

Assessing the Robustness of the Negative Binomial Multiple Change Point Algorithm Using Synthetic Data

Shalyne Nyambura, Anthony Waititu, Antony Wanjoya, Herbert Imboga

School of Mathematics & Physical Sciences, Jomo Kenyatta University of Agriculture & Technology, Kenya
Email: nshalyne@gmail.com, agwaititu@gmail.com, awanjoya@gmail.com, imbogaherbert@jkuat.ac.ke

How to cite this paper: Nyambura, S., Waititu, A., Wanjoya, A. and Imboga, H. (2024) Assessing the Robustness of the Negative Binomial Multiple Change Point Algorithm Using Synthetic Data. *Open Journal of Statistics*, 14, 775-789.
<https://doi.org/10.4236/ojs.2024.146036>

Received: November 26, 2024

Accepted: December 23, 2024

Published: December 26, 2024

Copyright © 2024 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 Negative Binomial Multiple Change Point Algorithm is a hybrid change detection and estimation approach that works well for overdispersed and equidispersed count data. This simulation study assesses the performance of the NBMCPA under varying sample sizes and locations of true change points. Various performance metrics are calculated based on the change point estimates and used to assess how well the model correctly identifies change points. Errors in estimation of change points are obtained as absolute deviations of known change points from the change points estimated under the algorithm. Algorithm robustness is evaluated through error analysis and visualization techniques including kernel density estimation and computation of metrics such as change point location accuracy, precision, sensitivity and false positive rate. The results show that the model consistently detects change points that are present and does not erroneously detect changes where there are none. Change point location accuracy and precision of the NBMCPA increases with sample size, with best results for medium and large samples. Further model accuracy and precision are highest for changes located in the middle of the dataset compared to changes located in the periphery.

Keywords

Kernel Density Estimation, Precision, Change point Location Accuracy, Sensitivity, Negative Binomial Multiple Change Point Algorithm

1. Introduction

The Negative Binomial Multiple Change Point Algorithm (NBMCPA) developed by [1] was found to work well for count datasets that exhibit over-dispersion where the variance exceeds the mean. The algorithm formulation also works well

for discrete count data that are equidispersed, as an alternative to the Poisson model for count data as discussed in [2]. The objective of this study is to conduct an evaluation of the NBMCPA robustness across different data scenarios which will mirror its applicability in real-life situations. As such, it is crucial to check how well the change point algorithm detects and estimates change points across varying data features prior to applying the model to real-life situations [3]. Highlighting the strengths of the NBMCPA alongside its limitations and computational complexities will provide users and future researchers in the evolving change point analysis field with the necessary information to select the most suitable algorithm available literature for their data requirements. Additionally, this work provides a platform for potential research into the application of the NBMCPA and the design of new algorithms for multiple change detection in over-dispersed count datasets.

In this study, we illustrate through Monte Carlo simulations how the NBMCPA behaves in different sample sizes and true change point locations. We check how the algorithm performance is affected by altering sample size and varying the location of known change points between the initial and last data point in a posteriori change setting.

Most change point models in available literature work well for medium-to-large datasets, with biased estimates produced for small-sized data sets [3]-[5]. The current study assesses the NBMCPA capability to correctly detect and locate a set of known (pre-defined) change points for various small and medium over-dispersed discrete data samples.

The sample sizes considered are $n = 12, 20, 60, 100, 200$ and 500 while the known change point locations are set at $n/4$, $n/2$, and $3n/4$. The locations of the change points are chosen so as to compare the performance of the algorithm when true changes are located closer to the first data point (at position $n/4$), midway through the data (at position $n/2$), and farther away from the first data point (at position $3n/4$). Past studies on change detection models have indicated that changepoints that are located near the middle of a dataset are detected with the most accuracy [6]-[8].

The algorithm is fitted over 1000 iterations to synthetic data generated under the Negative Binomial (r, p) distribution. Two data scenarios are considered: when there is no change in both distribution parameters, and when there exists a single simultaneous change in the dispersion parameter, r , and the success probability, p . In the latter scenario, parameters for the first segment are chosen as $r = 8$, $p = 0.4$ while the second segment takes the parameters $r = 5$, $p = 0.2$. The choice of the parameters matches those selected in simulation study presented in [1]. Many change point models in literature are capable of detecting changes in mean or dispersion parameters where the changes between adjacent data segments are sufficiently large, or somewhat obvious to the human eye when the time series is visualized. Further, changes are more easily detected when the dataset is large compared to small and medium datasets [9].

In this simulation study, we focus our investigations on the NBMCP algorithm's detection and estimation ability for small changes that are not immediately apparent on a time plot, yet not too small to be easily missed as change points by the algorithm. Visualization of changepoint estimates produced by the NBMCPA in the simulation study allows for visual evaluation of the bias (systematic deviation from the true change point) and precision (variability of the estimates) of the model under different conditions of sample size and change locations.

2. Methodology

Evaluation of performance of the NBMCPA for detecting and locating change points considers the view of changepoint detection similar to a binary classification problem, where each data point is expected to belong to either the change-point class or non-change point class as discussed in [5]. The step-wise recursive binary segmentation procedure is applied to simulated datasets, and the change detection test is conducted on each segment formed with the boundary at some value k , the possible change point. If a variation in the dispersion parameter is found to exist between the two data segments, then K falls into the detected change-points class, otherwise it belongs in the non-change point class. The detected change points may or may not coincide with the true changepoints that are fixed in the data simulations; hence, there is a need to check how well the algorithm estimates change points.

In this study, the NBMCPA is evaluated through the calculation of various metrics based on the concepts of a confusion matrix for the binary classification problem in the change-point detection setting. The metrics, including sensitivity or recall, accuracy, precision, and empirical coverage, are used in the analysis of change detection errors. In addition to numerical measures, the kernel density estimate (KDE) of the distribution of estimated change points is obtained and visualized. The density curve is used to analyze the distribution of estimated change-points in contrast to the true change points. The algorithm performance is checked by visually inspecting the features of the KDE curve at varying locations of true change points while controlling for the sample size and magnitude of the distinction between data segments.

2.1. NBMCPA Performance Metrics

2.1.1. True Positive Rate (Sensitivity)

In the context of change point detection, a True Positive (TP) refers to a correctly detected change point, while a False Negative (FN) is a missed changepoint or a true changepoint that was not detected by the model. Sensitivity, evaluates how well the model detects change points that are truly present. For each iteration, if the model identifies a change point at position k , where $k = n/4, n/2$ or $3n/4$ or within a defined tolerance, the change point estimate counts as a true positive. If the algorithm fails to detect the preset change point, then the event counts as a false negative. A high sensitivity indicates that the model is good at detecting

change points. Mathematically, the true positive rate is calculated according to [10] as:

$$\begin{aligned}\text{True Positive Rate} &= \frac{\text{Number of correctly detected change points}}{\text{Total number of true change points}} \\ &= \frac{\text{True Positives}}{\text{True Positives} + \text{False Negatives}}\end{aligned}\quad (1)$$

2.1.2. False Positive Rate (FPR)

This metric determines the rate of false alarms in change point detection. It measures how often the model incorrectly detects a change point where there is none. In the simulation study, a false alarm is found to occur if the model detects a change point outside of the expected region, so that the estimated change point is not near positions $k = n/4, n/2$ or $3n/4$ respectively. A low FPR is desirable, meaning the model does not falsely locate a change point too frequently.

$$\text{False Positive Rate} = \frac{\text{Number of false positives}}{\text{Total number of iterations}} \quad (2)$$

2.1.3. Change Point Location Accuracy

This metric evaluates how well the algorithm locates a true change point, particularly when it correctly detects a change point. For each iteration, we calculate the absolute difference between the estimated change point and the true change point, then find the location accuracy as the average of these differences over all iterations. Smaller values of mean absolute error (MAE) are more accurate in locating the change points, and if there is a consistently low error, then the model is consistently accurate. Mathematically, the MAE is calculated according to [10] and [11] as:

$$\text{Mean Absolute Error (MAE)} = \frac{1}{T} \sum_{i=1}^T |\hat{c}_i - c| \quad (3)$$

where c is the true/known change point, \hat{c}_i is the estimated changepoint for iteration i , and T is the total number of iterations where a change point is detected.

In addition to MAE, the median and interquartile range of the absolute errors are calculated. A narrow interquartile range and a median error close to 0 would indicate that the model is precise and accurate. A wider range or a high median error would indicate issues with precision or bias. Unlike the MAE, the Median Absolute Error is less sensitive to outliers; hence it gives the typical magnitude of error in a model's predictions, providing a better measure of change location accuracy in the presence of outliers.

2.1.4. Precision

In the context of change point analysis, precision measures how often the detected changepoints are correct or rather, how often an estimated change point matches the true changepoint. Precision is computed as the ratio of true positives to the

total number of all detected change points (among true positives and false positives). High precision means that when the model detects a change point, it is usually correct. Mathematically, the precision of the algorithm is computed according to [10] as:

$$\text{Precision} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Positives}} \quad (4)$$

2.1.5. F1 Score

The F1 score is a combined measure of precision and true positive rate, providing a single metric to evaluate the model's performance. An F1 score closer to 1 indicates a good balance between precision and recall, suggesting that the model is both accurate and reliable in detecting change points. Mathematically, the F1 score is computed from the model precision and sensitivity according to [5] [12] as:

$$\text{F1 Score} = 2 \times \frac{\text{Precision} \times \text{Sensitivity}}{\text{Precision} + \text{Sensitivity}} \quad (5)$$

2.1.6. Empirical Coverage

This is the proportion of times the true change point falls within a certain region, or a window around the true change point, where the model estimates a change point. In the evaluation of the change point detection ability of the NBMCPA, or other change detection models using classification metrics such as precision and empirical coverage, it is common to define a margin of error or threshold around the known change location to allow for minor discrepancies ([4] [12]). The tolerance level defines the allowable error in the change detection so that estimated change points that are close, though not exact, to the real change point are not classified as missed changes. Given a tolerance range, say within $\pm k$ positions of $n/2$, we compute the proportion of iterations in which the model estimates a change point within this range. A higher empirical coverage indicates that the model correctly detects change points close to the true location most of the time. Empirical coverage is computed according to [5] as:

$$\text{Empirical coverage} = \frac{\text{Number of detected change points within tolerance range}}{\text{Number of iterations}} \quad (6)$$

Selection of the threshold k depends on the scale of the data and magnitude of changes. If data spans a wide range or if changes are subtle, a larger threshold may be set to account for small variations that should not be penalized as false detection. For smaller datasets or datasets where changes are more pronounced, a smaller threshold could be used to ensure that the model's detected change points are very close to the true changepoint. The threshold can also be chosen based on how it affects performance metrics such as the F1 score. In this study, the changes considered are subtle with the Negative Binomial parameters remaining constant throughout the simulation study. A threshold of $k = 1$ is used for small samples of size $n \leq 60$, for sample sizes between $n = 60$ and $n = 200$ a threshold of $k = 3$ is used, while a tolerance of $k = 5$ is applied for sample sizes $200 < n < 500$.

2.2. Visual Analysis of Estimation Results

2.2.1. Histograms of Estimated Change Points

Frequency histograms are plotted to show the general distribution of the estimated change point locations across all iterations. The charts give an idea of the scatter or spread of the model's estimates around the true location. The tallest bar on the chart shows the modal change point estimate(s), while the general shape of the histogram helps to infer the distribution of estimated change points around the true/known value. Bars located farther from the true value point to the existence of false positives.

2.2.2. Kernel Density Estimate of Change Points

A Kernel Density Estimate (KDE) is a non-parametric way to estimate the probability density function of a random variable. For the change point estimation problem, when comparing the estimated change points against the true change point, the KDE helps to visualize how the estimated change points are distributed relative to the true change point [13]. In this study, we graph the Kernel density estimate of the detected change points across all iterations to provide a smoothed estimate of the data distribution. Mathematically, the KDE is computed as:

$$\hat{f}_h(x) = \frac{1}{nh} \sum_{i=1}^n K\left(\frac{x - X_i}{h}\right) \quad (7)$$

where:

$\hat{f}_h(x)$ is the estimated probability density function at point x .

n is the number of estimated changepoints.

X_i is the estimated changepoint from iteration i .

h is the bandwidth parameter, controlling for the smoothness of the density estimate.

$K(\cdot)$ is the kernel function.

In this study, we used the Gaussian (Normal) kernel defined by:

$$K(u) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}u^2} \quad (8)$$

Plugging the Gaussian kernel in Equation (8), the KDE in Equation (7) becomes:

$$\hat{f}_h(x) = \frac{1}{nh\sqrt{2\pi}} \sum_{i=1}^n e^{-\frac{1}{2}\left(\frac{x-X_i}{h}\right)^2} \quad (9)$$

This kernel ensures that each estimated change point contributes to the density estimate according to the normal distribution centered at X_i , with a spread controlled by the bandwidth h . If the model is performing well, the KDE will show a peak near the true change point. The width of the peak depends on the chosen bandwidth parameter h : a small bandwidth results in a more jagged estimate (with sharper peaks and valleys), while a large bandwidth gives a smoother, broader peak [14].

By visualizing the KDE, we assess whether the estimated change points are concentrated around the true change point or if there is significant bias or variability

in the model's estimates.

The bandwidth h is a critical parameter in KDE. Too small a bandwidth leads to an over-fitting problem, where the density estimate is too noisy, while too large a bandwidth results in under-fitting, where the estimate is too smooth. This study utilizes the Silverman's Rule of Thumb to identify the optimal bandwidth parameter h as:

$$h = 0.9 \cdot \min\left(\text{std}(X), \frac{\text{IQR}(X)}{1.34}\right) \cdot n^{-1/5} \quad (10)$$

where $\text{std}(X)$ is the standard deviation, $\text{IQR}(X)$ is the interquartile range, and n is the number of change points estimated.

3. Results and Discussion

3.1. Particulars of the Simulation Study

The Negative Binomial Change Point Algorithm was tested using synthetic data from a Negative Binomial distribution (r, p) . Monte Carlo simulations were performed using the R software to showcase the model performance in detecting changepoints in a case of single change in both distribution parameters. The study considered small and medium samples of sizes $n=12$, $n=20$, $n=60$, $n=100$, $n=200$ and $n=500$. The change points were set at each of the locations: $n/4$, $n/2$, and $3n/4$ where parameters r and p were assumed to change simultaneously, while the distributional form remained the same throughout the samples similar to [1].

The parameters for the two segments were chosen to ensure a clear but subtle distinction between data segments as follows:

Segment 1: $r=8$ and $p=0.4$

Segment 2: $r=5$ and $p=0.2$

The choice of model parameters was such that the change is not visually obvious on a time series plot, while still ensuring that the change is not too small so that it is easily missed by the algorithm. This section gives a summary of results for the simulation study and visual analysis showcasing performance metrics of the NBMCPA given different sample sizes, as well as the location of change points. For each sample size considered, the true changepoints are set at positions $n/4$, $n/2$ and $3n/4$. The distribution of estimated change points is visualized using histograms and plots of the kernel density estimates. The vertical red dashed lines mark the position of the true change points on the kernel density curves.

3.2. When the Sample Size Is $n=20$

Figures 1-3 show the distribution of the estimated change points for a small sample of size $n=20$. From the histogram, the estimated change points are scattered at different positions with a slight concentration around the true change point for cases where the true changes are located at position $n/4$ and $3n/4$. The data distributions displayed on the histograms in **Figure 1** and **Figure 3** reveal that the algorithm has lower accuracy and precision for small sample sizes when the true

changes are located near the endpoints of the dataset. The KDE curves show two relatively broad peaks, with the higher peaks showing a higher concentration of the estimates around the true change points. The lower peaks are an indication of false alarms in change detection. On the other hand, when the true change point is located in the middle of the dataset (at $n/2$), the histogram shows a peak at around the middle position, with fewer estimates in regions away from $n/2$. In addition, the KDE curve has a single peak near, but not coinciding with the true change location. This result indicates that the model accuracy and precision is higher for changes located at $n/2$ compared to when changes are located at $n/4$ and $3n/4$. However, there is a notable misclassification rate, as evidenced by the departure of the peak of the KDE curve from the true change point.

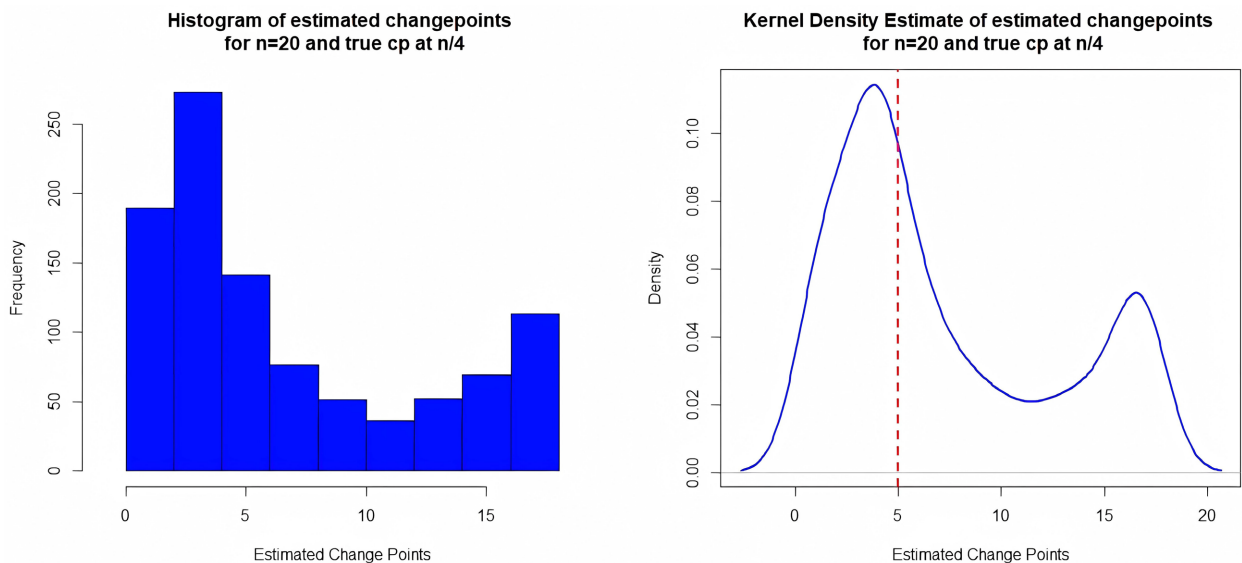


Figure 1. Distribution of change points for a case of a single change at $n/4$ for a sample size $n = 20$.

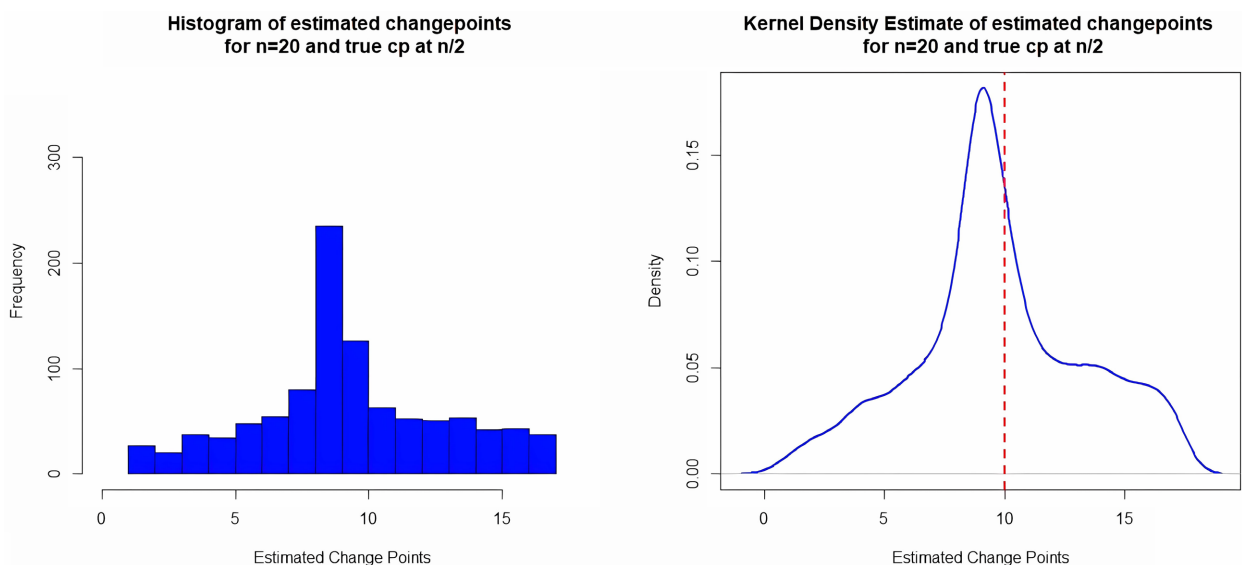


Figure 2. Distribution of change points for a case of a single change at $n/2$ for a sample size $n = 20$.

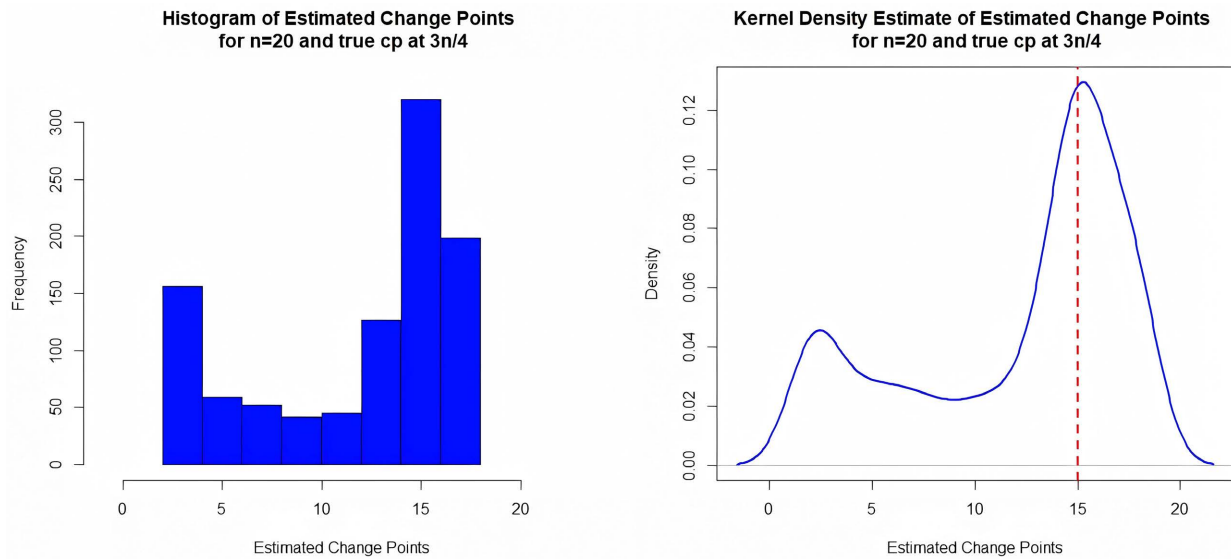


Figure 3. Distribution of change points for a case of a single change at $3n/4$ for a sample size $n = 20$.

3.3. When the Sample Size Is $n = 200$

Figures 4-6 show the distribution of the estimated change points for a sample of size $n = 200$. The histogram shows that estimated change points are concentrated around the true change positions for all three locations, indicating a high precision and accuracy of the model across all positions of actual change points. This finding is further supported by the KDE curve which has a narrow peak in all three cases indicating that there is no significant bias or variability in the model's estimates for the given sample size. However, the distribution of estimated change points has longer tails to the right (for true change and $n/4$) and to the left (for true change and $3n/4$) pointing to the presence of false alarms and lower precision in the change point detection before and after the true change points respectively.

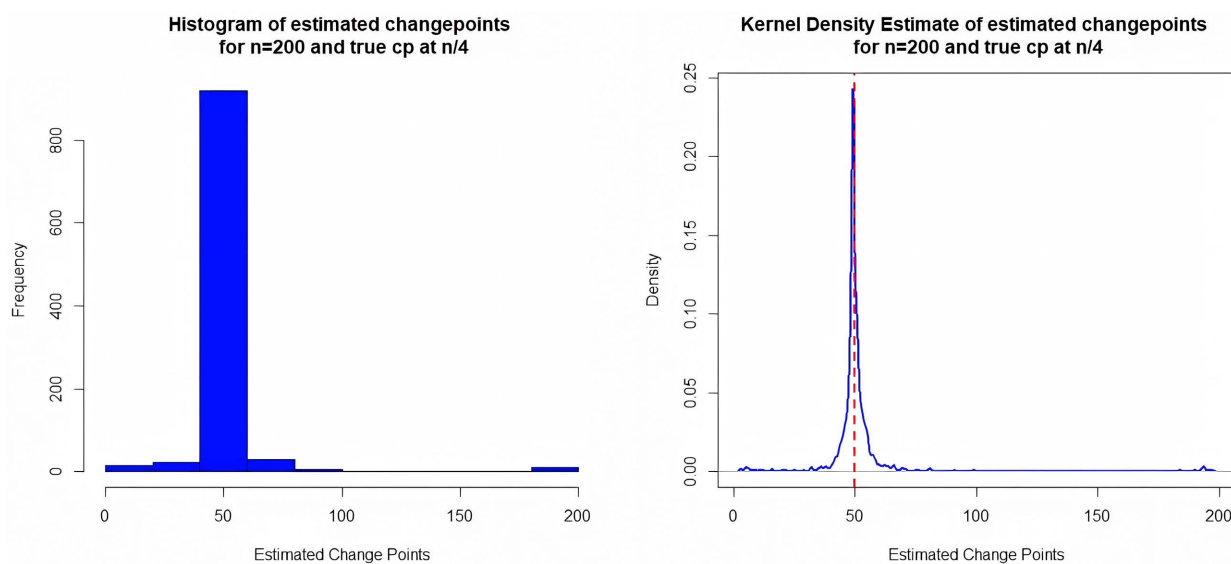


Figure 4. Distribution of change points for a case of a single change at $n/4$ for a sample size $n = 200$.

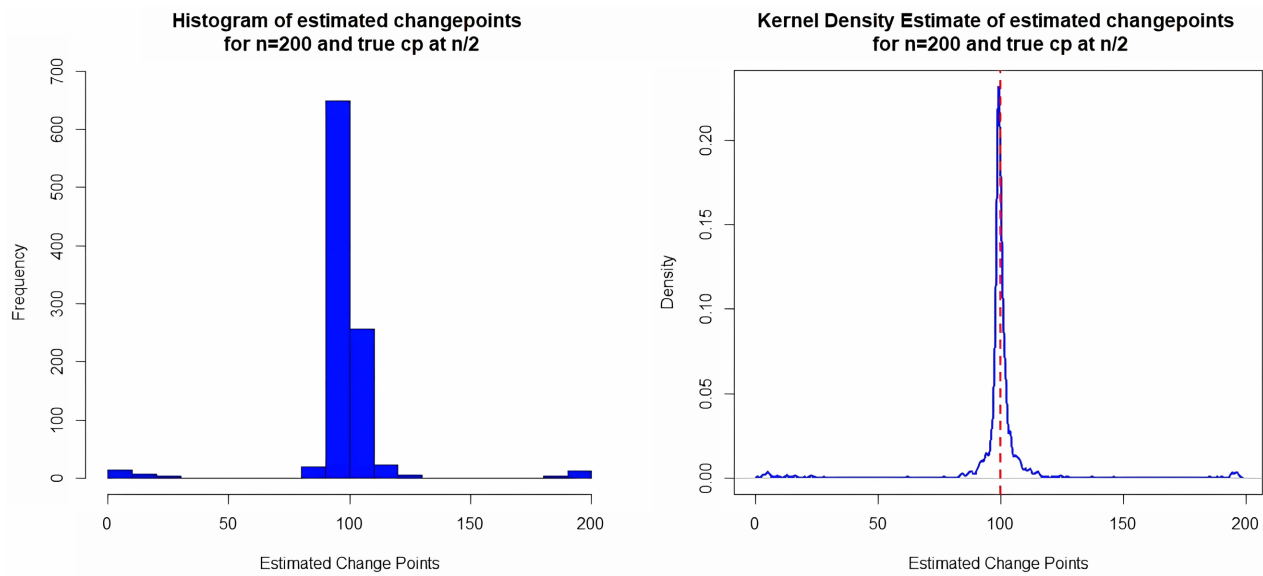


Figure 5. Distribution of change points for a case of a single change at $n/2$ for a sample size $n = 200$.

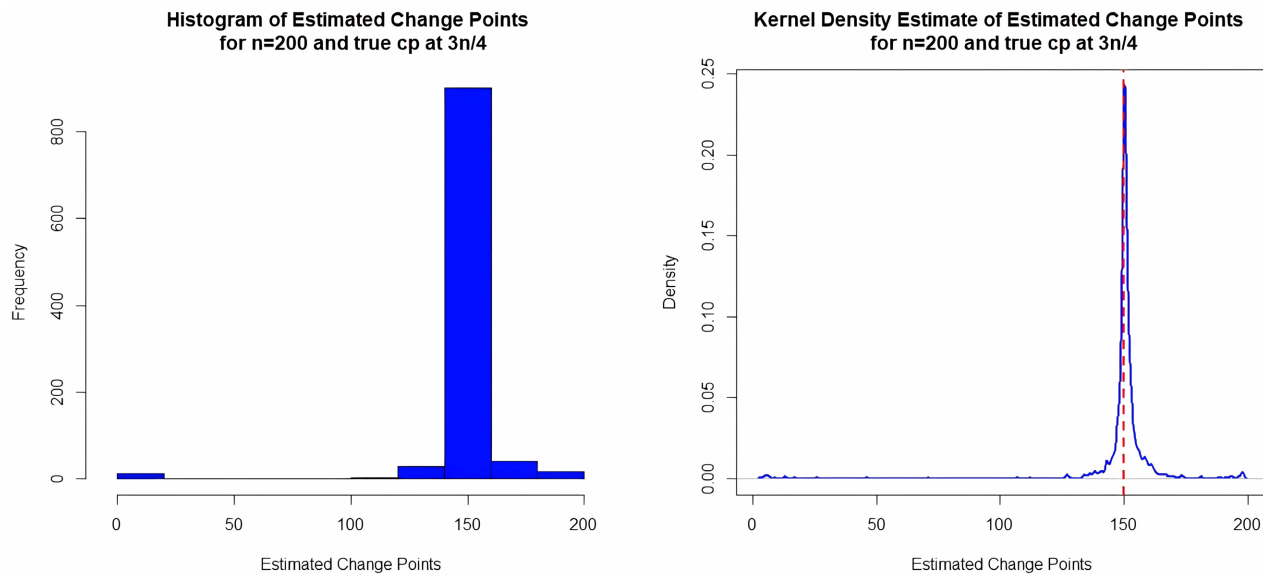


Figure 6. Distribution of change points for a case of a single change at $3n/4$ for a sample size $n = 200$.

3.4. When the Sample Size Is $n = 500$

Figures 7-9 show the distribution of the estimated change points for a sample of size $n = 500$. From the histogram, the estimated change points are concentrated around the true change positions for all three locations indicating a high precision and accuracy of the model. This finding is further supported by the KDE curve which has a narrow peak in all three cases indicating that there is no significant bias or variability in the model’s estimates for the given sample size.

Table 1 gives a summary of performance metrics based on estimated change points for different sample sizes ($n = 12, 20, 60, 100, 200, 500$) and true change point locations ($n/4, n/2, 3n/4$). The change point estimation error was calculated

as the absolute difference between the estimated change point and the true change point. The results showed that the NBMCPA locates changes with a median absolute error close to 0, with a narrow (between 2 and 7) interquartile range irrespective of the sample size and true change location. Further, the mean absolute error is fairly small (ranging between 1.7 and 14.9) across all sample sizes. Therefore, the algorithm is found to be consistently accurate.

With regard to model precision, the model's capability to detect changes and correctly estimate change points increases with the sample size. In addition, true changes located in the middle of the dataset (at position $n/2$) were estimated with higher precision in comparison to changes located at positions $n/4$ and $3n/4$. The model performance in terms of the empirical coverage was similar to the precision.

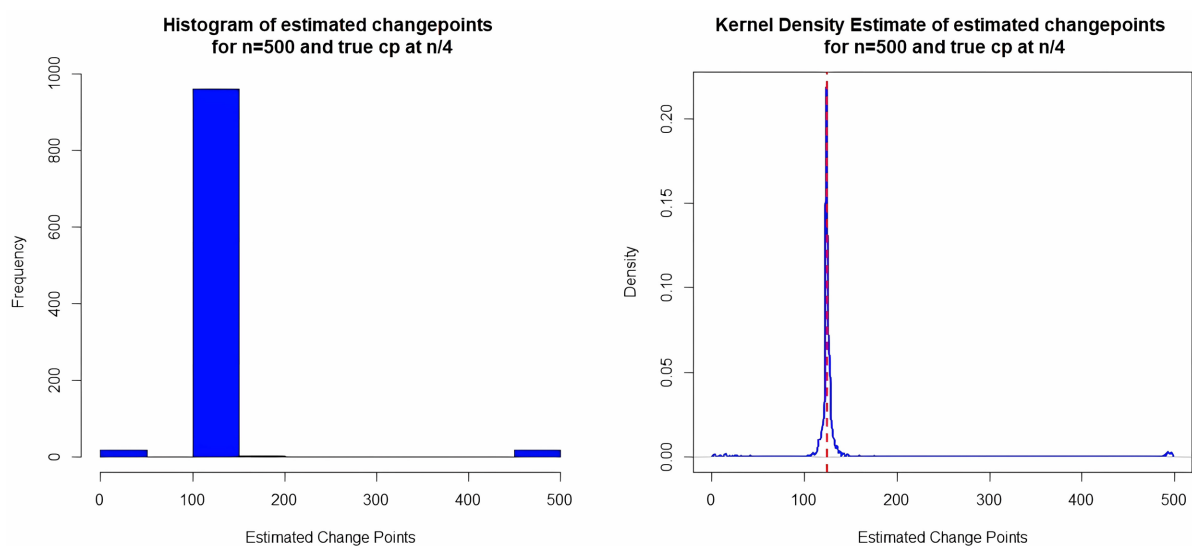


Figure 7. Distribution of change points for a case of a single change at $n/4$ for a sample size $n = 500$.

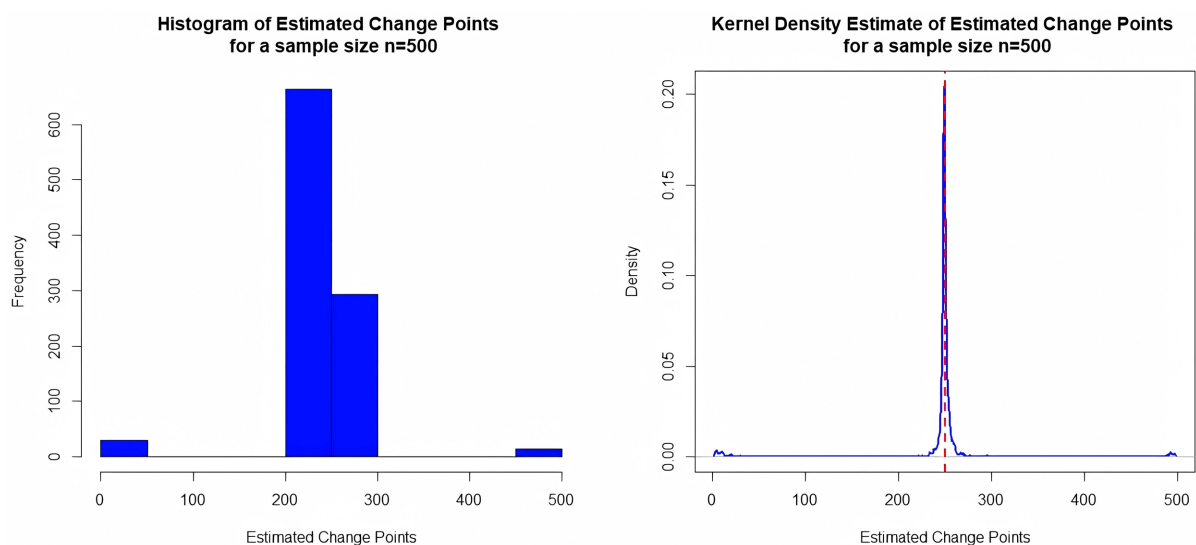


Figure 8. Distribution of change points for a case of a single change at $n/2$ for a sample size $n = 500$.

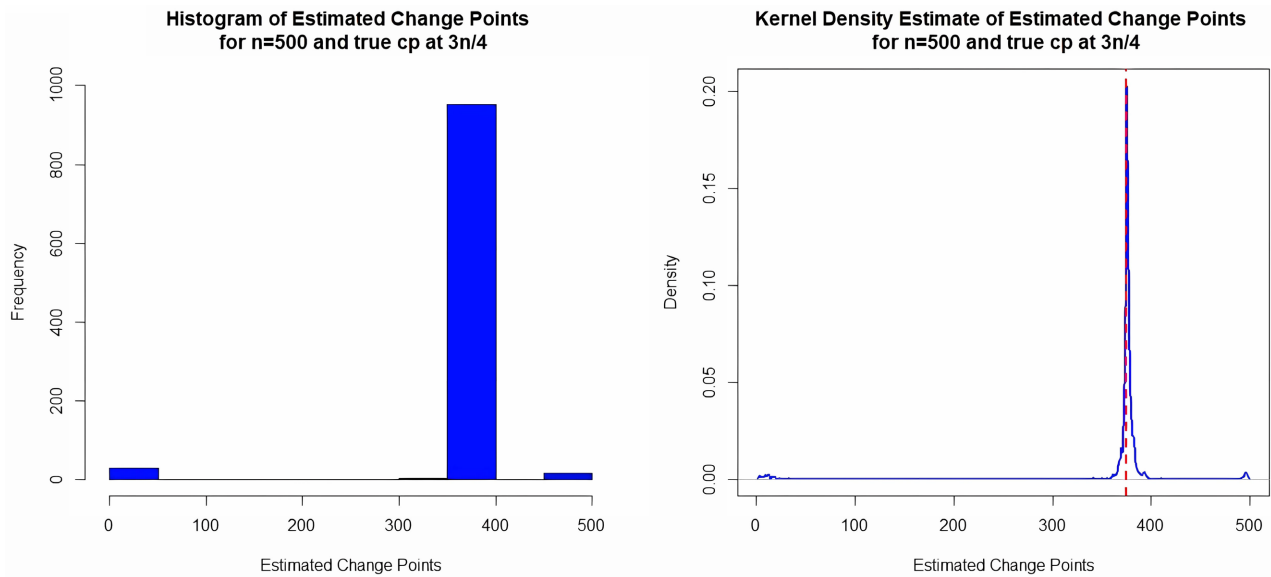


Figure 9. Distribution of change points for a case of a single change at $3n/4$ for a sample size $n = 500$.

Table 1. Model performance metrics for varying sample sizes and change locations.

Sample size	True cpt	IQR AE	MedAE	MAE	Precision	Sensitivity	FPR	F1 Score
$n = 12$	$n/4$	3	2	2.53	0.385	0.096	0.614	0.154
	$n/2$	2	1	1.77	0.503	0.137	0.497	0.162
	$3n/4$	5	0	2.27	0.521	0.104	0.478	0.113
$n = 20$	$n/4$	6	3	4.29	0.328	1	0.672	0.494
	$n/2$	3	2	2.76	0.424	1	0.576	0.596
	$3n/4$	6	2	4.08	0.403	1	0.597	0.574
$n = 60$	$n/4$	7	2	7.17	0.407	1	0.593	0.579
	$n/2$	4	2	4.79	0.455	1	0.545	0.625
	$3n/4$	6	2	6.27	0.434	1	0.566	0.605
$n = 100$	$n/4$	4	2	6.14	0.673	1	0.327	0.805
	$n/2$	4	2	5.38	0.695	1	0.327	0.820
	$3n/4$	5	2	5.52	0.669	1	0.331	0.802
$n = 200$	$n/4$	3	1	4.94	0.730	1	0.270	0.844
	$n/2$	3	1	6.43	0.737	1	0.263	0.849
	$3n/4$	3	1	6.35	0.719	1	0.281	0.837
$n = 500$	$n/4$	3	1	11.82	0.741	1	0.259	0.851
	$n/2$	3	1	13.21	0.820	1	0.172	0.906
	$3n/4$	4	1	14.96	0.806	1	0.194	0.893

The algorithm detects changes whenever true variations in model parameters exist, irrespective of whether they are correctly located or not. The true positive rate equals 1 across all sample sizes and true change positions, with the exception

of very small sample sizes, $n = 12$ where the sensitivity ranges between 0.096 and 0.137. Based on the simulation results, the algorithm misses true change points in some iterations, which is attributed to a low signal-to-noise ratio given the subtle change occurring in a few observations.

The false positive rate decreases with increasing sample size (FPR = 0.497 for $n = 12$ and FPR = 0.172 for $n = 500$ when the true change point is located at $n/2$). However, similar to model's accuracy, the FPR is lower for changes located in the middle of the dataset compared to quarter-way and three-quarter-way.

The F1 score provides a balance, in terms of harmonic mean, between the sensitivity and precision of the model. The results showed that F1 score increases with the sample size, and similar to precision, it is highest for changes located midway through the data compared to the true change locations of $n/4$ and $3n/4$. Located further away from the initial data point were more accurately estimated for small sample sizes.

4. Conclusion and Recommendations

This simulation study considered a single change detection problem for datasets with sizes ranging between $n = 12$ and $n = 500$ while controlling for the magnitude of change. Given the foregoing results, the NBMCPA consistently detects and estimates change points across varying sample sizes and true change point locations. In particular, model's ability to accurately locate the true change point varies depending on the location of the true change point, relative to the size of the dataset, similar to the results discussed in [4]. In this study, the model performance was compared across three different true change point locations: at $n/2$, $n/4$, and $3n/4$. The simulation results revealed that the algorithm has the highest precision when the true changepoint is located at the middle (point $n/2$) of the dataset consistent with [6] [8] [10]. A symmetrical distribution with equal amount of data on either side of the true change point makes it easier for the model to detect the change point because the magnitude of change in the data before and after the true change point is likely to be roughly similar, allowing the model to detect subtle abrupt shifts.

When the true changepoint is located at position $n/4$ or $3n/4$, there is a data imbalance such that the true change point is located closer to the beginning (at $n/4$) or end (at $3n/4$) of the dataset. The smaller data segments before or after the change may have less data to provide evidence of the change, contingent on the sample size, which could result in lower detection accuracy.

In a nutshell, the model performance is optimal when the true changepoint is located at $n/2$ because the data is balanced, and the shift is clear on both sides of the change point, which results in a high precision and detection accuracy, both of which increase with sample size. Further, the rate of false alarms is highest for changes located quarter-way, followed by changes at $3n/4$, and lowest for changes located midway through the dataset. This simulation study considered model performance under a single-change setting. However, as the NBMCPA is

designed to detect a single change at a time, starting with the most pronounced, the results of this study may be generalized for the multiple changepoint setting. Further investigations can be done to investigate the algorithm performance where there are multiple subtle changes in close proximity to each other within small and medium datasets.

Conflicts of Interest

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

References

- [1] Nyambura, S., Waititu, A., Wanjoya, A. and Imboga, H. (2024) A Likelihood-Based Multiple Change Point Algorithm for Count Data with Allowance for Over-Dispersion. *Open Journal of Statistics*, **14**, 518-545. <https://doi.org/10.4236/ojs.2024.145023>
- [2] Wilken, B.A. (2007) Change-Point Methods for Overdispersed Count Data. Master's Thesis, Air Force Institute of Technology. <https://scholar.afit.edu/etd/3092/>
- [3] Gupta, M., Wadhvani, R. and Rasool, A. (2024) Comprehensive Analysis of Change-Point Dynamics Detection in Time Series Data: A Review. *Expert Systems with Applications*, **248**, Article ID: 123342. <https://doi.org/10.1016/j.eswa.2024.123342>
- [4] Truong, C., Oudre, L. and Vayatis, N. (2020) Selective Review of Offline Change Point Detection Methods. *Signal Processing*, **167**, Article ID: 107299. <https://doi.org/10.1016/j.sigpro.2019.107299>
- [5] Van den Burg, G.J.J. and Williams, C.K.I. (2020) An Evaluation of Change Point Detection Algorithms. arXiv: 2003.06222.
- [6] Mundia, S., Gichuhi, A. and Kihoro, J. (2014) The Power of Likelihood Ratio Test for a Change-Point in Binomial Distribution. *Journal of Agriculture, Science and Technology*, **16**, 105-123.
- [7] Dorcas Wambui, G. (2015) The Power of the Pruned Exact Linear Time(pelt) Test in Multiple Changepoint Detection. *American Journal of Theoretical and Applied Statistics*, **4**, 581-586. <https://doi.org/10.11648/j.ajtas.20150406.30>
- [8] Nyambura, S. (2016) Estimation of Change Point in Poisson Random Variables Using the Maximum Likelihood Method. *American Journal of Theoretical and Applied Statistics*, **5**, 219-224. <https://doi.org/10.11648/j.ajtas.20160504.18>
- [9] Scott, M.J. and Gorman, T.H. (2015) Optimal Detection of Multiple Change Points. *SpringerLink*, **19**, 1-19.
- [10] Aminikhanghahi, S. and Cook, D.J. (2016) A Survey of Methods for Time Series Change Point Detection. *Knowledge and Information Systems*, **51**, 339-367. <https://doi.org/10.1007/s10115-016-0987-z>
- [11] Chander, S., Liang, R. and Chen, J.P. (2024) Change Point Detection for Automatic Time Series Forecasting. *Journal of Student Research*, **13**, 1-15. <https://doi.org/10.47611/jsrhs.v13i2.6597>
- [12] Killick, R., Fearnhead, P. and Eckley, I.A. (2012) Optimal Detection of Changepoints with a Linear Computational Cost. *Journal of the American Statistical Association*, **107**, 1590-1598. <https://doi.org/10.1080/01621459.2012.737745>
- [13] Chen, Y. (2017) A Tutorial on Kernel Density Estimation and Recent Advances. *Biostatistics & Epidemiology*, **1**, 161-187. <https://doi.org/10.1080/24709360.2017.1396742>

- [14] Kang, Y. and Noh, Y. (2019) Comparison Study of Kernel Density Estimation According to Various Bandwidth Selectors. *Journal of the Computational Structural Engineering Institute of Korea*, **32**, 173-181.
<https://doi.org/10.7734/coseik.2019.32.3.173>