OF MARSHALL – OLKIN GENERALIZED DISTRIBUTION

**T.Sukeerthi[1*], Dr.A.Jyothi Babu[2]**

[1*]Department of Statistics, Mohan Babu University, Tirupati, Andhra Pradesh, India.
tptsukeerthi2309@gmail.com
[2]Department of Computer Applications, Mohan Babu University, Tirupati, Andhra Pradesh, India.
jyothibabu.a@vidyanikethan.edu

**Abstract**

{A novel change point detection framework based on the Marshall – Olkin Generalized Distribution integrates parametric and non-parametric methods, improving detection under skewed and heavy – tailed data.}

This study proposes a novel framework for Change Point Detection (CPD) using the Marshall-Olkin Generalized Distribution (MOGD), a flexible distribution family that accommodates skewness and heavy tailed behaviour common features in real world data. Traditional CPD methods often rely on symmetric or normal distribution assumptions, which may fail to detect structural changes accurately in such contexts. The proposed approach integrates both parametric and non-parametric techniques, employing the Likelihood Ratio Test and Bayesian inference to identify change points in time series data modeled by MOGD. This dual method strategy enhances sensitivity to shifts in distributional behaviour, especially in datasets with asymmetry or extreme values. The framework is validated through simulation studies and real-world applications, demonstrating its robustness and improved accuracy over classical CPD methods. By addressing limitations of traditional models, this work introduces a more adaptive and statistically sound methods for detecting distributional changes in complex time series.

**Keywords:** Change Point Detection, Statistical Process Control, Marshall- Olkin Distribution, Skewed Distributions, Likelihood Ratio Test, Modified Information Criterion.

## 1. Introduction

Change Point Detection (CPD) plays a vital role in identifying structural changes in time series data, with broad applications in environmental studies, finance, genetics and quality control. A change point represents a time instance where the statistical properties of the data undergo a significant shift, such as in mean, variance or distribution shape.

Classical CPD approaches often assume that data follow symmetric or normal distributions. However, real world data frequently deviate from these assumptions, exhibiting skewness or heavy tails. This limitation has motivated the use of more flexible models and criteria, such as the Akaike Information Criterion (AIC), Bayesian Information Criterion (BIC) and Modified Information Criterion (MIC), to improve detection performance.

Despite the progress, limited attention has been given to CPD using skewed distributions, particularly under the Marshall –Olkin Generalized Distribution (MOGD) framework. MOGD offers greater flexibility in modelling asymmetry and tail behaviour. This study introduces a new CPD methodology by leveraging MOGD and integrating it with both likelihood based and Bayesian techniques. This hybrid approach enables the detection of subtle distributional changes, enhancing CPD performance in data environments where traditional assumptions fail.

Through simulation and application to real data, the proposed method demonstrates improved accuracy and robustness, offering a significant advancement in CPD methodology.

Received: August 07, 2025

International Journal of Applied Mathematics
Volume 38 No. 11s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

## 2. Related Work

The study of change point detection (CPD) has evolved significantly since Page's (1954) pioneering contribution, which introduced a simple method to detect a single change point. This work laid the foundation for subsequent developments that aimed to identify structural changes in data sequences. Chernoff and Zacks (1964) extended CPD techniques by focusing on changes in the mean of normally distributed data. Gardner (1969) and Sen and Srivastava (1975) expand this work to account for variance, enabling the detection of changes in both mean and variability.

Worsley (1979) introduced further enhancements to detect multiple change points, particularly within normal distributions. In the following decades, researchers developed information based model selection techniques to improve accuracy. Notably, Yao (1988) introduced the Bayesian Information Criterion (BIC), followed by Jones and Dey (1995), who refined the Akaike Information Criterion (AIC) for multiple change points. Chen and Gupta (1997) applied BIC under fixed mean assumptions to address data uncertainty.

Later developments explored alternative distributions. Jandhyala and Fotopoulos (1999) used the Weibull distribution, while Ninomiya (2005) extended CPD to the Gaussian model. Chen et at. (2006) introduced the Modified Information Criterion (MIC), enhancing detection through refined penalty terms. Zhang and Siegmund (2007) applied MIC to high-dimensional genomic data.

More recent work by Said et al. (2019) and Cai et al. (2016,2017) applied MIC to skewed and exponential distributions, highlighting its utility in reliability and failure rate models. Bayesian methods have also gained prominence for their flexibility and robustness in identifying multiple change points under model uncertainty.

Despite these advancements, limited attention has been paid to CPD using skewed distributions such as the Marshall-Olkin Generalized Distribution (MOGD), which this study aims to address.

## 3. Marshall-Olkin Exponential Lomax distribution

Bayesian methods have gained prominence due to their flexibility and robustness in identifying multiple change points under model uncertainty.

Let `X' be a random variable following the Exponential Lomax distribution with parameters α (shape), β (scale), and λ (location). The cumulative distribution function (CDF) and probability density function (PDF) of the Exponential Lomax distribution (denoted as X ~ $E_X L(\alpha,\beta,\lambda)$) are given by El-Bassiony et al. (2015) as:

$$F_{E_X L}(x) = 1 - exp\left(-\lambda\left(\frac{\beta}{x+\beta}\right)^{-\alpha}\right)$$

$$f_{E_X L}(x) = \frac{\lambda\alpha}{\beta}\left(\frac{\beta}{x+\beta}\right)^{-\alpha+1} exp\left(-\lambda\left(\frac{\beta}{x+\beta}\right)^{-\alpha}\right);x \geq -\beta, (\alpha,\beta,\lambda) > 0$$

Several extensions to the EXL distribution have been proposed in the literature, including the Transmuted Exponential Lomax by Ab-dullahi and Ieren [2018] and Exponentiated Exponential Lomax by Adeleke et al. [2019].

The general form of the Marshall-Olkin family of distributions is defined as:

$$F(x) = \frac{G(x)}{1 - \bar{\theta}\bar{G}(x)}$$

and

$$f(x) = \frac{\theta g(x)}{[1 - \bar{\theta}\bar{G}(x)]^2};\ -\infty < x < \infty$$

Received: August 07, 2025 

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

Where $G(x)$ and $g(x)$ are the baseline CDF and PDF respectively, and $\bar{G}(x) = 1 - G(x)$ substituting the Exponential Lomax functions into these expressions gives rise to the Marshall-Olkin Exponential Lomax (MOELX) distribution with the following CDF and PDF:

$$F_{MOE_XL}(x) = \frac{1 - exp\left(-\lambda\left(\frac{\beta}{x+\beta}\right)^{-\alpha}\right)}{1 - \theta exp\left(-\lambda\left(\frac{\beta}{x+\beta}\right)^{-\alpha}\right)}$$

and

$$f(x) = \frac{\frac{\theta\lambda\alpha}{\beta}\left(\frac{\beta}{x+\beta}\right)^{-\alpha+1} \exp -\alpha\left(\frac{\beta}{x+\beta}\right)^{-\alpha}}{\left[1 - \bar{\theta}\exp -\alpha\left(\frac{\beta}{x+\beta}\right)^{-\alpha}\right]^2} \; ; x > -\beta, (\alpha, \beta, \theta, \lambda) > 0$$

## 4. Application of Change Point Methodology

To clarify the null and alternative hypotheses in the context of change point detection (CPD), we define them as follows:

**Null Hypothesis (H0):**

- The sequence of observations follows a single, unchanged distribution without any change points.
- There is no change point in the sequence, meaning the data comes from a homogeneous distribution throughout.
-

**Alternative Hypothesis (H1):**

- The sequence of observations has at most one change point, implying a shift in the distribution at one point in the data.
- The data is consistent with the presence of at most one change point, meaning the distribution may differ before and after that point.

These hypotheses define the framework for testing whether a change point exists.
Specifically:

$$H_0: \xi_1 = \xi_2 = \cdots = \xi_n \; Vs \; H_1: \xi_1 = \xi_2 = \cdots = \xi_k \neq \xi_{k+1} = \cdots = \xi_n$$

In change point detection, the sequence of observations $\xi_i; i = 1,2, \dots, n$ follows a specific distribution, and each observation has a corresponding parameter $\theta_i$. The primary goal is to identify the **presence** and **location** of any change points. Under the hypothesis testing framework:

- $H_0$ assumes no change points exist.
- $H_1$ allows for the presence of at most one change point.

If a change point is detected, we accept $H_1$; otherwise, we accept $H_0$.
Assuming a change point occurs at position $k^{th}$, the data can be divided into two segments:

- The first segment includes observations from 1 to k ,
- The second segment includes observations from (k+1) to n.

Detecting a change point involves identifying the position k that maximizes the likelihood under the assumption that a distributional change occurs at that point.
Two widely used methods to address this problem are:

- Likelihood Ratio Test (LRT)
- Bayesian procedures

These methods evaluate whether there is statistical evidence for a change point at a specific location in the sequence.

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

In addition to hypothesis testing methods, model selection criteria such as the **Akaike Information Criterion (AIC)** and **Bayesian Information Criterion (BIC)** are commonly employed. These criteria help identify the model that best balances data fit and model complexity, thus avoiding overfitting. In CPD context they are used to compare models with and without change points, facilitating the optimal model selection.

Foundational work on CPD was carried out by **Chen and Gupta (1997),** focusing on changes in variance in normally distributed data using BIC. Later, **Chen et al. (2006)** observed that BIC's penalty term was not well suited to change point problems and proposed the **Modified Information Criterion (MIC)**. MIC refines the penalty term to better account for the complexity introduced by change points, improving model selection accuracy.

## 5. The Binary Segmentation Algorithm

Binary segmentation, introduced by Vostrikova (1981), is a widely used procedure for detecting multiple change points in a sequence. It simplifies the complex task of identifying several changes by breaking it into a series of steps, each aimed at detecting a single change point. This approach is particularly effective when the number and locations of change points are unknown.

The process begins by testing whether the data fits a specified distribution. Initially, the entire dataset is analyzed to detect a potential change point. If a change is identified, the dataset is split into two segments: one before the change point and one after. The same procedure is then applied recursively to each segment to detect further changes.

At every step, the algorithm evaluates whether the data within the current segment aligns well with the chosen distribution. If a deviation is detected, indicating a change point, the segment is divided, and the binary segmentation process continues on each resulting sub-segment. This iterative process is repeated until no further changes are found or a pre-defined stopping criterion is met.

Binary segmentation is highly valued for its simplicity, computational efficiency and effectiveness especially when the number of change points is relatively small.

The steps of the binary segmentation algorithm are as follows:

**Step 1:**

Define the null hypothesis $H_0$ (no change point) and the alternative hypothesis $H_1$ (presence of a change point). Let $X \sim f_x(.)$ denote the data. The log-likelihoods under each hypothesis are:

- Under $H_0$: $\log L(X, \xi)$
- Under $H_1$: $\log L(X, \xi', \xi'')$

Here, $\xi$ and $\xi'$ represent the parameter sets for the two segments.

**Step 2:**

Calculate the Modified Information Criterion (MIC) for both hypotheses.

- $\text{MICH}_0 + m \log n$
- $\text{MICH}_1 + \left(2m + \left(\frac{2m}{n} - 1\right)^2\right) \log n$

Where:

- m: number of parameters
- n: number of observations in the segment.

**Step-3:**

Compare the MIC Values. If $MIC_n < \min_{1 \le k < n} MIC_k$, then accept $H_0$ (no change point). Otherwise, accept $H_1$, indicating the presence of a change point at the location corresponding to the minimum MIC value. Split the segment at this point.

**Step-4:**

Repeat the above steps on each resulting segment until no further change points are detected (i.e.,$H_0$ is accepted for all segments).

International Journal of Applied Mathematics
Volume 38 No. 11s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

In the following sections, we apply this binary segmentation procedure to data modelled using skewed distributions.

## 6. Illustrations

The practical utility of the proposed Change Point detection method based on the MOELX distribution is demonstrated using two real datasets: (1) gauge length measurements and (2) annual flood discharge data.

### 6.1 Guage Length Data

We consider gauge length measurements (20mm and 101mm) collected from a single fiber, as presented in Bader and priest (1982). The dataset consists of 40 samples observations, including values such as: 1.6, 2.0, 2.6, 3.0, 3.5, 3.9, 4.5, 4.6, 4.8, 5.0, 5.1, 5.3, 5.4, 5.6, 5.8, 6.0, 6.0, 6.1, 6.3, 6.5, 6.5, 6.7, 7.0, 7.1, 7.3, 7.3, 7.3, 7.7, 7.7, 7.8, 7.9, 8.0, 8.1, 8.3, 8.4, 8.4, 8.5, 8.7, 8.8, and 9.0. Applying our binary segmentation algorithm yields the following results:
The stopping condition $MIC_n > \min_k (MIC_{k})$ is satisfied (175.8928 > 139.9510) indicating the existence of a change point. As shown in the Fig.1 the results are given below.

1. **First change point**: k = 22, splitting the dataset into two segments of 22 and 18 observations.
2. **Second change point**: Within the first segment (1-22), a change point is identified at k=6 ($MIC_k$ = 73.9435).
3. **Third change point:** Within the second segment (23-40), a change point occurs at position k=31 relative to the full dataset ($MIC_k$ = 37.2057).
4. **Fourth change point:** A further change point is detected between positions 7 and 22, specifically at k = 13 ($MIC_k$ = 34.2534).
5. No additional change points are found beyond this level of segmentation.

**Summary of Detected Change Points:**

- k = 22 (approximately 6.7mm)
- k = 6 (approximately 3.9mm)
- k = 31 (approximately 7.9mm)
- k = 13 (approximately 5.4mm).

**Figure 1** illustrates the Minimum Information Criterion (MIC) plots for each stage of segmentation, with vertical lines indicating the detected change points.

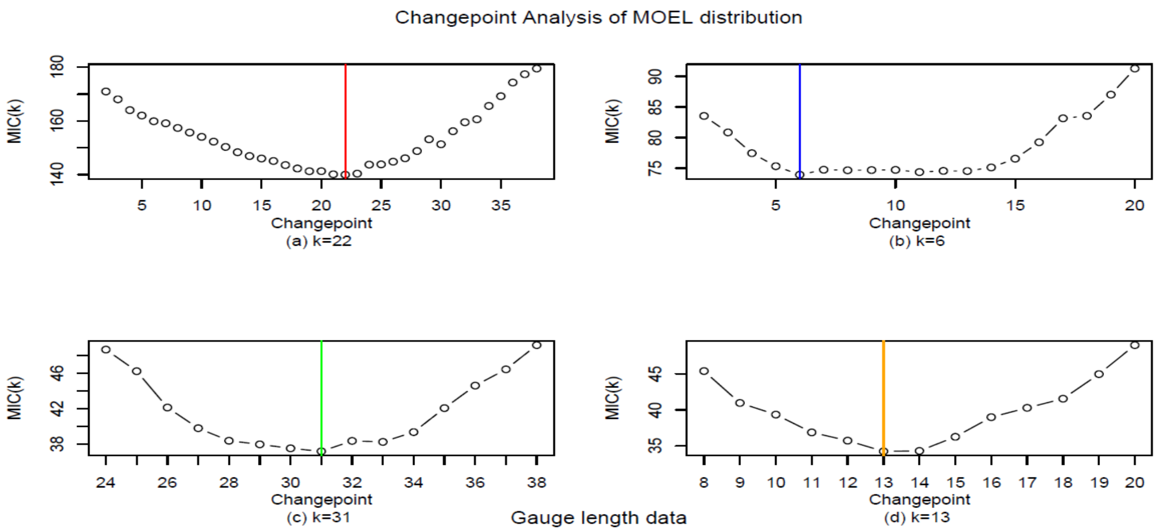


**Figure 1: MIC plots for guage length data with detected Change Points using the MOELX based method.**

### 6.2 Flood Discharge Data (Floyd River)

The second dataset comprises the annual flood discharge rates (in $ft^3/s$) of the Floyd River, recorded every 10 years from 1935 to 1973 (Mudholkar & Hutson, 1996). A total of 39 observations are analyzed, grouped by decade in Table 1.

**Table 1**. Floyd River Annual Flood Discharge Data (in $ft^3$ /s)

| **Years** | **Flood Discharge ($ft^3/s$)** |
|---|---|
| **1935-1944** | 1460, 4050, 3570, 2060, 1300, 1390, 1720, 6280, 1360, 7440 |
| **1945-1954** | 5320, 1400, 3240, 2710, 4520, 4840, 8320, 13900, 71500, 6250 |
| **1955-1964** | 2260, 318, 1330, 970, 1920, 15100, 2870, 20600, 3810, 726 |
| **1965-1973** | 7500, 7170, 2000, 829, 17300, 4740, 13400, 2940, 5660 |

Applying the binary segmentation procedure:

- The stopping condition $MIC_n > \min\{MIC_k\ (1 \le k < n)\}$ is satisfied (774.1057 > 744.1057).
- **First change point:** k = 15, dividing the dataset into two segments of 15 and 24 observations.
- **Second change point:** within the first segment (1-15), a change point is found at k=4 (relative to full data; $MIC_K = 237.2374$)
- **Third change point:** within the second segment (16-39), a change occurs at k=33 (relative to full data; $MIC_K = 476.6589$).
- **Fourth change point:** Between positions 16 and 33,a further change point is identified at k = 22 ($MIC_k = 322.5852$).
- No further change points are detected.

**Summary of Detected Change Points:**

- k = 15 (approximately 4520 $ft^3/s$)
- k = 4 (approximately 2060 $ft^3/s$)
- k = 33 (approximately 2000 $ft^3/s$)
- k = 22 (approximately 318 $ft^3/s$).

**Figure 2** displays the MIC plots with marked vertical lines denoting each change point.

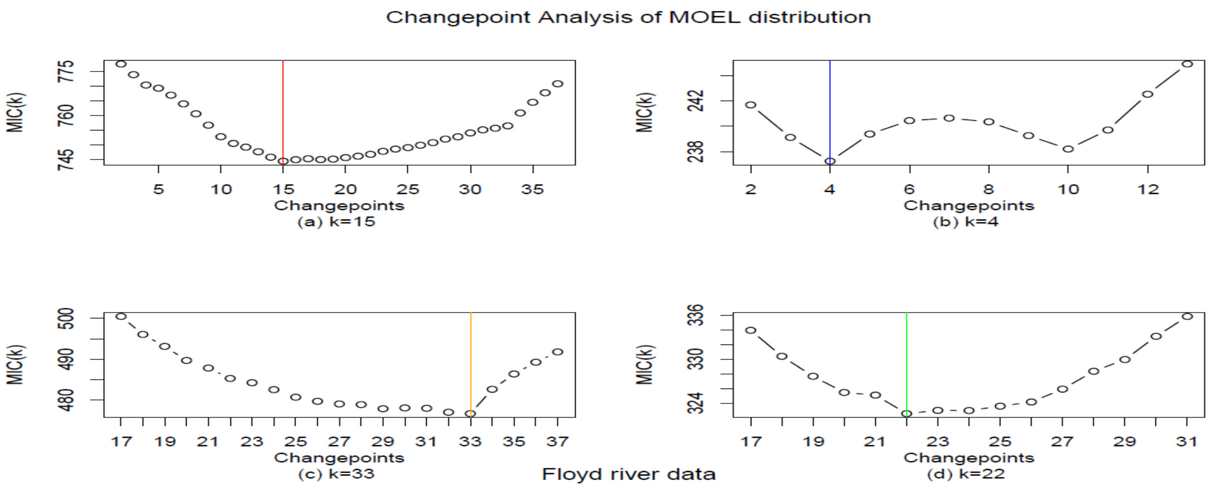


**Figure 2: MIC plots for Floyd River flood discharge data with detected change points.**

**Conclusion:**

The binary segmentation method, when integrated with the Modified Information Criterion (MIC), provides a robust and efficient approach for detecting change points across various types of data.

Received: August 07, 2025 

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

This methodology proves particularly effective in identifying significant shifts in the distribution of data, whether it involves physical measurements such as gauge lengths of wire or environmental parameters like flood discharge rates.

Through the analysis of the Gauge Length Data and Floyd River Data, the practical relevance of change point detection is clearly demonstrated. In both datasets, multiple change points were identified, signifying substantial shifts. These changes serve as indicators of underlying transformations in the data generating processes, such as variations in manufacturing conditions for the wires or climatic and environmental affecting flood patterns.

Overall, change point detection emerges as a valuable tool across diverse research and application domains. It enables a clearer understanding of when and where significant changes occurs within time series or sequential data. The combination of MIC and binary segmentation offers an interpretable and statistically sound framework that facilities data driven decision-making and deeper insights into dynamic systems.

**Conflict of Interest:** “The authors declare that they have no conflict of interest”.

**References:**

1. Abdullahi, U. K., & Ieren, T. G. (2018). On the inferences and applications of transmuted exponential lomax distribution. *Int. J. of Adv. Prob. Stat, 6(1):30-36.*
2. Adeleke, M., Ayodele, A. A., and Barnabas, A. F. (2019). Exponentiated exponential lomax distribution and its properties. *Mathematical Theory and Modeling, 9(1):1-13.*
3. Bader, M. and Priest, A. (1982). Statistical aspects of fibre and bundle strength in hybrid composites. *Progress in science and engineering of composites, 1129-1136.*
4. Cai, X., Said, K. K., and Ning, W. (2016). Change-point analysis with bath-tub shape for the exponential distribution. *Journal of Applied Statistics, 43(15):2740-2750.*
5. Cai, X., Tian, Y., and Ning, W. (2017). Modified information approach for detecting change points in piecewise linear failure rate function. *Statistics & Probability Letters, 125:130-140.*
6. Chen, G. and Balakrishnan, N. (1995). A general purpose approximate goodness-of-_t test. *Journal of Quality Technology, 27(2):154-161.*
7. Chen, J., Gupta, A., and Pan, J. Information criterion and change point problem for regular models. *Sankhy_a: The Indian Journal of Statistics, 252-282, 2006.*
8. Chernoff, H. and Zacks, S. (1964). Estimating the current mean of a normal distribution which is subjected to changes in time. *The Annals of Mathematical Statistics, 35(3):999-1018.*
9. Erdman, C. and Emerson, J. W. (2008). A fast bayesian change point analysis for the segmentation of microarray data. *Bioinformatics, 24(19):2143-2148.*
10. Gardner, L. (1969). On detecting changes in the mean of normal variates. *The Annals of Mathematical Statistics, 40(1):116-126.*
11. Jandhyala, B. and Fotopoulos, S. B. (1999). Capturing the distributional behaviour of the maximum likelihood estimator of a changepoint. *Biometrika,86(1):129-140.*
12. Jones, R. H. and Dey, I. (1995). Determining one or more change points. *Chemistry and Physics of LIPIDS, 76(1):1-6.*
13. Mudholkar, G. S. and Hutson, A. D. (1996). The exponentiated weibull family: some properties and flood data application. *Communications in Statistics Theory and Methods, 25(12):3059-3083.*
14. Ngunkeng, G. and Ning, W. (2014a). Information approach for the change-point detection in the skew normal distribution and its applications. *Sequential Analysis, 33(4):475-490.*
15. Ning, W. and Gupta, A. K. (2009). Change point analysis for generalized lambda distribution.*Communications in Statistics-Simulation and Computation, 38(9):1789-1802.*
16. Ninomiya, Y. (2005). Information criterion for gaussian change-point model. *Statistics & probability letters, 72(3):237-247.*

**International Journal of Applied Mathematics**
**Volume 38 No. 11s, 2025**
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

17. Page, E. S. Continuous inspection schemes. *Biometrika, 41(1/2):100-115, 1954.*
18. Rashwan, N. I. (2018). Moments of order statistics from nonidentically distributed lomax, exponential lomax and exponential pareto variables. *Journal of Advances in Math, 14(1):7431-7438.*
19. Said, K. K., Ning,W., and Tian, Y. (2019). Modified information criterion for testing changes in skew normal model. *Brazilian Journal of Probability and Statistics, 33(2):280-300.*
20. Sen, A. and Srivastava, M. S. (1975). On tests for detecting change in mean. *The Annals of statistics, pages 98-108.*
21. Vostrikova, L. Y. (1981). Detecting disorder in multidimensional random processes. *In Doklady Akademii Nauk, volume 259, pages 270-274. Russian Academy of Sciences.*
22. Worsley, K. (1979). On the likelihood ratio test for a shift in location of normal populations. *Journal of the American Statistical Association, 74(366a):365-367.*
23. Yao, Y.-C. (1988). Estimating the number of change-points via schwarz'criterion. *Statistics & Probability Letters, 6(3):181-189.*
24. Zhang, N. R. and Siegmund, D. O. (2007). A modi_ed bayes information criterion with applications to the analysis of comparative genomic hybridization data. *Biometrics, 63(1):22-32.*