Statistical Inference for Multiple Change-Point Models

Wu Wang[a], Xuming He[b], Zhongyi Zhu[c]

[a]*Statistics Program, King Abdullah University of Science and Technology*

[b]*Department of Statistics, University of Michigan*

[c]Department of Statistics, Fudan University

Abstract

In this paper, we propose a new technique for constructing confidence intervals for the mean of a noisy sequence with multiple change-points. We use the weighted bootstrap to generalize the bootstrap aggregating or bagging estimator. A standard deviation formula for the bagging estimator is introduced, based on which smoothed confidence intervals are constructed. To further improve the performance of the smoothed interval for weak signals, we suggest a strategy of adaptively choosing between the percentile intervals and the smoothed intervals. A new intensity plot is proposed to visualize the pattern of the change-points. We also propose a new change-point estimator based on the intensity plot, which has superior performance in comparison with the state-of-the-art segmentation methods. The finite sample performance of the confidence intervals and the change-point estimator are evaluated through Monte Carlo studies and illustrated with a real data example.

This is the author manuscript accepted for publication and has undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the Version of Record. Please cite this article as doi: 10.1111/sjos.12456

Keywords: bagging estimator, binary segmentation, bootstrap, copy number variation, multiple change-points.

1 Introduction

Changes are frequently occurring in the real world. External influences in the environment are likely to induce instability in the underlying stochastic systems. A common form of the change-point model is that the parameters in the stochastic system are piece-wise constant. We consider a model with multiple change-points,

$$X_t = \mu_t^0 + \epsilon_t,\ t = 1, \ldots, n, \tag{1}$$

where the mean $\{\mu_t^0\}_{t=1}^n$ is a piece-wise constant signal with change-points at $t_1^0, \ldots, t_N^0$, that is $\mu_t^0 = \beta_{i+1}^0$ for $t_i^0 + 1 \leq t \leq t_{i+1}^0,\ i = 0, \ldots, N$, where $t_0^0 = 0$, $t_{N+1}^0 = n$ and $\beta_i^0 \neq \beta_{i+1}^0, i = 1, \ldots, N$. The superscript ^{0} is used to indicate the true values of parameters. The errors $\{\epsilon_t\}_{t=1}^n$ are independent and identically distributed (i.i.d.) zero-mean random variables.

Change-points detection is closely related to the model selection problem. Some researchers proposed to obtain change-point estimates by penalizing the number of the change-points in the model through a Schwarz-like criterion (Yao, 1988; Killick et al., 2012), or by penalizing the total variation of the piecewise constant signal using the Fused Lasso (Harchaoui & Lévy-Leduc, 2010). Fast computation is critical in change-point detection. The computational complexity of the binary segmentation (BS) algorithm introduced by Vostrikova (1981) is $O(n \log(n))$. Refinements of BS, such as circular BS (Olshen et al., 2004), and wild binary segmentation (Fryzlewicz, 2014) have been successfully applied to DNA copy number studies. The screening and ranking algorithm (SaRa) proposed by Niu & Zhang (2012) explored the idea of local statistics for detecting change-points. The computational complexity of SaRa is $O(n)$.

In this paper, we are interested in constructing confidence intervals for $\mu_t^0, t = 1, \ldots, n$ of model (1). Siegmund (1988) discussed confidence sets estimation for parameters of the exponential family, but only one change-point is allowed. Frick et al. (2014) and Pein, Sieling & Munk (2017) proposed the simultaneous multiscale change-point estimator (SMUCE) and constructed confidence sets for the change-points and model parameters. Frick et al. (2014) relied on the consistency of the estimated change-points to ensure good performance of the confidence sets. They did not tackle the problem that the number of change-points might be misspecified in finite samples. When the jump size and the spacing between change-points are small, the probability of correctly estimating the number of change-points might be significantly lower than 1. Thus their confidence sets might have low coverage for weak signals in finite samples. Consider the simulation in Figure 1, where the true change-points are at $40, 80, 120$, the empirical coverage probabilities of the SMUCE intervals represented by line 5 are lower than the nominal level 95%. See Section 4 for the setup of the simulation.

[Figure 1 about here.]

We emphasize that the randomness in change-point estimation must be appropriately dealt with when constructing confidence intervals for the mean. Confidence sets for the mean implicitly depend on the underlying change-point estimator, ignoring randomness in the change-point estimates will have harmful effects in confidence sets estimation as illustrated by Figure 1.

To incorporate the randomness in change-point estimation, we use the weighted bootstrap to generalize the bootstrap aggregating or bagging estimator of Efron (2014). We develop a standard deviation formula for the bagging estimator, based on which confidence intervals are constructed, that we define to be the *smoothed* interval. The bagging estimator smooths the discontinuities in the process of change-point detection by averaging over

the bootstrap replications. The bagging estimator and the corresponding standard error formula can successfully capture the randomness in estimating the number and location of the change-points.

The bootstrap is a popular resampling method for obtaining the distribution of estimators and test statistics, see Efron (1982) for an introduction. In time series analysis, the block bootstrap (Carlstein et al., 1986) can be used to obtain bootstrap samples from a stationary sequence without specifying the dependence structure. Suppose the observed data is $(X_1, \ldots, X_n)$, and denote a block of data as $\mathbf{X}_m^j = \{X_{j+1}, \ldots, X_{j+m}\}$, where m is the length of the block. For simplicity, assume that $n = mb$ where both m and b are integers. Bootstrap samples are obtained by sampling b blocks with replacement from $\{\mathbf{X}_m^{jm}, 0 \leq j \leq b-1\}$ and $\{\mathbf{X}_m^j, 0 \leq j \leq n-m\}$ for non-overlapping (Carlstein et al., 1986) and overlapping (Kunsch, 1989) block bootstrap, respectively. If the time series follows a suitable model, e.g., $X_t = g(\mathbf{Z}_t) + \epsilon_t$, where $\mathbf{Z}_t$ is a covariate vector, the residual bootstrap can be used to obtain bootstrap samples (Freedman, 1984) by sampling from the centered residuals. See Horowitz (2019) for a recent overview.

In the change-point literature, bootstrap methods are used to approximate the critical values of test statistics for the existence of a single change-point (Kirch, 2007). In online change-point monitoring problems, Hlávka et al. (2016) considered the problem of monitoring changes with bootstrapped critical values in time series model and regression model. Besides hypotheses testing, Hušková & Kirch (2010) obtained confidence interval for a single change-point in time-series models. We contribute by considering the problem of doing statistical inference for the mean of a multiple change-point model based on the bootstrapping methods, which is largely ignored in the change-point literature.

In the simulation studies, we found that the smoothed interval has coverage probabilities close to the nominal level when the signals are moderate or strong, but may have

coverage probabilities lower than the nominal level otherwise. As indicated by Fryzlewicz (2014), the square root of the minimal spacing between change-points multiplied by the size of jumps determines how easily change-points can be detected. When these two are small or the variance of the noise is high, the probability of underestimating the number of change-points is high. To correct the coverage probability of the smoothed interval, we introduce *the intensity score, the bootstrap change-point (BootCp) estimator, and the adaptive interval*. The intensity scores are defined as the frequencies of change-points occurring in the bootstrap replications. The plot of the intensity scores versus the locations is defined as an intensity plot. Intuitively, the intensity scores should be large when a location is close to a true change-point, and close to 0 otherwise. The lower panel of Figure 1 illustrates the intensity plot with simulated data. The intensity plot can also be used to discover possible missing change-points. If some regions of the intensity plot are 'significant' to human eyes, but no change-points are detected by an algorithm, the researcher can further investigate those regions either by data analysis or expert's knowledge.

Based on the intensity score, we define a new change-point estimator, the BootCp estimator. The BootCp estimator is essentially a subset of the local maxima of the intensity scores. The unique feature of the BootCp estimator is that every change-point estimate is coupled with an intensity score, which indicates the plausibility of a change-point occurring in the sample. The BootCp estimator performs well compared to state-of-art segmentation methods in the simulation studies.

With the BootCp estimator and the intensity scores, we define the adaptive interval as a data-driven choice between the percentile interval and the smoothed interval. The percentile interval is constructed using the empirical quantiles of the bootstrap samples of the mean estimator. The percentile interval is more robust to misspecification of the change-points than the smoothed interval because it does not rely on a point estimate or a

standard error estimate. The disadvantage of the percentile interval is that it is much wider than the smoothed interval. For an estimated change-point with a low intensity score, the bagging estimator of the mean for nearby locations will be severely biased. Thus the resulting smoothed interval can not cover the true mean with the nominal probability. To protect the confidence interval from low coverage probabilities, we use the percentile interval to replace the smoothed interval. This is the intuition behind the adaptive intervals. We show by simulation examples that the coverage probability of the adaptive interval is closer to the nominal level compared to the smoothed interval, and the average length of the adaptive interval is shorter compared to the percentile interval.

Although the model we consider is simple, the method can be generalized to more complicated models. The methodology essentially consists of three elements, a valid bootstrapping method for the observed data, a multiple change-point detection procedure which can be applied to bootstrapped data, and the standard deviation of the bagging estimator. With the three elements, one can proceed exactly as we do in our paper to derive confidence intervals for interested parameters in more general settings.

Bayesian analysis is another approach for quantifying uncertainty in change-point models. Application of Bayesian analysis in change-point model can be date back to Shiryaev (1963) and Chernoff & Zacks (1964), also see Kim et al. (2009) for subsequent developments. For multiple change-points models, prior distributions on the number and location of change-points penalize the complexity of the studied model, see Du et al. (2016). Hidden Markov Model (HMM) is a popular formulation in Bayesian change-point analysis, in which a latent state variable is used to indicate the segment from which a particular observation has been drawn, see Chib (1998) and Rozenholc & Nuel (2013). The Bayesian method is a promising candidate for the inference problem, although it needs further research.

We organize the rest of the paper as follows. In Section 2, we describe the model and the proposed confidence intervals. Large sample properties are investigated in Section 3. Simulation studies are conducted in Section 4 to illustrate the finite sample performance of the proposed procedure. In Section 5, we apply our proposed method to DNA copy number data. Section 6 summarizes the paper.

2 Model and Method

2.1 The sequential binary segmentation

In this paper, we implement the BS algorithm sequentially as advocated by Bai (1997), which is more convenient for theoretical analysis. We first describe the BS algorithm for estimating $m \geq 1$ change-points. Afterwords, we will discuss how to select the number of change-points. For two integers $0 \leq k_1 < k_2 \leq n$, let $\bar{X}_{k_1,k_2} = (k_2 - k_1)^{-1} \sum_{t=k_1+1}^{k_2} X_t$ and $S_n(k_1, k_2) = \sum_{t=k_1+1}^{k_2} (X_t - \bar{X}_{k_1,k_2})^2$. Consider one change-point at $k_1 < k < k_2$, define the residual sum of squares for the data segment $(X_{k_1+1}, \ldots, X_{k_2})$ as $L_n(k; k_1, k_2) = S_n(k_1, k) + S_n(k, k_2)$. In each iteration of the sequential BS, only one change-point is added to the set of the estimated change-points such that the total residual sum of squares is minimized. For instance, at the beginning of the ith round, $i < m$, we already have $i - 1$ estimated change-points, say $0 = \hat{t}_0 < \hat{t}_1 < \cdots < \hat{t}_{i-1} < \hat{t}_i = n$. The original sample $\{X_t\}_{t=1}^n$ is partitioned into i segments, i.e., $\{X_t\}_{t=1}^{\hat{t}_1}, \ldots, \{X_t\}_{t=\hat{t}_{i-1}+1}^{n}$, by the estimated change-points. Then, the ith detected change-point is $\hat{t} = \hat{k}_{\hat{j}}$, where $\hat{j} = \arg\min_{j=1,\ldots,i}\{S_n(\hat{t}_{j-1}, \hat{t}_j) - L_n(\hat{k}_j; \hat{t}_{j-1}, \hat{t}_j)\}$, and $\hat{k}_j = \arg\min_{k=\hat{t}_{j-1}+1,\ldots,\hat{t}_j-1} L(k; \hat{t}_{j-1}, \hat{t}_j)$.

To select the number of change-points, we use a BIC criterion as in Yao (1988) and Fryzlewicz (2014). For a candidate model with m estimated change-points $0 = \hat{t}_0 < \hat{t}_1 < \cdots < \hat{t}_m < \hat{t}_{m+1} = n$, define an estimate of μ_t^0 as $\hat{\mu}_t(m) = \bar{X}_{\hat{t}_i,\hat{t}_{i+1}}$ for $t \in [\hat{t}_i + 1, \hat{t}_{i+1}]$,

let $\hat{\sigma}_m^2 = n^{-1}\sum_{t=1}^{n}(X_t - \hat{\mu}_t(m))^2$. The BIC criterion is defined as

$$\mathrm{BIC}(m) = \frac{n}{2}\log\hat{\sigma}_m^2 + m\log n.$$

The BIC criterion is based on an assumption of Gaussianity and common variance. Denote $\hat{N} = \arg\min_{m\leq\mathcal{N}}\mathrm{BIC}(m)$, where $\mathcal{N}$ is a known upper bound for the number of change-points N. In practice, we can proceed by setting $\mathcal{N}$ at a relatively large number, e.g., $\mathcal{N} = n/10$. Because of the sequential nature of the BS algorithm, the computational burden increases only linearly with $\mathcal{N}$.

We are interested in constructing confidence intervals for $\mu_t^0, t = 1, \ldots, n$. Without loss of generality, we will focus on constructing a 95% confidence interval for μ_1^0, the confidence interval for other μ_t^0 can be constructed similarly. Conditional on the estimated location of the change-points, we can estimate μ_1^0 by $\hat{\mu}_1 = (\hat{t}_1)^{-1}\sum_1^{\hat{t}_1} X_t$, and the corresponding standard error can be estimated by $(\hat{\sigma}_{\hat{\mu}_1})^2 = (\hat{t}_1^2)^{-1}\sum_1^{\hat{t}_1}(X_t - \hat{\mu}_1)^2$. We define the 95% unsmoothed confidence interval for μ_1^0 as

$$\left[\hat{\mu}_1 - 1.96\hat{\sigma}_{\hat{\mu}_1},\ \hat{\mu}_1 + 1.96\hat{\sigma}_{\hat{\mu}_1}\right]. \tag{2}$$

Note that the quantiles of the normal distribution are used in (2). As pointed out by a reviewer, the confidence interval can be constructed using the quantiles t-distribution as well, especially for a small $\hat{t}_1$. The unsmoothed interval ignores the randomness for estimating $\hat{t}_1$. If the change-point estimates miss t_1^0 frequently, $\hat{\mu}_1$ will be severely biased, and the unsmoothed interval is problematic, see Figure 1 line 6 for an illustration. The unsmoothed interval works well only when the number and the location of the change-points are estimated accurately.

2.2 The smoothed interval

Efron (2014) showed the power of bootstrap in post model selection inferences, he developed a closed-form formula for the standard deviation of the bagging estimator, and con-

structed confidence intervals for interested model parameters. In this section, we use the weighted bootstrap to generalize Efron (2014)'s method and construct confidence intervals for $\mu_t^0,\ t = 1, \ldots, n$ of model (1). The weighted bootstrap is widely applied in confidence interval estimation and hypothesis testing. Chatterjee et al. (2005) studied the weighted bootstrap for estimators obtained by solving estimating equations. Spokoiny et al. (2015) considered the weighted bootstrap for constructing confidence sets for parameters in a possibly mis-specified model by a quasi-likelihood method. See also Chernozhukov et al. (2013) for studies of the weighted bootstrap when the dimension of the parameters is high.

Let $\mathbf{w} = (w_1, \ldots, w_n)$ be a random vector with i.i.d. elements. The elements w_i are non-negative with mean 1 and variance 1. In the simulation study and data analysis, we use weights that are independently generated from the exponential distribution with mean 1. We apply the sequential BS algorithm described in Section 2.1 to the weighted sample, with the residual sum of squares replaced by the weighted sum of squares,

$$L_n^w(k; k_1, k_2) = S_n^w(k_1, k) + S_n^w(k, k_2),$$

where $S_n^w(k_1, k_2) = \sum_{t=k_1+1}^{k_2} w_t(X_t - \bar{X}_{k_1,k_2}^w)^2$, and $\bar{X}_{k_1,k_2}^w = (\sum_{t=k_1+1}^{k_2} w_t)^{-1} \sum_{t=k_1+1}^{k_2} w_t X_t$ for integers $0 \le k_1 \le k < k_2 \le n$. The number of change-points is also determined by the BIC criterion. For a candidate model with m estimated change-points denoted as $0 = \hat{t}_0^w < \hat{t}_1^w < \cdots < \hat{t}_m^w < \hat{t}_{m+1}^w = n$, denote $(\hat{\sigma}_m^w)^2 = (\sum_{t=1}^n w_t)^{-1} \sum_{t=1}^n w_t(X_t - \hat{\mu}_t^w(m))^2$ where $\hat{\mu}_t^w(m) = \bar{X}_{\hat{t}_i^w, \hat{t}_{i+1}^w}^w$. The BIC criterion is defined as

$$\mathrm{BIC}(m) = \frac{n}{2} \log(\hat{\sigma}_m^w)^2 + m \log n.$$

We obtain the number of change-points by $\hat{N}^w = \arg\min_{m \le \mathcal{N}} \mathrm{BIC}(m)$. The ordered change-point estimator for the weighted sample is denoted as $0 = \hat{t}_0^w < \hat{t}_1^w < \cdots < \hat{t}_{\hat{N}^w}^w <$

$\hat{t}^w_{\tilde{N}^w+1} = n$, thus the bootstrap version of $\hat{\mu}_1$ is

$$\hat{\mu}^w_1 = \left(\sum_{t=1}^{\hat{t}^w_1} w_t \right)^{-1} \sum_{t=1}^{\hat{t}^w_1} w_t X_t. \tag{3}$$

By repeatedly sampling from the distribution of the random weights $\mathbf{w} = (w_1, \ldots, w_n)$, for example B times, we obtain bootstrap replications $\hat{\mu}^w_{1,1}, \ldots, \hat{\mu}^w_{1,B}$. Confidence intervals for μ^0_1 can be constructed based on these bootstrap replications. First, we can use the empirical standard deviation of $\hat{\mu}^w_{1,1}, \ldots, \hat{\mu}^w_{1,B}$, denoted as $\hat{\sigma}^w_{\hat{\mu}_1}$, as the bootstrap estimate of standard error of $\hat{\mu}_1$, then the 95% naive bootstrap interval can be constructed as

$$\left[\hat{\mu}_1 - 1.96\hat{\sigma}^w_{\hat{\mu}_1},\ \hat{\mu}_1 + 1.96\hat{\sigma}^w_{\hat{\mu}_1}\right]. \tag{4}$$

The naive interval also inherits the instability of the estimator $\hat{\mu}_1$, see Figure 1.

We use the bagging estimator,

$$\tilde{\mu}_1 = \frac{1}{B} \sum_{b=1}^{B} \hat{\mu}^w_{1,b}, \tag{5}$$

to overcome the instability of the estimator $\hat{\mu}_1$ by averaging over the bootstrap replications. Efron (2014) developed a standard deviation formula for the bagging estimator based on the infinitesimal jackknife method. The infinitesimal jackknife, originally proposed by Jaeckel (1972), is a tool for approximating the variance of an estimator. Using the idea of the infinitesimal jackknife, Efron (1992) obtained accuracy measures of various bootstrap estimators. Recently, Giordano et al. (2019) and Wager et al. (2014) used the infinitesimal jackknife to estimate the variability of machine learning algorithms. Next, we generalize this formula in the weighted bootstrap sampling scheme. To illustrate the ideas, we will consider exponential weights.

Proposition 1. *Suppose the weights $w_i, i = 1, \ldots, n$ are independent and identically distributed as standard exponential distribution $Exp(1)$, the variance of $\tilde{\mu}_1$, approximated by*

the infinitesimal jackknife method, is,

$$(\tilde{\sigma}_{\tilde{\mu}_1})^2 = \sum_{j=1}^{n} cov_j^2, \tag{6}$$

where $cov_j = \mathrm{cov}(\hat{\mu}_1^w, \bar{w} - w_j | \boldsymbol{X})$ *and* $\bar{w} = n^{-1}\sum_{t=1}^{n} w_t$.

The proof of Proposition 1 is presented in the appendix. The covariance cov_j can be easily estimated by the sample covariance of $\hat{\mu}_{1,b}^w,\ b = 1, \ldots, B$ and the jth component of the weights. Denote the bth replication of the bootstrap weights as $\mathbf{w}_b = (w_{b1}, \ldots, w_{bn})$, and its average is denoted as $\bar{w}_b = n^{-1}\sum_{j=1}^{n} w_{bj}$, then $\widehat{cov}_j = B^{-1}\sum_{b=1}^{B}(\hat{\mu}_{1,b}^w - \tilde{\mu}_1)(\bar{w}_b - w_{bj})$. With the standard deviation of $\tilde{\mu}_1$, we define the 95% smoothed interval as

$$\left[\tilde{\mu}_1 - 1.96\tilde{\sigma}_{\tilde{\mu}_1},\ \tilde{\mu}_1 + 1.96\tilde{\sigma}_{\tilde{\mu}_1}\right]. \tag{7}$$

To determine whether the bootstrap sample size B is large enough to guarantee the accuracy of $\tilde{\sigma}_{\tilde{\mu}_1}$, we can use the so called *jackknife-after-bootstrap* method. We refer the reader to Efron (2014, page 996) and Davison & Hinkley (1997, Section 3.10) for details.

Another promising confidence interval is the percentile interval,

$$\left[\hat{\mu}_1^{w,0.025},\ \hat{\mu}_1^{w,0.975}\right], \tag{8}$$

where $\hat{\mu}_1^{w,0.025}$ and $\hat{\mu}_1^{w,0.975}$ are the empirical 0.025 and 0.975 quantiles of the bootstrap replications $\hat{\mu}_{1,1}^w, \ldots, \hat{\mu}_{1,B}^w$, respectively. Note that we do not need a point estimate of μ_1^0 when constructing the percentile interval, and the percentile interval is more robust in misspecification of the change-points. In Section 3, we will prove the asymptotic consistency of the percentile interval and the smoothed interval. As demonstrated by the simulation study, in small samples, the percentile interval is generally wider than the smoothed interval, while it has more conservative empirical coverage probabilities.

2.3 The intensity plot, BootCp estimator and Adaptive choice of intervals

We found in the simulation studies that the smoothed interval has coverage probability close to the nominal level when the signals are moderate or strong, but may have low coverage probabilities otherwise. To correct the coverage probability of the smoothed interval, we introduce the intensity score, the bootstrap change-point (BootCp) estimator and the adaptive interval.

We first define the intensity score. For $t \in \{1, \ldots, n\}$ and $b = 1, \ldots, B$, let $I_{t,b} = 1$, if t is estimated as a change-point in the b-th bootstrap replication, otherwise let $I_{t,b} = 0$. The intensity score of a location $t \in \{1, \ldots, n\}$ is defined as $\tilde{p}_t = B^{-1} \sum_{b=1}^{B} I_{t,b}$. The intensity score $\tilde{p}_t$ reflects the possibility that a location t is, or is close to, a change-point. By Theorems 1 and 2 in Section 3, the estimated change-points in the bootstrap replications will be concentrated in $o(\log(n))$ neighborhoods of the true change-points almost surely. Thus the intensity score $\tilde{p}_t$ is significantly greater than 0 when t is near a true change-point, and is close to 0 when t is far away from any true change-points.

The intensity scores $\tilde{p}_t, t \in \{1, \ldots, n\}$ are usually grouped into clusters around true change-points. To visualize the pattern of the estimated change-points, we define the intensity plot as a plot of $\tilde{p}_t$ versus t. The lower panel of Figure 1 gives an illustration of the intensity plot. We can see that the intensity scores are large for locations that are close to the true change-points $40, 80$ and 120, and small otherwise. The intensity plot can also be used to discover possible missing change-points. If some regions of the intensity plot are 'significant' to human eyes, but no change-points are detected by an algorithm, the researcher can investigate those regions either by data analysis or expert's knowledge. We illustrate the usage of the intensity plot in Figure 2, where the data is simulated from Model 2 of Section 4. The change-points 421 and 491 are not detected by the BootCp estimator defined in the following, while the intensities of the locations that are close to

these change-points are not zero.

[Figure 2 about here.]

We propose a new change-point estimator by identifying local maximizers of the intensity scores, we call it the BootCp estimator. Define the location t as a h-local maximizer of $\tilde{p}_s, s = 1, \ldots, n$ if $\tilde{p}_t \geq \tilde{p}_s$, for $|s - t| \leq h$. For a fixed $h > 0$, denote the set of all the h-local maximizers as $\mathcal{L}(h)$. For a threshold value $\lambda > 0$, let $\mathcal{T}(h, \lambda)$ be a subset of $\mathcal{L}(h)$, such that $\tilde{p}_t > \lambda$ for any $t \in \mathcal{T}(h, \lambda)$. The locations in the set $\mathcal{T}(h, \lambda)$ are defined as the BootCp change-point estimators. This algorithm is an adaptation of the screening and ranking algorithm (SaRa) of Niu & Zhang (2012). The difference is that we replace the local statistics of Niu & Zhang (2012) by the intensity scores. The computational complexity of this algorithm is $O(n)$ as shown by Niu & Zhang (2012).

The parameter h works as a bandwidth parameter. As proved in Section 3, all change-points can only be estimated up to an order of $o_{\mathbb{P}_w}(\log(n))$. Thus any location in a neighborhood of a true change-point will have a positive probability to be estimated as a change-point in the bootstrap replications. The parameter h determines the size of the neighborhood. The parameter λ is a threshold parameter, and it serves as a lower bound of the intensity scores that are considered as significant.

We use an information criterion function to select the parameters h and λ as in Niu & Zhang (2012). It has been shown that under mild conditions the BIC criterion leads to consistent estimation of the number of change-points (Yao, 1988; Fryzlewicz, 2014). Denote the elements of $\mathcal{T}(h, \lambda)$ as $\mathcal{T}(h, \lambda) = \{\tilde{t}_1 < \tilde{t}_2 < \cdots < \tilde{t}_{\tilde{N}}\}$. Let $\tilde{\mu}_t = (\tilde{t}_{k+1} - \tilde{t}_k)^{-1} \sum_{\tilde{t}_k+1}^{\tilde{t}_{k+1}} X_t$, for $t \in [\tilde{t}_k + 1,\ \tilde{t}_{k+1}]$. The BIC criterion for selecting the parameters h, λ is defined as

$$\mathrm{BIC}(h, \lambda) = \frac{n}{2} \log \left(\frac{1}{n} \sum_{t=1}^{n} (X_t - \tilde{\mu}_t)^2 \right) + \tilde{N} \log(n), \tag{9}$$

where $\tilde{\mu}_t$ and $\tilde{N}$ are estimates based on (h, λ). In practice, we use a two dimensional grid search to minimize the BIC criterion function. In our application, the SaRa algorithm has two nesting properties which make the parameter tuning very fast. That is, we have $\mathcal{L}(h_2) \subset \mathcal{L}(h_1)$ if $h_2 > h_1$, and $\mathcal{T}(h, \lambda_2) \subset \mathcal{T}(h, \lambda_1)$ if $\lambda_2 > \lambda_1$.

Another strategy for selecting h and λ is the multi-bandwidth SaRa proposed by Niu & Zhang (2012). In the multi-bandwidth SaRa, a pool of candidate change-points is first created by identifying local maxima of the intensity scores using multiple values of the bandwidth h and the threshold λ. In the second step, the best subset selection along with the BIC criterion (9) can be applied to screen the candidate pool and obtain the final change-point estimates. The multi-bandwidth SaRa can adapt to more complicated data but may need more computation because in the second step the best subset selection is used. In the following, we stick to the basic SaRa for simplicity.

For the BootCp estimator, along with the change-point estimates $\tilde{t}_k$, we have the corresponding intensity scores $\tilde{p}_{\tilde{t}_k}$. A low value of $\tilde{p}_{\tilde{t}_1}$ indicates that we can not detect the true change-point t_1^0 with a high probability in the bootstrap replications, the bagging estimator $\tilde{\mu}_1$ are biased, and the smoothed interval for μ_1^0 will then have coverage probability lower than the nominal level. We propose to use the percentile interval to replace the smoothed interval in this circumstance, because the percentile interval is more robust to the misspecification of the change-points. For a location $t \in \{1, \ldots, n\}$, we construct a confidence interval for μ_t^0 as follows. For $t \in [\tilde{t}_k + 1,\ \tilde{t}_{k+1}]$, if $\min(\tilde{p}_{\tilde{t}_k}, \tilde{p}_{\tilde{t}_{k+1}}) > 0.5$ then we use the smoothed interval as the confidence interval of μ_t^0, otherwise we use the percentile interval. We call this strategy adaptive interval estimation. From Figure 1, we can see that the empirical coverage probabilities of the adaptive interval are close to the nominal level 95% except for those locations in the neighborhood of a true change-point.

In practice, the intensity scores of the BootCp estimator can help the practitioner ex-

plaining why an interval is chosen. For an estimated change-point with a low intensity score, the bagging estimator of the mean for nearby locations will be severely biased. Thus the resulting smoothed interval can not cover the true mean with the nominal probability. Therefore, we replace the smoothed interval with the percentile interval. For change-point estimates with high intensity scores, we trust the smoothed interval which is shorter than the percentile interval. The adaptive intervals ensure good empirical coverage probabilities compared to the smoothed interval. A schematic outline of the proposed procedure is listed in Algorithm 1.

Algorithm 1: Change-points detection and confidence interval construction

Input: Data: $\mathbf{X} = \{X_1, \ldots, X_n\}$; Number of bootstrap: B; Maximum number of change-points: $\mathcal{N}$; Confidence level: $1 - \alpha$.

Output: Confidence intervals; BootCp estimator; Intensity plot.

for $i = 1 : B$ **do**

Sample $\mathbf{w}_i = (w_{i1}, \ldots, w_{in})$;

Estimate change-points $\hat{t}^w_{i1} < \cdots < \hat{t}^w_{i\hat{N}^w_i}$ using the sequential BS;

Estimate $\hat{\mu}^w_{1,i}$ using (3);

end

Estimate $\alpha/2$ and $1 - \alpha/2$ sample quantiles of $(\hat{\mu}^w_{1,1}, \ldots, \hat{\mu}^w_{1,B})$. Output the percentile interval;

Estimate the bagging estimator $\tilde{\mu}_1$ and its standard error using (5) and (6), respectively. Output the smoothed interval;

Calculate the intensity score $\tilde{p}_t, t = 1, \ldots, n$, and output the intensity plot;

Calculate the BootCp estimator $\tilde{t}_1 < \cdots < \tilde{t}_{\tilde{N}}$. Output the BootCp estimator and the adaptive interval.

3 Asymptotic Theory

In this section, we establish the large-sample property of the weighted bootstrap change-point estimator and the asymptotic validity of the proposed confidence intervals. All the proofs are collected in the supplementary material. We need the following assumptions.

Assumption A1, *The error terms* $\epsilon_t, t = 1, \ldots, n$ *are i.i.d. with mean 0 and finite variance* σ^2*, and the fourth moment is finite* $E|\epsilon_t|^4 < \infty$.

Assumption A2, *The true change-points are* $t_i^0 = \lfloor n\tau_i^0 \rfloor, i = 1, \ldots, N$*, with* $0 < \tau_1^0 < \tau_2^0 <, \ldots, \tau_N^0 < 1$*, and* $\beta_i^0 \neq \beta_{i+1}^0, i = 1, \ldots, N$*. The number of change-points* N *does not change with the sample size* n*.*

Assumption A3, *The bootstrap weights* $(w_1, \ldots, w_n)$ *are i.i.d., and strictly positive random variables. The bootstrap weights are independent of the data* $X_t, t = 1, \ldots, n$*. The mean and variance of* w_t *are 1,* $E|w_t|^{2+\varepsilon} < \infty$ *for some* $\varepsilon > 0$*.*

Assumptions A1 and A2 are standard in the literature, see Bai (1997). We focus on independent data in this paper, as stated in assumption A1. Extending the current methodology to time series data is challenging, and we leave that for future research. Assumption A2 implies that the number of change-points is finite, and each segments contain a positive fraction of the whole sample. This assumption is appropriate when the change-points are fixed, e.g., dates or locations, and it is possible to collect more observations between change-points. For example, in copy number variation studies, the resolution of comparative genomic hybridization (CGH) data improves as technologies rapidly develop, while the locations of the change-points, which are the locations of the copy number variations, can be considered as fixed. See also Perron (2006) for an in-depth discussion of assump-

tion A2. Assumption A3 is commonly assumed in the weighted bootstrap literature, see van der Vaart & Weller (1996).

The large sample theory of binary segmentation algorithm is developed in Bai (1997) and Fryzlewicz (2014), where consistency and rate of convergence for the change-point estimator are established. We will study the asymptotic behavior of the bootstrap change-point estimator and the asymptotic consistency of the proposed confidence intervals. We first work under the assumption that the number of change-points is known. Later we will show the consistency of BIC criterion in selecting the number of change-points.

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a fixed probability space, all the random variables are defined on this space. In this paper, the probability measure $\mathbb{P}$ is understood to be the product measure of the bootstrap weights and the observations. For convenience, let $\mathbb{P}_w(.)$ and $\mathbb{E}_w(.)$ denote the conditional probability measure and the conditional expectation with respect to the bootstrap weights $\{w_t\}_{t=1}^n$ conditional on the data $\{X_t\}_{t=1}^n$, respectively. The abbreviation $a.s.$ stands for almost surely as usual. A sequence of random variables Z_n is denoted as $o_{\mathbb{P}_w}(1)$ if $\lim_{n\to\infty} \mathbb{P}_w(|Z_n| > \epsilon) = 0$ almost surely for any $\epsilon > 0$. Similarly, $Z_n = O_{\mathbb{P}_w}(1)$ if for any $\epsilon > 0$, there exists $M > 0$, such that $\limsup_{n\to\infty} \mathbb{P}_w(|Z_n| > M) < \epsilon, a.s.$. Denote $Z_n(k)$ as a measurable function of $\{w_t\}_{t=1}^n$, $\{X_t\}_{t=1}^n$, and an integer $k \in [a(n), b(n)]$, where $1 \le a(n) \le b(n) \le n$. We say that $Z_n(k)$ is $o_{\mathbb{P}_w}(1)$ uniformly over $k \in [a(n), b(n)]$ if $\sup_{k\in[a(n),b(n)]} |Z_n(k)| = o_{\mathbb{P}_w}(1)$. Denote $\sigma(X_1, \ldots, X_n)$ as the sigma field generated by $X_1, \ldots, X_n$. By a straightforward argument, if Z_n is $\sigma(X_1, \ldots, X_n)$ measurable, then $Z_n = o_{\mathbb{P}_w}(1)$ is equivalent to Z_n converges to zero almost surely, see Lemma 8(iv) of the supplementary material for a proof.

Theorem 1. *Assume assumptions A1-A3, and that the true number of change-points N is known. The change-points estimators $\hat{t}_1^w, \ldots, \hat{t}_N^w$ have the convergence rate* $\max_{1\le i\le N} |\hat{t}_i^w -$

$t_i^0| = o_{\mathbb{P}_w}(\log(n))$. *That is, for every* $\varepsilon > 0$,

$$\lim_{n\to\infty} \mathbb{P}_w \left(\bigcup_{i=1}^{N} \{|\hat{t}_i^w - t_i^0| > \varepsilon \log(n)\} \right) = 0 \quad a.s.$$

Theorem 1 establishes that $\hat{t}_i^w$ will be concentrated in an $o_{\mathbb{P}_w}(\log(n))$ neighborhood of the true change-point. The asymptotic results on the convergence rate of change-point estimators in the bootstrap setting is rare. We can compare with the convergence rates of change-point estimators in the non-bootstrap setting. Bai (1997) obtained a convergence rate of $O_{\mathbb{P}}(1)$ for the change-points estimator under similar assumptions. Fryzlewicz (2014) derived a convergence rate of $O_{\mathbb{P}}(n^2\delta_n^{-2}(\underline{f}_n)^{-2}\log(n))$ for the binary segmentation algorithm, and a convergence rate of $O_{\mathbb{P}}((\underline{f}_n)^{-2}\log(n))$ for the wild binary segmentation algorithm, where $\min_{i=0,\ldots,N} |t_{i+1}^0 - t_i^0| \geq \delta_n$ and $\min_{i=0,\ldots,N} |\beta_{i+1}^0 - \beta_i^0| \geq \underline{f}_n$. If assumption A2 is true, these two convergence rates are both of the order $O_{\mathbb{P}}(\log(n))$. The convergence rate we proved is sharp compared to these results up to a factor of $\log(n)$.

The intensity score $\tilde{p}_k$ defined in Section 2.3 will converge to 0 almost surely if k is not in a $\log(n)$ neighborhood of a true change-point. To see this, for some k where $\min_{i=1,\ldots,N} |k - t_i^0| > \log(n)$, note that

$$\tilde{p}_k = \mathbb{P}_w \left(\cup_{i=1}^{N} \{\hat{t}_i^w = k\} \right) \leq \mathbb{P}_w(\cup_{i=1}^{N} \{|\hat{t}_i^w - t_i^0| > \log(n)\}) \to 0 \quad a.s.$$

Thus, empirically $\tilde{p}_k$ can serve as an indicator of whether the location k is, or close to, a change-point. If $\tilde{p}_k$ is significantly greater than 0, there may exist a change-point close to k. On the contrary, if $\tilde{p}_k$ is close to 0, there are no change-points around k.

Theorem 1 is proved under the assumption that we know the true number of change-points N. In practice, we use the BIC criterion to determine the number of change-points as described in Section 2.2. Recall that $\hat{N}^w = \arg\min \mathrm{BIC}(m)$, the following results state the consistency of $\hat{N}^w$.

Theorem 2. *Under assumptions A1-A3, assume that the number of change-points is finite, then*

$$\lim_{n\to\infty} \mathbb{P}_w(\hat{N}^w = N) = 1, \quad a.s.$$

Yao (1988) established the consistency of a similar criterion for estimating the number of change-points N in a least squares regression framework. Fryzlewicz (2014) proved the consistency of a strengthened Schwartz Information Criterion for selecting the number of change-points with a binary segmentation algorithm. These results do not apply to the bootstrap procedure. We generalize the consistency results for the BIC criterion to the weighted bootstrap estimator. By Theorem 1 and 2, we can prove the asymptotic properties of the percentile confidence interval and the smoothed confidence. A confidence interval $[\hat{\mu}_{n,1}, \hat{\mu}_{n,2}]$ for a parameter μ is asymptotically consistent at level $1-\alpha$ if $\liminf_{n\to\infty} \mathbb{P}(\hat{\mu}_{n,1} \leq \mu \leq \hat{\mu}_{n,2}) \geq 1-\alpha$, see Section 23.2 of Van der Vaart (2000).

Theorem 3. *Under assumptions A1-A3, for any location $t_n \in \{1, \ldots, n\}$ such that $\min_{i=1,\ldots,N} |t_n - t_i^0| > \log(n)$, the percentile confidence interval, the smoothed confidence interval, and the adaptive confidence interval for μ_t^0 are all asymptotically consistent.*

Theorem 3 establishes the asymptotic consistency of the confidence intervals for μ_t^0 when t is not close to any true change-points. Theorem 1 implies that in a $\log(n)$ neighborhood of the true change-points, an estimator of the mean defined as the average of the observations between two adjacent estimated change-points is severely biased. This is the intuition why we can only establish the consistency of the proposed confidence intervals for the locations that are not in a $\log(n)$ neighborhood of the true change-points.

In practice, we can construct confidence intervals for μ_t^0 using either the percentile confidence interval, the smoothed confidence interval or the adaptive confidence interval as long as t is away from any estimated change-points. For any location t that is close

to a true change-point, the confidence interval for μ_t^0 will generally be biased, that is, the empirical coverage probabilities will be lower than the specified level, because we can not estimate μ_t^0 consistently. As we can see from the simulation study in Section 4, the length of the confidence interval for μ_t^0 will be wide and the empirical converge probability will be lower than the specified level for any location t that is close to a true change-point.

4 Simulation Study

In this section, we use simulation studies to illustrate the finite sample performance of the proposed method. In Example 1, we compare the BootCp estimator with the state-of-art change-point detection algorithms in the literature. In Example 2, we study the performance of the proposed confidence intervals. In all the experiments, the weights are sampled from the standard exponential distribution.

Example 1. Consider a piece-wise constant signal of length n, the errors are independently normally distributed with mean 0 and standard deviation σ. We simulate 500 datasets from each of the following models. See Figure 3 for the mean of the models and one sample of each model.

Model 1. The sample size $n = 497$ and $\sigma = 0.3$, the change-points are (139, 226, 243, 300, 309, 333), the means between change-points are (-0.18, 0.08, 1.07, -0.53, 0.16, -0.69, -0.16).

Model 2. The sample size $n = 560$ and $\sigma = 4$, the change-points are (11, 21, 41, 61, 91, 121, 161, 201, 251, 301, 361, 421, 491), the means between change-points are (7, -7, 6, -6, 5, -5, 4, -4, 3, -3, 2, -2, 1, -1).

Model 3. The sample size $n = 140$ and $\sigma = 0.4$, the change-points are (11, 21, 31, 41, 51, 61, 71, 81, 91, 101, 111, 121, 131), the means between change-points are (0, 1, 0, 1, 0, 1, 0, 1, 0, 1, 0, 1, 0, 1).

Model 4. The sample size $n = 150$ and $\sigma = 0.3$, the change-points are (11, 21, 31, 41, 51, 61, 71, 81, 91, 101, 111, 121, 131, 141), the means between change-points are (1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15).

[Figure 3 about here.]

We compare BootCp with the following change-point detection methods. The competing methods are implemented by the R (R Core Team, 2019) packages mentioned below, most of which are available on CRAN. See the supplementary material for a detailed description of how do we implement the competing methods.

- the binary segmentation (BS) and wild binary segmentation (WBS) algorithm of Fryzlewicz (2014), which are implemented in the R package `wbs` (Baranowski & Fryzlewicz, 2015);
- the fused lasso (FLASSO) of Harchaoui & Lévy-Leduc (2010), which is implemented in the R package `genlasso` (Arnold & Tibshirani, 2014);
- the SMUCE procedure of Frick et al. (2014), which is implemented in the R package `stepR` (Pein, Hotz, Sieling & Aspelmeier, 2017);
- the cumSeg procedure of Muggeo & Adelfio (2011), which is implemented in the R package `cumSeg` (Muggeo, 2012);
- the PELT procedure of Killick et al. (2012), which is implemented in the R package `changepoint` (Killick et al., 2016);
- the S3IB method, implemented in the R package `Segmentor3IsBack` (Cleynen et al., 2016);
- the screening and ranking algorithm (SaRa) of Niu & Zhang (2012).

For the proposed BootCp estimator, the number of bootstrap samples is $B = 5000$ and the parameters λ and h are selected by the BIC criterion described in Section 2.3.

Besides summary statistics of $\hat{N} - N$ for all methods, we also use the Hausdorff distance d_H to measure the estimation accuracy of the change-points. Let A and B be two nonempty sets of integers, the Hausdorff distance d_H is defined as

$$d_H(A, B) = \max\{\sup_{a \in A} \inf_{b \in B} |a - b|, \sup_{b \in B} \inf_{a \in A} |a - b|\}.$$

We evaluate the mean and standard deviation of the Hausdorff distance between the set of true change-points and the set of estimated change-points.

The results are shown in Table 1 and Table 2. The proposed BootCp estimator has the best overall performance, followed by three other methods WBS, S3IB, PELT, and SaRa. For Model 1, BootCp recovers the true number of change-points in 94.2% of the replications, and it is comparable to that of WBS and S3IB. For the estimation accuracy d_H, BootCp is also close to the best competitors WBS and S3IB. For Models 2-4, BootCp outperforms all the other methods in recovering the number of change-points and in estimation accuracy. Model 2 is challenging for all the methods. BootCp detects the true number of change-points in 49% of the replications. The mean of d_H is 51.12, which is the smallest among all the procedures. For Model 3 and Model 4, BootCp recovers the true number of change-points around 90% of the times. Several procedures have large biases in estimating the number of change-points. The average computational time (in seconds) are listed in Table 3. Since the bootstrap method is used in our approach, it is slower than other methods. With modern computing devices, it is easy to parallelize the bootstrapping method such that it can work for massive data sets. More simulation studies with correlated errors are shown in the supplementary material.

[Table 1 about here.]

[Table 2 about here.]

[Table 3 about here.]

Example 2. The observations are a sequence of length $n = 160$, with the change-points at $40, 80, 120$. The means between the change-points are $\{1, 2, 1, 2\}$, respectively. The errors are normally distributed or double exponentially distributed. The standard deviation σ of the signal is 0.75 or 1. For the proposed adaptive interval, the parameters h, λ are selected by the BIC criterion proposed in Section 2.3. We simulate 500 data sets under each setting. The bootstrap sample size is $B = 5000$. We construct 95% confidence intervals for $\mu_t^0, t = 1, \ldots, 160$ by using the smoothed intervals, the percentile intervals and the adaptive type intervals. We also compare our methods with the SMUCE interval proposed by Frick et al. (2014). The empirical coverage probabilities of the confidence intervals for $\mu_t^0, t = 1, \ldots, 160$ are shown in Figure 4, and the average lengths of the constructed intervals are depicted in Figure 5. Other methods do not provide confidence intervals, and therefore are not included in the comparison.

Generally speaking, the adaptive interval outperforms the other methods. The coverage probabilities of the adaptive interval are the closest to the nominal level 95% for all four cases. The SMUCE interval does not cover the true value in a high frequency, its coverage probabilities are generally around 0.7, except for normal errors with $\sigma = 0.75$, where its coverage probabilities are close to 0.8. The coverage probabilities of the smoothed confidence intervals are lower than the nominal size 95% in general, especially in the case where $\sigma = 1$. The coverage probabilities of the percentile interval are generally higher than the nominal level 95%. The average length of the smoothed interval and the SMUCE interval are the shortest among the four intervals. The average length of the percentile intervals are twice the average length of the smoothed interval. As predicted by the theory,

the methods do not give satisfactory confidence intervals for μ_t^0 when t is close to one of the true change-points $40, 80, 120$, in which case the estimator for μ_t^0 is generally biased.

[Figure 4 about here.]

[Figure 5 about here.]

5 Application to the DNA copy number data

In this section, we will analyze the comparative genomic hybridization (CGH) data to illustrate the possible applications of the proposed procedure. Copy number variations (CNV) refer to amplification and deletions of chromosome segments, which constitute a major source of variation between individual humans and contribute to many diseases, see Hastings et al. (2009). The alteration of the copy number for specific genes may cause genomic disorders, which are related to the formulation and progression of cancer and many other diseases. Advances in technologies such as microarray comparative genomic hybridization (CGH) make it possible to compare the cancer cell lines to the normal cell lines with high resolution. The CGH data record log ratio of the testing sample copy number to the reference sample copy number. A value higher than 0 in regions of a chromosome indicates amplification of the copy number to the reference, while a value less than 0 indicates the deletion of the copy number to the reference.

CGH data are often modeled as a piece-wise constant function, and segmentation methods can be implemented to detect the possible amplifications and deletions in a CGH signal, see Niu & Zhang (2012) and Frick et al. (2014). In this section, we apply the proposed method to detect the change-points in a CGH data studied by Snijders et al. (2001) who analyzed the genomic aberrations of breast cancer. The original data are shown in the upper panel of Figure 6, and the adaptive confidence intervals for the mean are plotted in shaded areas. The intensity plot is depicted in the lower panel of Figure 6. From

the intensity plot, we can readily see the major locations of the change-points (Figure 6, lower panel). For example, we find some obvious change-points, such as genome position 123841 and 228492 at chromosome 1, genome position 198680 and 224174 at chromosome 2 and genome position 42560 and 63000 at chromosome 7.

In array CGH data analysis, we are interested in determining segments of genome locations on which the log ratios differ from zero. The confidence sets can be used with the BootCp change-points estimator to determine these intervals. A segment shows a sign of amplification if the confidence band is above zero, and it shows a sign of deletion if the confidence band is below zero. The BootCp algorithm detects 23 change-points. Moreover, we use the confidence bands to determine the significant segments of genes. From Figure 6, we conclude that the amplification segments are genome positions 123841 to 228492 at chromosome 1, genome positions 86135 to 10050 at chromosome 13 and genome positions 57063 to 86000 at chromosome 18. The deletion segments are genome positions 198680 to 224174 at chromosome 2, genome positions 42560 to 63000 at chromosome 7 and genome positions 3292 to 56563 at chromosome 14. These findings are consistent with those of Snijders et al. (2001). Compared to other segmentation algorithms, we not only detect these segments but also provide uncertainty assessment through confidence intervals.

[Figure 6 about here.]

6 Conclusion

The inherent uncertainty in change-point detection makes it challenging to construct confidence intervals for the mean of a noisy sequence. We used the weighted bootstrap to generalize the bagging estimator and developed a standard deviation formula for the proposed estimator. We proposed an adaptive interval which chooses between the smoothed

confidence interval and the percentile interval to correct the coverage probability of the smoothed interval. We proposed a new intensity plot, which is a visualization tool for the change-point detection process. Through the intensity plot, we have a second chance to detect possible missing change-points. Based on the intensity plot, we proposed a new change-point estimator, the bootstrap change-point estimator (BootCp). The BootCp estimator has excellent performance in the simulation study compared to some state-of-art segmentation methods.

Acknowledgments

We thank the associate editor and two anonymous reviewers for constructive comments. This work was partially supported by the Natural Science Foundation of China (11671096, 11731011, 11690013).

Supporting Information

Additional information for this article is available online. Web Appendix A contains details of the simulation study and additional simulation results. Web Appendix B contains all the proofs.

References

Arnold, T. B. & Tibshirani, R. J. (2014). *genlasso: Path algorithm for generalized lasso problems*. R package version 1.3.
URL: *https://CRAN.R-project.org/package=genlasso*

Bai, J. (1997). Estimating multiple breaks one at a time, *Econometric Theory* **13**(3): 315–352.

Baranowski, R. & Fryzlewicz, P. (2015). *wbs: Wild Binary Segmentation for Multiple*

Change-Point Detection. R package version 1.3.
URL: *https://CRAN.R-project.org/package=wbs*

Carlstein, E. et al. (1986). The use of subseries values for estimating the variance of a general statistic from a stationary sequence, *Ann. Statist.* **14**(3): 1171–1179.

Chatterjee, S., Bose, A. et al. (2005). Generalized bootstrap for estimating equations, *Ann. Statist.* **33**(1): 414–436.

Chernoff, H. & 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.

Chernozhukov, V., Chetverikov, D., Kato, K. et al. (2013). Gaussian approximations and multiplier bootstrap for maxima of sums of high-dimensional random vectors, *Ann. Statist.* **41**(6): 2786–2819.

Chib, S. (1998). Estimation and comparison of multiple change-point models, *J. Econometrics* **86**(2): 221–241.

Cleynen, A., Rigaill, G. & Koskas, M. (2016). *Segmentor3IsBack: A Fast Segmentation Algorithm*. R package version 2.0.
URL: *https://CRAN.R-project.org/package=Segmentor3IsBack*

Davison, A. C. & Hinkley, D. V. (1997). *Bootstrap Methods and their Application*, Cambridge: Cambridge University Press.

Du, C., Kao, C.-L. M. & Kou, S. (2016). Stepwise signal extraction via marginal likelihood, *J. Amer. Statist. Assoc.* **111**(513): 314–330.

Efron, B. (1982). *The Jackknife, the Bootstrap, and other Resampling Plans*, Philadelphia: SIAM.

Efron, B. (1992). Jackknife-after-bootstrap standard errors and influence functions, *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **54**(1): 83–111.

Efron, B. (2014). Estimation and accuracy after model selection, *J. Amer. Statist. Assoc.* **109**(507): 991–1007.

Freedman, D. (1984). On bootstrapping two-stage least-squares estimates in stationary linear models, *Ann. Statist.* **12**(3): 827–842.

Frick, K., Axel, M. & Hannes, S. (2014). Multiscale change point inference, *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **76**(3): 495–580.

Fryzlewicz, P. (2014). Wild binary segmentation for multiple change-point detection, *Ann. Statist.* **42**(6): 2243–2281.

Giordano, R., Stephenson, W., Liu, R., Jordan, M. & Broderick, T. (2019). A swiss army infinitesimal jackknife, *The 22nd International Conference on Artificial Intelligence and Statistics*, Vol. 89, pp. 1139–1147.

Harchaoui, Z. & Lévy-Leduc, C. (2010). Multiple change-point estimation with a total variation penalty, *J. Amer. Statist. Assoc.* **105**(492): 1480–1493.

Hastings, P., Lupski, J. R., Rosenberg, S. M. & Ira, G. (2009). Mechanisms of change in gene copy number, *Nat. Rev. Genet.* **10**(8): 551–564.

Hlávka, Z., Hušková, M., Kirch, C. & Meintanis, S. G. (2016). Bootstrap procedures for online monitoring of changes in autoregressive models, *Comm. Statist. Simulation Comput.* **45**(7): 2471–2490.

Horowitz, J. L. (2019). Bootstrap methods in econometrics, *Annual Review of Economics* **11**: 193–224.

Hušková, M. & Kirch, C. (2010). A note on studentized confidence intervals for the change-point, *Comput. Statist.* **25**(2): 269–289.

Jaeckel, L. (1972). The infinitesimal jackknife, *Memorandum MM72-1215-11*, Bell Laboratories, Murray Hill.

Killick, R., Fearnhead, P. & Eckley, I. (2012). Optimal detection of changepoints with a linear computational cost, *J. Amer. Statist. Assoc.* **107**(500): 1590–1598.

Killick, R., Haynes, K. & Eckley, I. A. (2016). *changepoint: An R package for changepoint analysis*. R package version 2.2.2.
URL: *https://CRAN.R-project.org/package=changepoint*

Kim, C., Suh, M.-S. & Hong, K.-O. (2009). Bayesian changepoint analysis of the annual maximum of daily and subdaily precipitation over South Korea, *J. Climate* **22**(24): 6741–6757.

Kirch, C. (2007). Block permutation principles for the change analysis of dependent data, *J. Statist. Plann. Inference* **137**(7): 2453–2474.

Kunsch, H. R. (1989). The jackknife and the bootstrap for general stationary observations, *Ann. Statist.* **17**(3): 1217–1241.

Muggeo, V. M. (2012). *cumSeg: Change point detection in genomic sequences*. R package version 1.1.
URL: *https://CRAN.R-project.org/package=cumSeg*

Muggeo, V. M. & Adelfio, G. (2011). Efficient change point detection for genomic sequences of continuous measurements, *Bioinformatics* **27**(2): 161–166.

Niu, Y. S. & Zhang, H. (2012). The screening and ranking algorithm to detect DNA copy number variations, *Ann. Appl. Stat.* **6**(3): 1306–1326.

Olshen, A. B., Venkatraman, E., Lucito, R. & Wigler, M. (2004). Circular binary segmentation for the analysis of array-based DNA copy number data, *Biostatistics* **5**(4): 557–572.

Pein, F., Hotz, T., Sieling, H. & Aspelmeier, T. (2017). *stepR: Multiscale change-point inference*. R package version 2.0-1.
URL: *https://CRAN.R-project.org/package=stepR*

Pein, F., Sieling, H. & Munk, A. (2017). Heterogeneous change point inference, *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **79**(4): 1207–1227.

Perron, P. (2006). Dealing with structural breaks, *in* K. Patterson & T. Mills (eds), *Palgrave Handbook of Econometrics*, Vol. 1, Palgrave-Macmillan, New York, pp. 278–352.

R Core Team (2019). *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria.
URL: *https://www.R-project.org/*

Rozenholc, Y. & Nuel, G. (2013). Fast estimation of posterior probabilities in change-point analysis through a constrained hidden Markov model, *Comput. Statist. Data Anal.* **68**: 129–140.

Shiryaev, A. N. (1963). On optimum methods in quickest detection problems, *Theory Probab. Appl.* **8**(1): 22–46.

Siegmund, D. (1988). Confidence sets in change-point problems, *Int. Stat. Rev.* **56**(1): 31–48.

Snijders, A. M., Nowak, N., Segraves, R., Blackwood, S., Brown, N., Conroy, J., Hamilton, G., Hindle, A. K., Huey, B. & Kimura, K. (2001). Assembly of microarrays for genome-wide measurement of DNA copy number, *Nat. Genet.* **29**(3): 263–264.

Spokoiny, V., Zhilova, M. et al. (2015). Bootstrap confidence sets under model misspecification, *Ann. Statist.* **43**(6): 2653–2675.

Van der Vaart, A. W. (2000). *Asymptotic Statistics*, Cambridge University Press.

van der Vaart, A. W. & Weller, J. A. (1996). *Weak Convergence and Empirical Processes: With Applications to Statistics*, Springer-Verlag.

Vostrikova, L. J. (1981). Detecting 'disorder' in multidimensional random process, *Soviet Mathematics Doklady* **24**: 55–59.

Wager, S., Hastie, T. & Efron, B. (2014). Confidence intervals for random forests: The jackknife and the infinitesimal jackknife, *J. Mach. Learn. Res.* **15**(1): 1625–1651.

Yao, Y.-C. (1988). Estimating the number of change-points via Schwarz'criterion, *Statist. Probab. Lett.* **6**(3): 181–189.

Corresponding author's address

Zhongyi Zhu,

Department of Statistics, Fudan University,

670 Guoshun Road, Shanghai, China.

Email:zhuzy@fudan.edu.cn.

Appendix

Proof of Proposition 1. We use the infinitesimal jackknife method as in Efron (2014) to derive a standard deviation formula for $\tilde{\mu}_1$, see also Chapter VI of Efron (1982). The parameters of the random vector $\mathbf{w} = (w_1, \ldots, w_n)$ is $\mathbf{p}_0 = (1, \ldots, 1)$, which is the mean of $\mathbf{w}$. We can also write $\mathbf{w}$ as $\mathbf{w}(\mathbf{p}_0)$ to emphasize $\mathbf{w}$ depends on the parameter $\mathbf{p}_0$. More generally, for a vector $\mathbf{p} = (p_1, \ldots, p_n)$, $\mathbf{w}(\mathbf{p}) = (w_1, ..., w_n)$ is defined as a vector of independent exponentially distributed random variables, with the marginal distribution of w_j being $Exp(p_j)$.

The population version of the bagging estimator $\tilde{\mu}_1$ is the expectation of $\hat{\mu}_1^w(\mathbf{w}, \mathbf{X}) = (\sum_{t=1}^{\hat{t}_1^w} w_t)^{-1} \sum_1^{\hat{t}_1^w} w_t X_t$ with respect to $\mathbf{w}$ conditional on the original observation $\mathbf{X}$, which we denoted as $\tilde{\mu}_1(\mathbf{p}_0) = E(\hat{\mu}_1^w(\mathbf{w}(\mathbf{p}_0), \mathbf{X})|\mathbf{X})$. We proceed similarly as that in Efron (2014) and perturb the distribution of the weights $\mathbf{w}$ to get the influence function of $\tilde{\mu}_1(\mathbf{p}_0)$. Now, we change the parameter $\mathbf{p}_0$ of $\mathbf{w}$ to $\mathbf{p} = (1 - \epsilon/n, \ldots, 1 + (n-1)\epsilon/n, \ldots, 1 - \epsilon/n)$ where the j-th element is $1 + (n-1)\epsilon/n$, and ϵ is a small perturbation. The influence function is defined as

$$\dot{\mu}_j = \lim_{\epsilon \to 0} \frac{\tilde{\mu}_1(\mathbf{p}) - \tilde{\mu}_1(\mathbf{p}_0)}{\epsilon}$$

where

$$\begin{aligned}
\tilde{\mu}_1(\mathbf{p}) &= E(\hat{\mu}_1^w(\mathbf{w}(\mathbf{p}), \mathbf{X})|\mathbf{X}) \\
&= \int_0^\infty \hat{\mu}_1^w(\mathbf{w}, \mathbf{X})(1 + (n-1)\epsilon/n)(1 - \epsilon/n)^{n-1} \times \\
&\quad \exp\left\{-(1-\epsilon/n)\sum_{i \neq j} w_i - (1 + (n-1)\epsilon/n)w_j\right\} d\mathbf{w} \\
&= \int_0^\infty \hat{\mu}_1^w(\mathbf{w}, \mathbf{X})(1 + (n-1)\epsilon/n)(1 - (n-1)\epsilon/n) \times \\
&\quad \exp\left\{-\sum_{i=1}^n w_i + \epsilon/n \sum_{i \neq j} w_i - (n-1)\epsilon/n w_j\right\} d\mathbf{w} + o(\epsilon) \\
&= \int_0^\infty \hat{\mu}_1^w(\mathbf{w}, \mathbf{X}) \exp(-\sum_{i=1}^n w_i) \exp(\epsilon(\bar{w} - w_j)) d\mathbf{w} + o(\epsilon) \\
&= \tilde{\mu}_1(\mathbf{p}_0) + \epsilon \int_0^\infty \hat{\mu}_1^w(\mathbf{w}, \mathbf{X})(\bar{w} - w_j) \exp(-\sum_{i=1}^n w_i) d\mathbf{w} + o(\epsilon),
\end{aligned}$$

and $\bar{w} = 1/n \sum_{i=1}^n w_i$. Thus we have $\dot{\mu}_j = cov(\hat{\mu}_1^w(\mathbf{w}, \mathbf{X}), \bar{w} - w_j|\mathbf{X})$. Using the formula (6.18) of Efron (1982), we obtain the standard deviation formula

$$\tilde{\sigma}^2_{\tilde{\mu}_1} = \sum_{j=1}^n cov(\hat{\mu}_1^w(\mathbf{w}, \mathbf{X}), \bar{w} - w_j|\mathbf{X})^2.$$

□

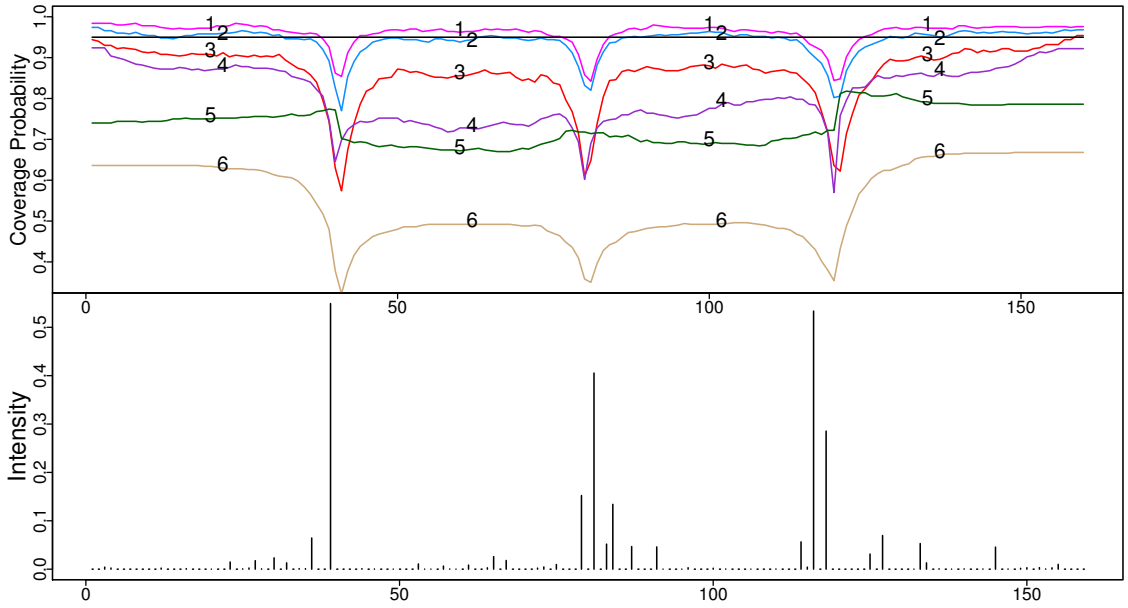


Figure 1: Upper panel: Empirical coverage probabilities of 95% confidence intervals for $\mu_1^0, \mu_2^0, \ldots, \mu_{160}^0$, line 1 for percentile interval, line 2 for adaptive interval, line 3 for smoothed interval, line 4 for naive bootstrap interval, line 5 for SMUCE interval, and line 6 for unsmoothed interval. Lower panel: intensity plot for a typical sample, locations close to the true change-points $40, 80, 120$ have high intensity scores.

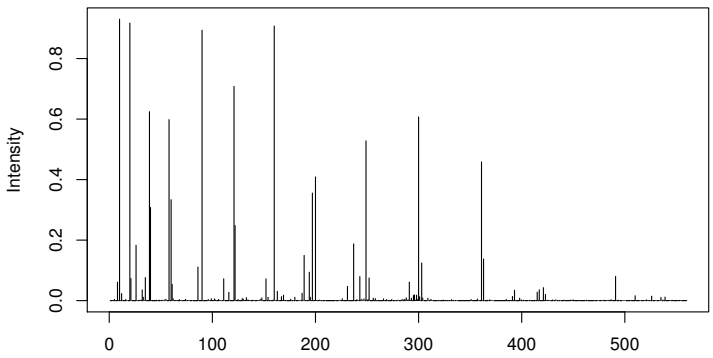


Figure 2: An illustration of the intensity plot. The BootCp estimator does not detect the change-points 421 and 491, while the intensities of the locations that are close to these change-points are not zero.

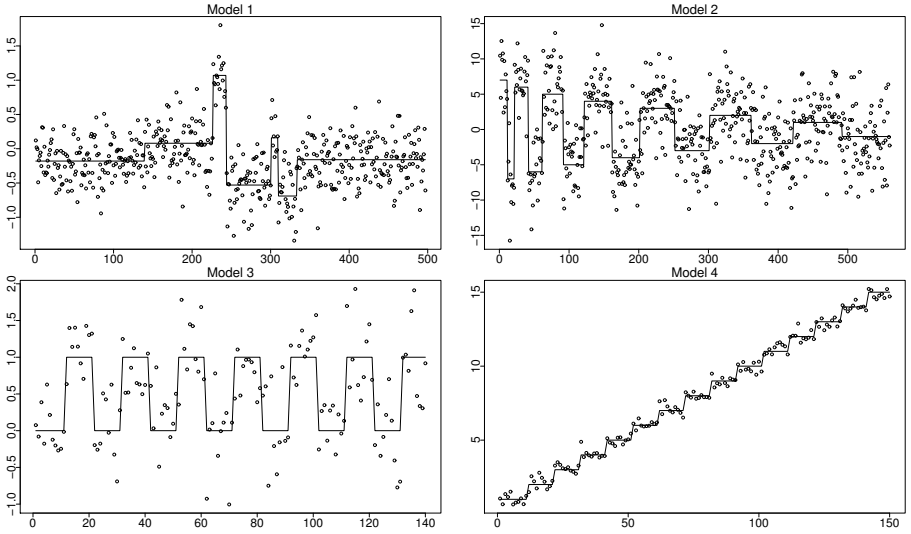


Figure 3: The signals used in example 1 with one simulated data.

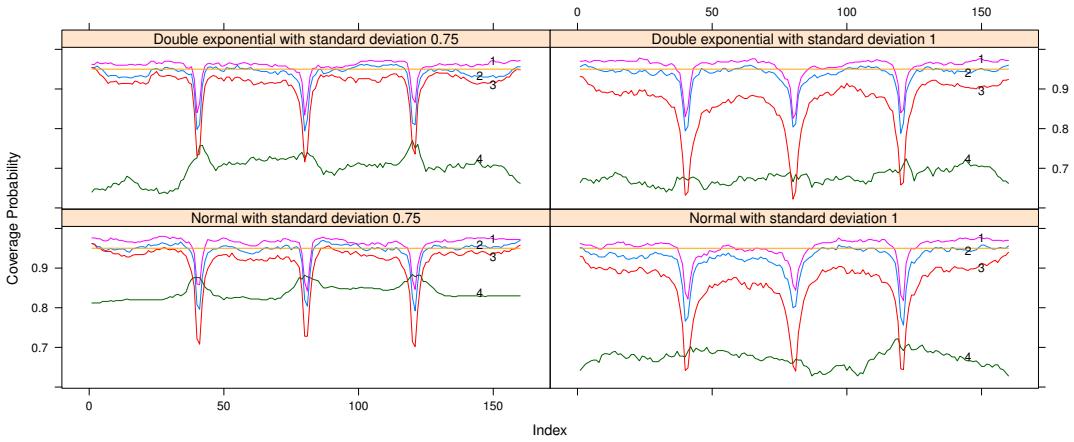


Figure 4: The empirical coverage probabilities of constructed confidence intervals based on 500 simulation runs, 1 for percentile intervals, 2 for adaptive intervals, 3 for smoothed intervals, 4 for SMUCE intervals.

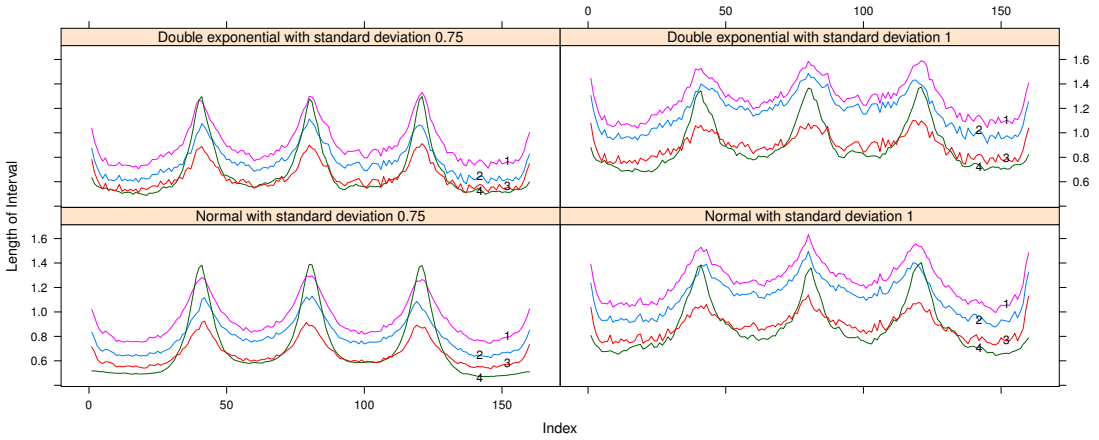


Figure 5: The average length of constructed confidence intervals based on 500 simulation runs, 1 for percentile intervals, 2 for adaptive intervals, 3 for smoothed intervals, 4 for SMUCE intervals.

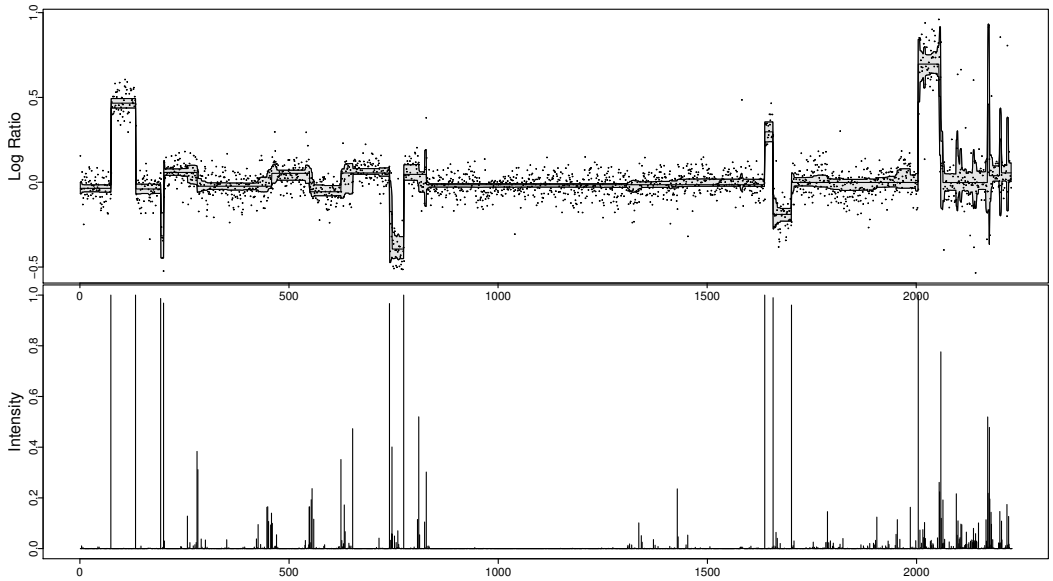


Figure 6: The S0034 cell line a-CGH data, the y-axis is the position in the genome. Upper panel, the original data with the adaptive intervals plotted in shaded areas. Bottom panel, the intensity plot.

Table 1: Summary statistics of $\hat{N} - N$ and the Hausdorff distance d_H for model 1 and 2.

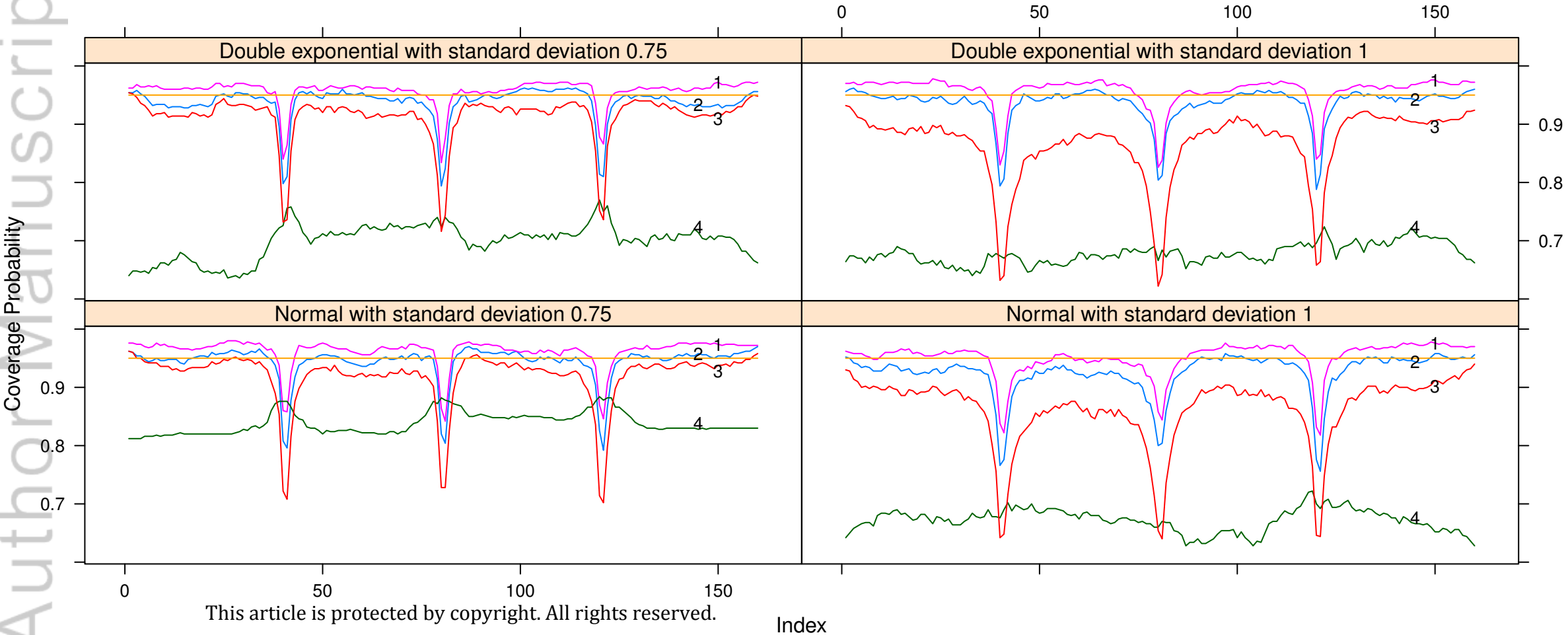
	$\hat{N} - N$							d_H	
	Frequency(%)								
Method	≤ -2	$= -1$	$= 0$	$= 1$	≥ 2	Mean	Med	Mean	SD
				Model 1					
BootCp	0.0	0.6	**94.2**	2.0	3.2	0.08	0	9.32	16.31
BS	34.0	1.6	60.4	4.0	0.0	-0.66	0	16.62	16.65
WBS	0.4	0.0	**95.4**	3.6	0.6	0.04	0	**7.24**	13.73
FLASSO	0.6	0.4	1.2	3.0	94.8	6.34	6	57.57	28.04
SMUCE	1.6	28.0	66.4	3.8	0.2	-0.27	0	10.99	21.05
cumSeg	75.2	4.0	18.4	2.4	0.0	-1.52	-2	30.15	13.33
PELT	0.6	0.0	92.0	6.4	1.0	0.07	0	8.50	19.43
S3IB	0.6	0.2	**96.8**	2.0	0.4	0.01	0	**6.28**	11.28
SaRa	10.4	36.8	51.2	1.2	0.4	-0.59	0	41.19	39.43
				Model 2					
BootCp	11.0	23.4	**49.0**	9.2	7.4	-0.19	0	**51.12**	44.53
BS	83.2	12.2	4.0	0.6	0.0	-2.69	-3	178.29	68.39
WBS	36.6	27.4	32.4	3.0	0.6	-1.04	-1	79.52	57.56
FLASSO	94.0	0.4	0.6	0.6	4.4	-11.80	-13	181.19	164.71
SMUCE	79.8	18.2	2.0	0.0	0.0	-2.23	-2	116.37	48.17
cumSeg	100.0	0.0	0.0	0.0	0.0	-8.85	-9	181.49	77.19
PELT	38.4	29.6	30.0	1.6	0.4	-1.12	-1	86.50	59.77
S3IB	46.0	28.8	24.2	1.0	0.0	-1.33	-1	96.83	61.50
SaRa	24.6	25.6	**40.4**	7.0	2.4	-0.67	-1	68.14	52.47

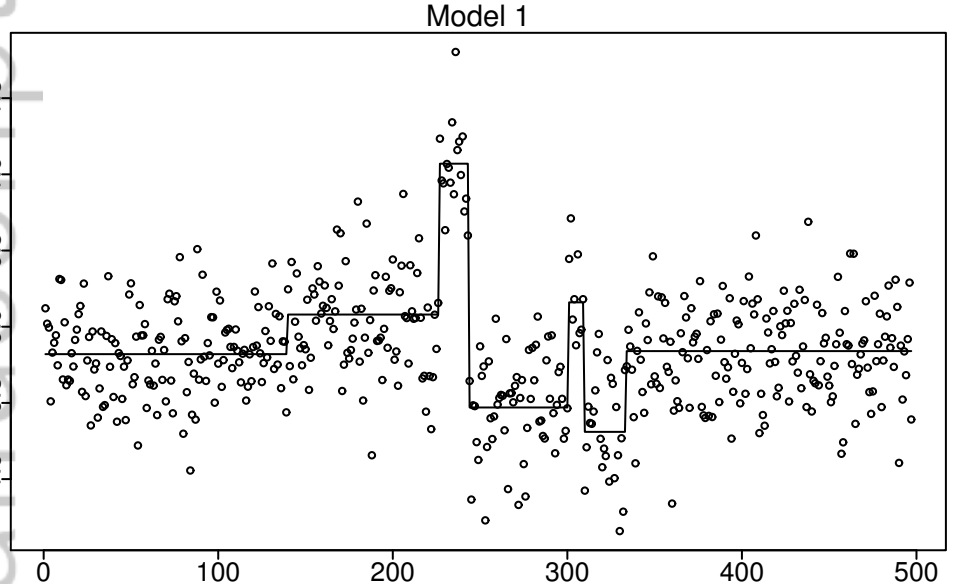
Table 2: Summary statistics of $\hat{N} - N$ and the Hausdorff distance d_H for model 3 and 4.

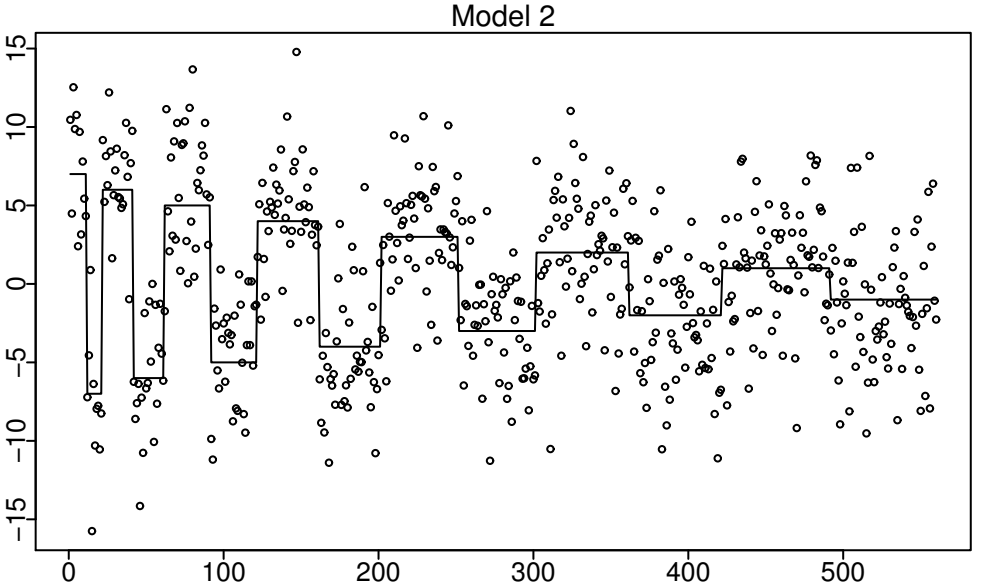
	$\hat{N} - N$							d_H	
	Frequency(%)								
Method	≤ -2	$= -1$	$= 0$	$= 1$	≥ 2	Mean	Med	Mean	SD
				Model 3					
BootCp	6.4	2.8	**87.2**	3.4	0.2	-0.17	0	**3.71**	6.43
BS	100.0	0.0	0.0	0.0	0.0	-12.86	-13	119.74	3.27
WBS	10.4	1.2	72.2	11.8	4.4	-0.49	0	4.91	12.99
FLASSO	100.0	0.0	0.0	0.0	0.0	-12.98	-13	106.50	24.28
SMUCE	96.2	2.6	1.2	0.0	0.0	-6.29	-6	21.27	15.55
cumSeg	100.0	0.0	0.0	0.0	0.0	-12.98	-13	119.90	0.32
PELT	28.2	3.4	62.8	5.2	0.4	-1.11	0	8.8	15.17
S3IB	75.0	2.0	22.8	0.2	0.0	-3.57	-3	24.75	25.49
SaRa	15.8	7.6	74.8	1.6	0.2	-0.46	0	4.34	6.19
				Model 4					
BootCp	0.0	0.8	**94.2**	4.2	0.8	0.05	0	**1.47**	1.45
BS	1.2	10.0	85.4	3.4	0.0	-0.09	0	2.37	2.44
WBS	0.0	0.0	63.0	29.4	7.6	0.48	0	2.12	1.53
FLASSO	11.0	1.2	2.4	7.4	78.0	2.45	4	37.48	6.89
SMUCE	73.4	16.6	10.0	0.0	0.0	-2.46	-2	7.34	2.59
cumSeg	3.2	13.2	75.4	8.2	0.0	-0.13	0	3.43	2.56
PELT	0.0	0.4	**93.6**	5.0	1.0	0.066	0	**1.34**	1.29
S3IB	0.2	6.6	**93.2**	0.0	0.0	-0.07	0	**1.63**	2.09
SaRa	0.2	11.6	84.8	3.4	0.0	-0.09	0	2.54	2.93

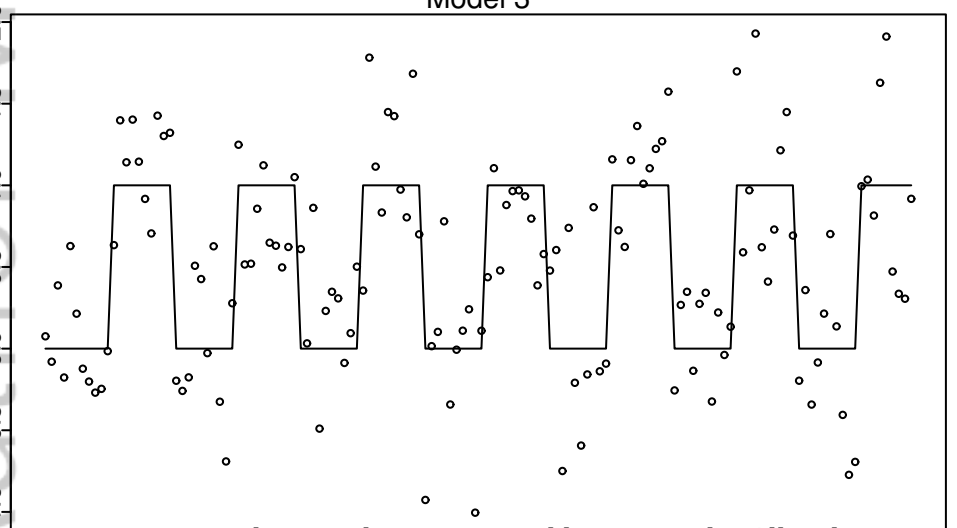
Table 3: The average computational time (in seconds) under model 1-4.

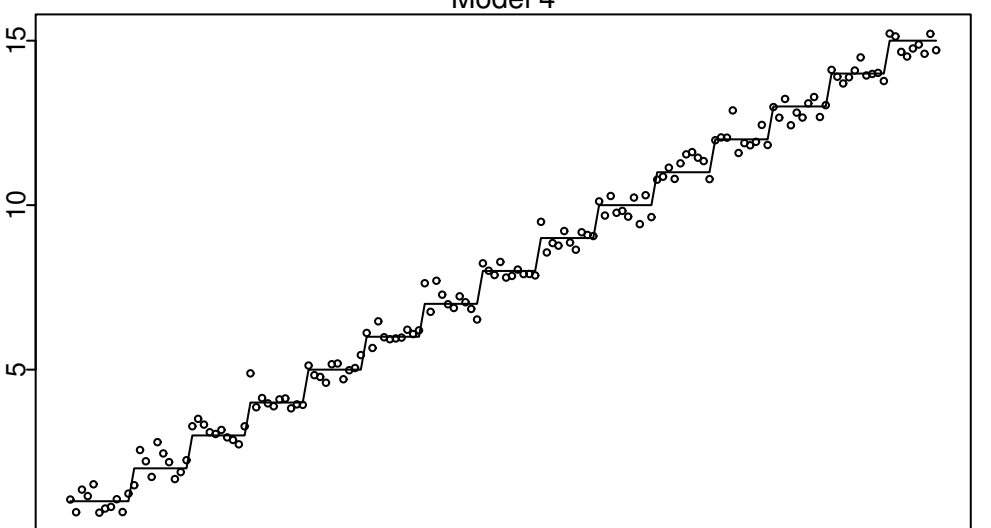
Method	Model 1	Model 2	Model 3	Model 4
BootCp	29.146	38.970	1.218	1.033
BS	0.018	0.023	0.001	0.001
WBS	0.038	0.043	0.003	0.003
FLASSO	5.741	5.913	0.434	0.474
SMUCE	0.528	0.532	0.011	0.013
cumSeg	0.162	0.098	0.007	0.009
PELT	0.107	0.004	0.001	0.001
S3IB	0.174	0.131	0.021	0.028
SaRa	0.182	13.880	0.034	0.050

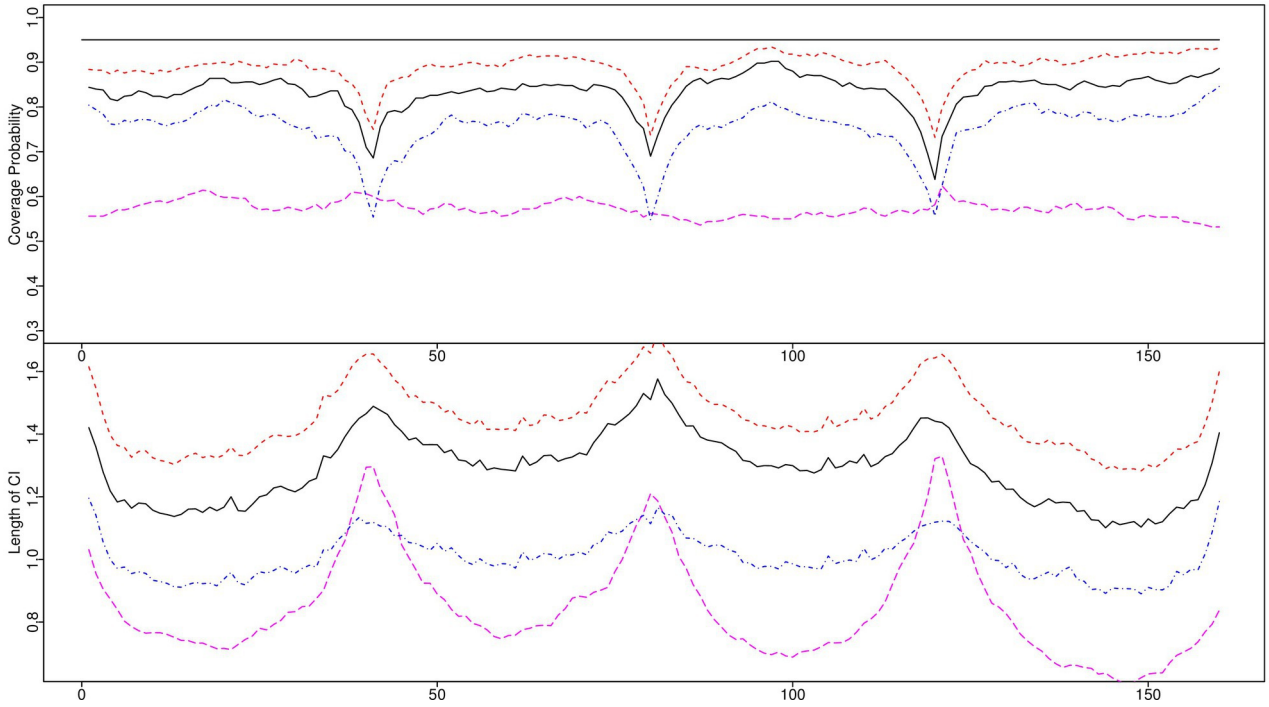




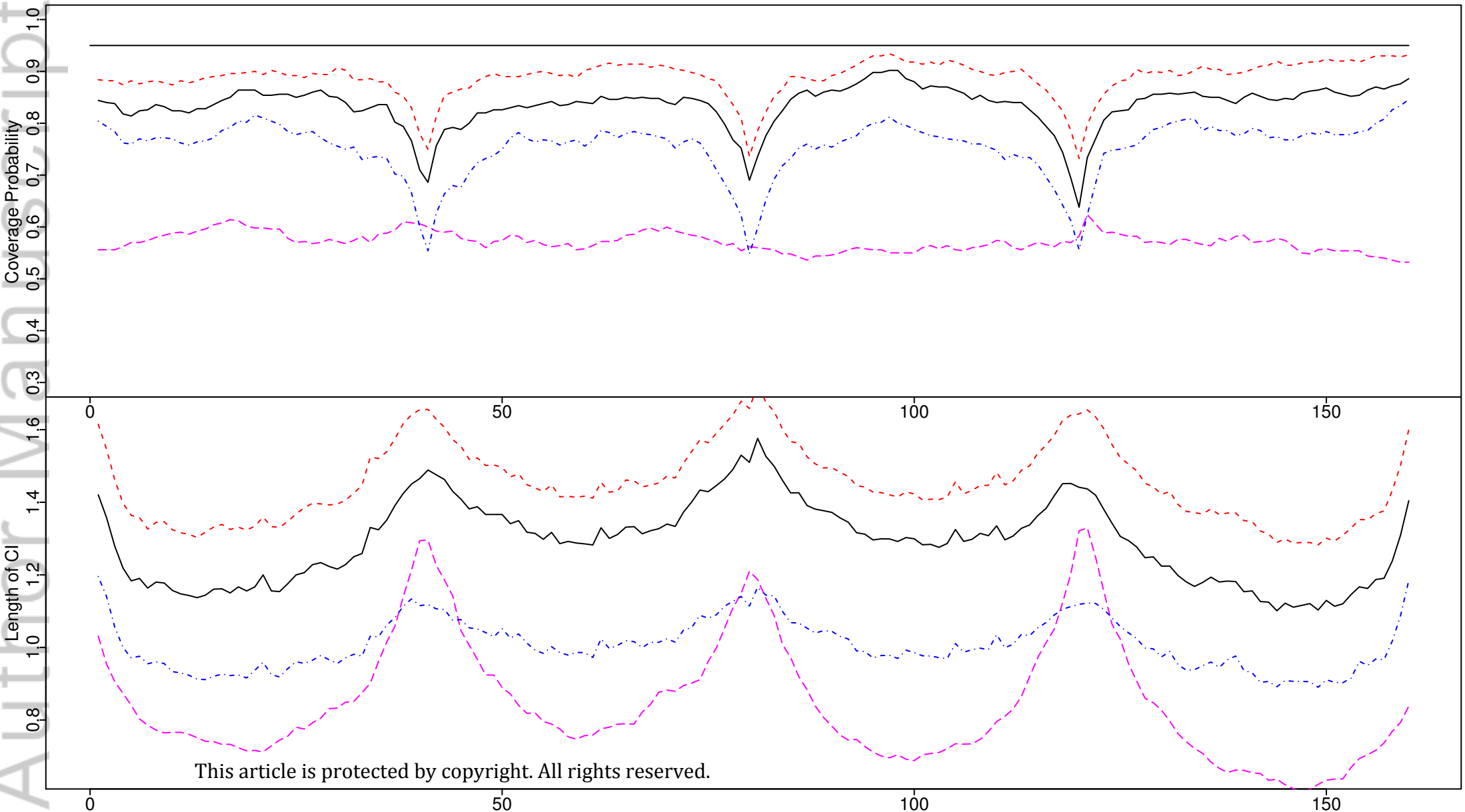


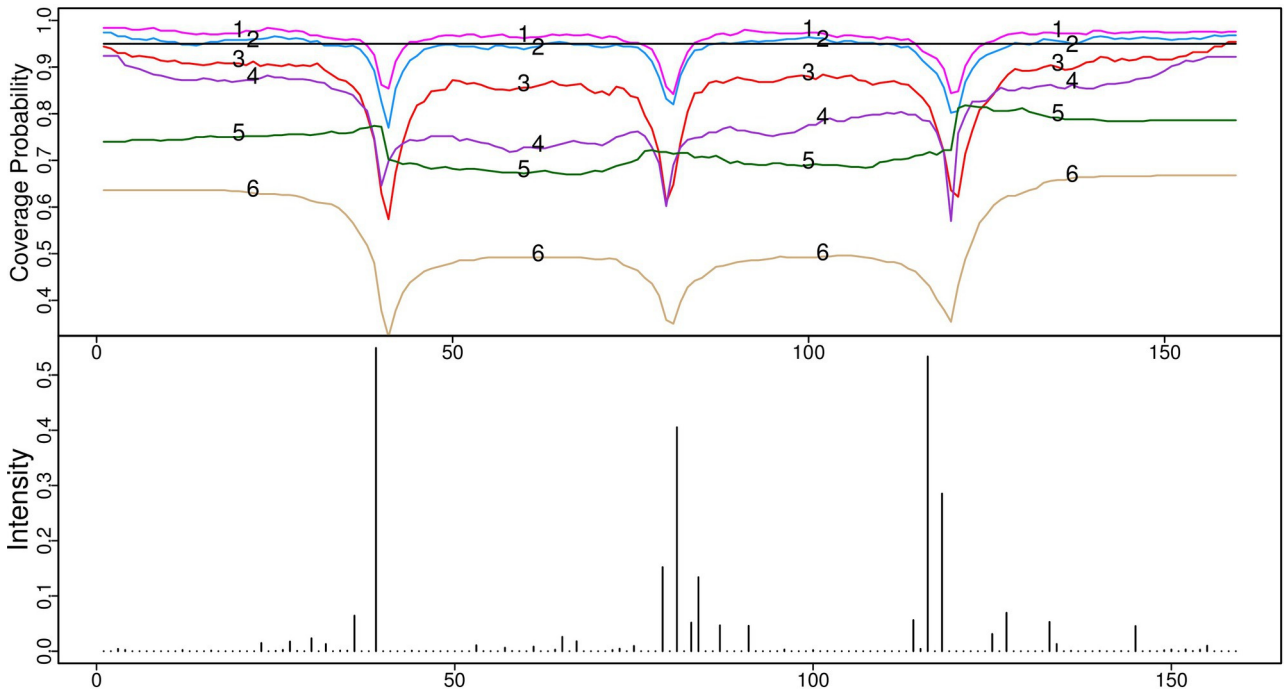




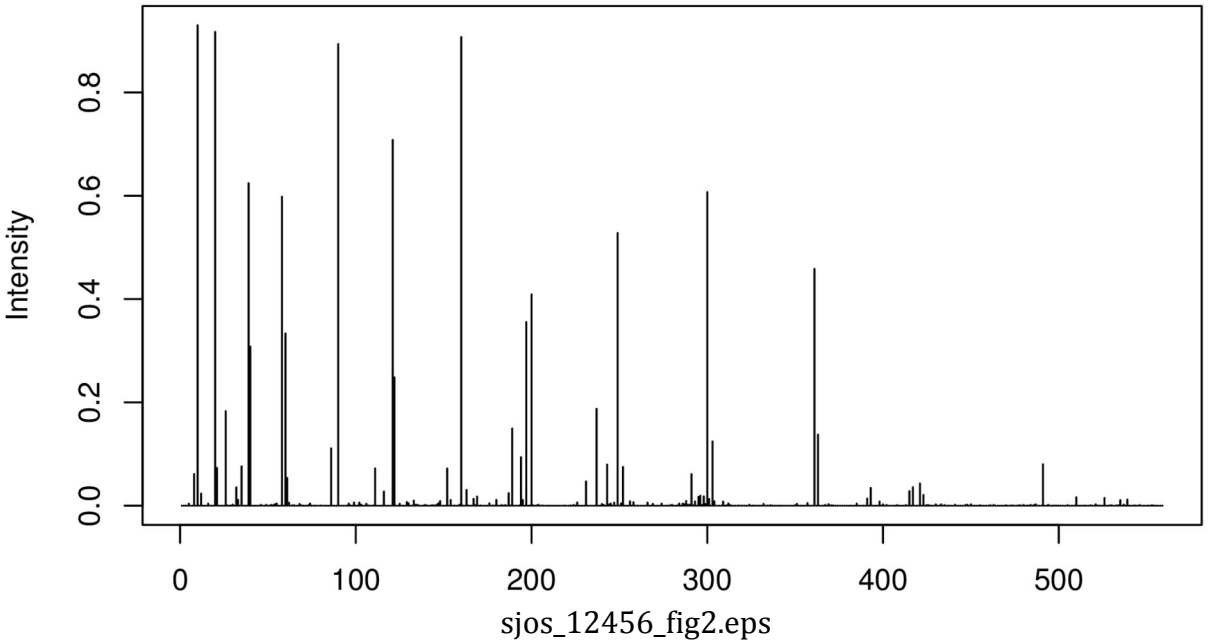


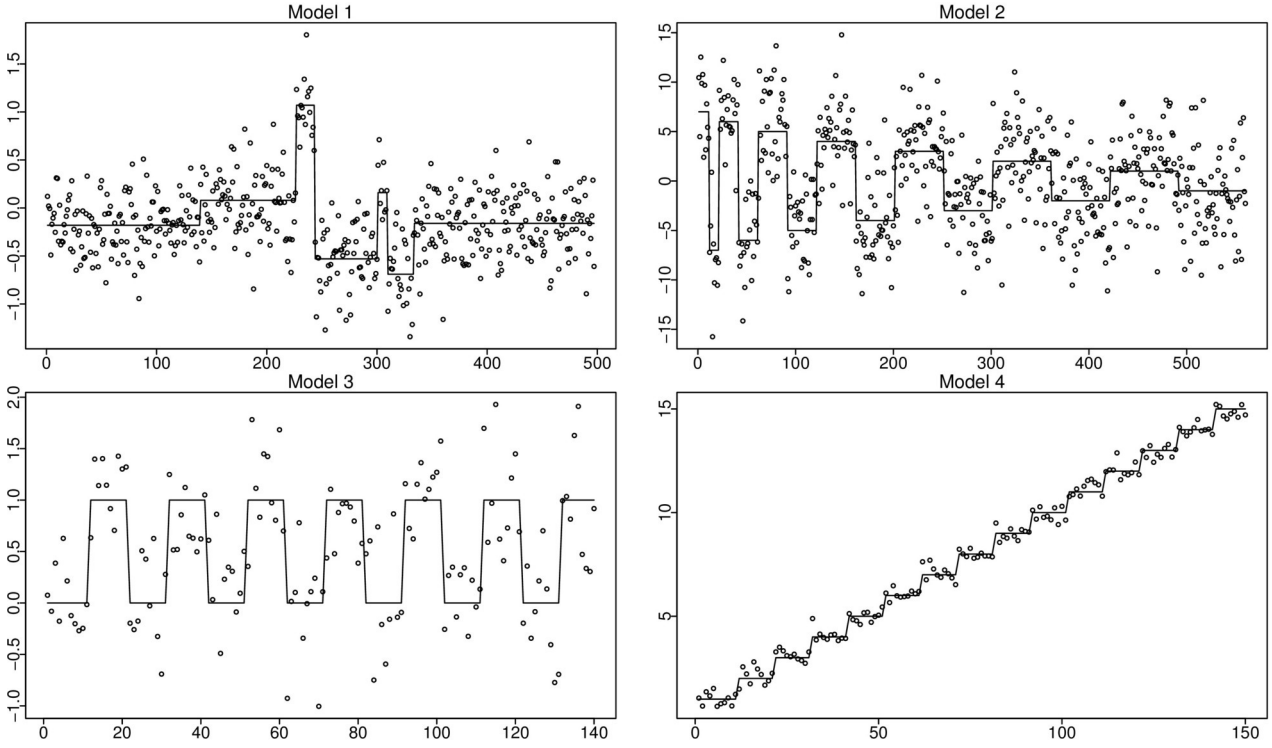
sjos_12456_ex2ar03.eps



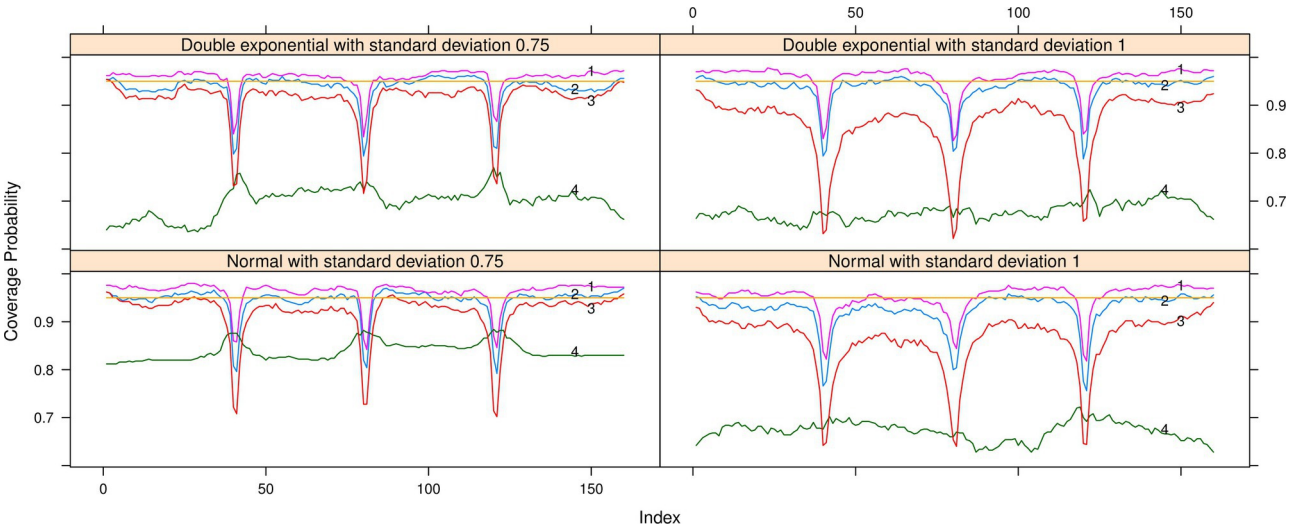


sjos_12456_fig1.eps

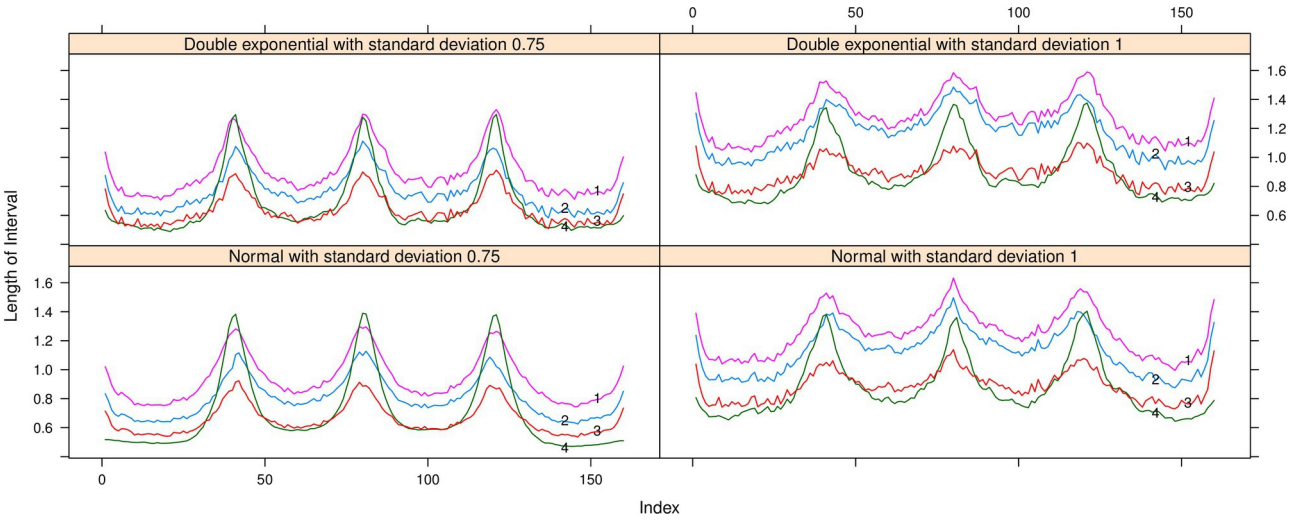




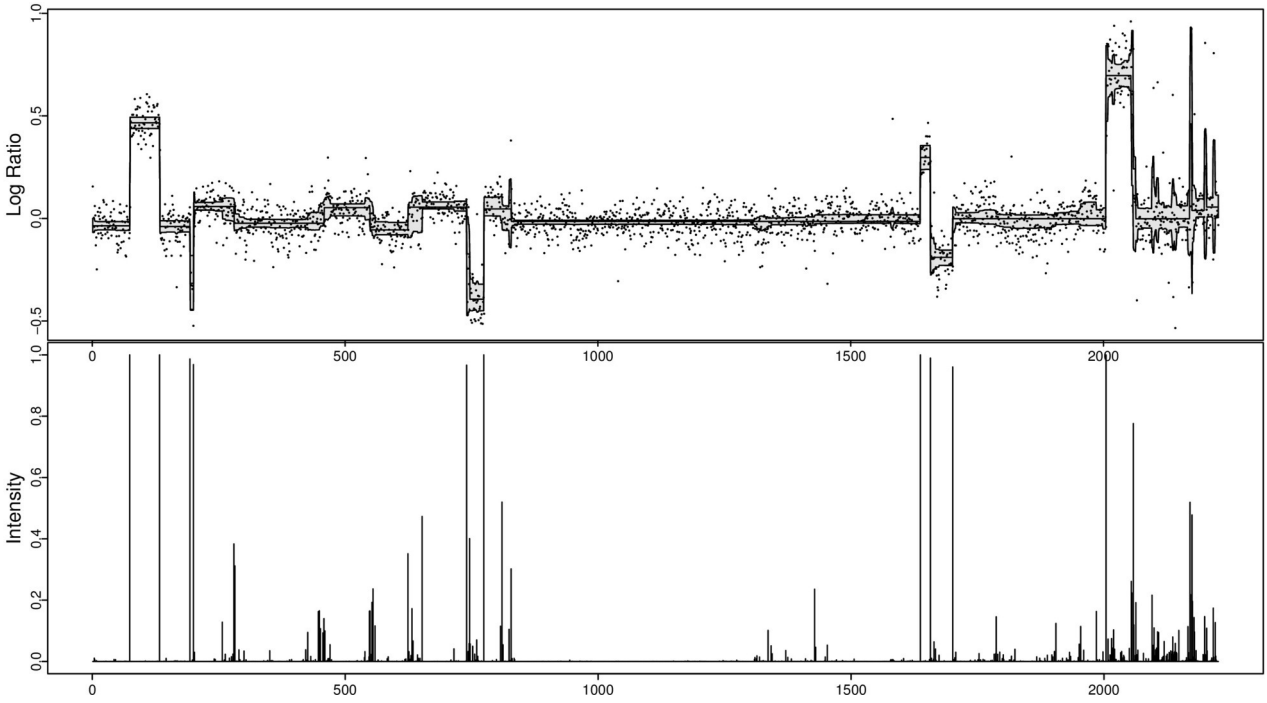
sjos_12456_fig3.eps



sjos_12456_fig4.eps



sjos_12456_fig5.eps



sjos_12456_fig6.eps