Multiscale change point detection
via gradual bandwidth adjustment
in moving sum processes

Tijana Levajković and Michael Messer
Vienna University of Technology

Abstract

A method for the detection of changes in the expectation in univariate sequences is provided. Moving sum processes (MOSUM) are studied. These rely on the selection of a tuning bandwidth. Here, a framework to overcome bandwidth selection is presented – the bandwidth adjusts gradually. For that, MOSUM are made dependent on both time and the bandwidth: the domain becomes a triangle. On the triangle, paths are constructed which systematically lead to change points. An algorithm is provided that estimates change points by subsequent consideration of paths. Strong consistency for the number and location of change points is shown. Simulation studies corroborate estimation precision and reveal competitiveness with state of the art change point detection methods. A companion R-package mscp is made available on CRAN.

Keywords: change point detection, moving sum, multiscale, gradual bandwidth, mscp.

MSC subject classifications: 62G20, 62M99.

1 Introduction

We contribute to the field of change point detection in stochastic sequences. Change point detection applies in various research areas, e.g., climatology (Reeves et al., 2007), speech recognition (Rybach et al., 2009), oceanography (Killick et al., 2010), neuroimaging (Aston and Kirch, 2012), virology (Kass-Hout et al., 2012) etc.

We consider $T$ univariate and independent random variables (RVs) $X_1, \ldots, X_T$, that are piecewise identically distributed, with existing $(2+p)$-th moments $(p > 0)$, without parametric assumptions. Multiple change points in expectation form a set $C$. See Figure 1 (bottom) for an example with $T = 200$, $X_i \sim N(\mu, \sigma^2)$, three change points $C = \{65, 105, 145\}$, and thus four sections with parameters $\mu = 1, 4, 1, -2$ and $\sigma = 1, 0.8, 1, 0.5$, i.e., changes in $\sigma$ may additionally occur when $\mu$ changes.

There is extensive literature that covers changes in expectation, e.g., methods based on likelihood ratios (Fang et al., 2020), empirical processes (Holmes et al., 2013, Horváth and Shao, 2007), $U$-statistics (Döring, 2010), Gombay and Horváth (2002), Horváth and Hušková (2005), least-squares (Harchaoui and Lévy-Leduc, 2010), Lavielle and Moulines (2000) and many more. We mention methodology based on CUSUM-statistics, e.g.,
by [Berkes et al. 2006; Dehling et al. 2017; Hinkley 1971; Page 1954]. For a general overview of change point methods see the textbooks of [Chen and Gupta 2000; Brodsky 2017; Csorgo and Horváth 1997]. In this paper, we aim to tackle multiple change points that may occur on different time scales, as considered e.g., in [Fryzlewicz 2014; Matteson and James 2014; Pein et al. 2017; Spokoiny 2009]. We study MOSUM, see e.g., [Antoch and Hušková 1999; Chu et al. 1995; Hušková and Slaby 2001; Steinebach and Eastwood 1995]. For that we select a window size (bandwidth) $h$ and the upper vertex describes the largest $\hat{c}$ windows, $j \in \{1, \ldots, T/2\}$ and define MOSUM $(D_{t,h})_t$ for index $t = h, \ldots, T-h$: for every time $t$ consider two adjacent windows of size $h$, left $\{t-h+1, \ldots, t\}$ (index $\ell$) and right $\{t+1, \ldots, t+h\}$ (index $r$), and set

$$D_{t,h} := \sqrt{h} \cdot \frac{\hat{\mu}_r - \hat{\mu}_\ell}{(\hat{\sigma}_r^2 + \hat{\sigma}_\ell^2)^{1/2}} \quad (1)$$

where $\hat{\mu}_j$ and $\hat{\sigma}_j^2$ denote the mean and empirical variance of the RVs whose indices lie in the windows, $j \in \{\ell, r\}$. $D_{t,h}$ is Welch’s $t$-statistic for two samples of size $h$. We typically find $|D_{t,h}| \approx 0$ if no change is involved, but $|D_{t,h}| > 0$ if there is a change nearby, $t \approx c \in C$. Thus, change point estimates may be obtained by argmax-estimation, see e.g., [Eichinger and Kirch 2018].

A major challenge lies in the choice of the window size $h$. An $h$ small enough is sensitive to rapid changes when it does not overlap subsequent change points, while a larger $h$ improves detection power of small effects as more RVs are evaluated. But note that $h$ too large may result in overlap of subsequent changes and thus in an estimation bias, or even a failure of detection at all. In order to account for change points that occur on multiple time scales, including rapid changes as well as small effects, methods that combine multiple windows were proposed, see e.g., [Cho and Kirch 2022; Messer 2022]. They work in two steps: first change point candidates are generated for every single $h$, and afterwards all sets are merged giving final estimates. Despite improvements, the methods demand the selection of a window set that best accounts for the location of unknown change points.

The aim of this paper is to provide a MOSUM framework that overcomes window selection, but nevertheless exploits multiple windows to address change point occurrences on multiple time scales, denoted multi-scale change point detection algorithm (MSCP), see Algorithm 4.5. For that we extend the MOSUM perspective: instead of considering $(D_{t,h})_t$ as a process of time $t$ only, we let it depend on both $t$ and $h$, i.e., $(D_{t,h})_{(t,h)}$, while the indices $(t, h)$ lie in a triangle $\Delta_\delta \subset \mathbb{R}^2$, see Figure 1 (top). We define $\Delta_\delta := \{(t, h) \mid t \in T_h, h \in H_\delta\}$, for which we consider first $\delta \in \{1, 2, \ldots, \lfloor T/2 \rfloor\}$ a fixed minimal window, while $\lfloor \cdot \rfloor$ denotes the floor function, second $H_\delta := [\delta, T/2]$ a window interval, and third $T_h := [h, T-h]$ a time interval. Note, $\Delta_\delta$ is a right-angled and isosceles triangle, and the hypotenuse is oriented as the lower edge and refers to the smallest window $h = \delta$. A higher horizontal slice refers to a larger $h$, and the upper vertex describes the largest $h = \lfloor T/2 \rfloor$.

The triangular structure follows from shrinkage of possible $t$-indices $h, \ldots, T-h$ when $h$ increases. In Figure 1 $D_{t,h}$ is color-coded with $D_{t,h} \approx 0$ green, $> 0$ red, and $< 0$ blue. At $c_1 = 65$ there is an increase in $\mu$ and thus $D_{t,h} > 0$, while at $c_2 = 105$ and $c_3 = 145$ a decrease in $\mu$ yields $D_{t,h} < 0$, at least when $h$ is not too large ($h < 40$) such that only a single $c_u$ is overlapped. The upper part of $\Delta_\delta$ refers to larger $h$ that result in an overlap of multiple change points. The area between $c_1$ and $c_2$ is green also for $h$ large, at $(t, h) \approx (85, 70)$, because the
parameters in the first and third section coincide an thus the effects cancel out. In contrast, the area between \( c_2 \) and \( c_3 \) is dark blue for \( h \) large, at \((t,h)\approx(120,70)\), as both changes at \( c_2 \) and \( c_3 \) are negative, which amplifies the effect and as a consequence simple argmax estimation would be flawed. Importantly, note that for smaller \( h \) large values of \(|D_{t,h}|\) concentrate around \( c_u \in C \).

MSCP subsequently acts on subsets of \( \Delta_\delta \) by locally exploiting \( (D_{t,h})_{(t,h)} \). The key ingredient is the construction of a zigzag-path, see Figure 2 (magenta): given a starting value \((t_s,h_s)\) (pink circle), the path leads towards the lower edge of \( \Delta_\delta \) to some point \((\hat{c},\delta)\), and \( \hat{c} \) functions as a change point estimate. The path evolves according to stepwise local argmax-estimation: in each instance the path moves one step downwards, i.e., \( h \) switches to \( h - 1 \) non-randomly. Then the path moves either one step left or right, or it stays, i.e., \( t \) switches to some value in \( \{t - 1, t, t + 1\} \), while the choice falls on the \( t \)-maximizer of \(|D_{t,h-1}|\).

The idea of MSCP is the following, see also Figure 3: given a set \( S \subset \Delta_\delta \) of possible starting values (pink circles), the one maximizing \( h^{-1/2} \cdot |D_{t,h}| \) (i.e., the strongest signal to noise ratio) is the first starting point considered. Then its path delivers the first change point estimate \( \hat{c} \). In order to avoid false positives, a breaking criterion is evaluated, which is computed from \( D_{t,h} \) on the path. If breaking is not demanded, then \( \hat{c} \) is accepted. In order to avoid multiple
detections of the same change point, all elements of $S$, whose paths could lead to $\hat{c}$, are deleted from $S$ (‘cut out cone’). Then estimation restarts, and iteratively change points are detected until the breaking criterion applies. In Figure 3 the algorithm estimates $c_3$, $c_1$ and $c_2$, and then the breaking criterion applies for the next candidate.

Figure 3: Schematic representation of MSCP. It is $C = \{c_1, c_2, c_3\}$. Subsequently $c_3$, $c_1$ and $c_2$ are estimated, and then MSCP breaks.

The main result states strong consistency of MSCP for both the number and location of estimated change points. The crucial technical foundation relies on the path behavior: if the starting point $(t_s, h_s)$ lies high enough on $\Delta_\delta$, i.e., $h_s$ large, then at some point the path enters a subset of $\Delta_\delta$ that we call the area of attraction $A_u$ of a change point $c_u \in C$, see Figure 2 (blue), and from this on it is systematically pushed towards the $c_u$.

For the process in Figure 1 MSCP yields estimates 145, 63 and 105 depicted in Figure 4, i.e., the number is correct, the smallest estimate at 63 is close to $c_1 = 65$ and the other two estimates even hit the true $c_2 = 105$ and $c_3 = 145$. MSCP is made available in the R-package mscp on CRAN [Levajković and Messer (2021)], which includes a summary and plotting routine while the latter created Figure 4: the top panel shows $\Delta_\delta$ including $S$, and the paths constructed with detection orders indicated by integers (green). The middle panel gives the process and the means of the RVs in the detected sections (red), and the bottom shows the empirical variances therein (blue).

We strengthen three important upsides of MSCP: first, the method is non-parametric with weak distributional assumptions and thus allows to analyze a high variety of data. Robustness against additional changes in variance could be helpful in practice, as e.g., an increase in the mean might be accompanied with an increase in volatility. Second, the problem of window selection is overcome. Third, the gradual window adjustment improves the simultaneous detection of change points on multiple time scales and of different signal to noise ratios.

Estimation precision is corroborated by simulation studies investigating different change point scenarios and effect sizes. Various distributions are considered, including normal, gamma, Poisson and binomial, and their combinations. A comparison with state of the art methods reveals strong competitiveness of MSCP.

We mention that MSCP could be generalized generically, i.e., extended for the detection of other quantities than $\mu$, e.g., higher moments, changes in slope, etc. For that, on the one hand
both MOSUM $D_{t,h}$ in [1] and also model formulation needed adjustment, e.g., by considering a difference of empirical higher moments or estimated slopes. On the other hand, the intrinsic structure of MSCP, i.e., the successive construction of paths on $\Delta_{\delta}$ would be maintained.

The paper is organized as follows: In Section 2 we specify the model, $\Delta_{\delta}$ and $D_{t,h}$. In Section 3 we study properties of $(D_{t,h})_{(t,h)\in\Delta_{\delta}}$ which are used for change point detection. In Section 4 we construct the zigzag-path and show that it yields proper estimates. We introduce the set of starting points $S$, define MSCP, and state consistency. In Section 5 we discuss the tuning parameters ($\delta$, $S$ and the breaking criterion), present the simulations, and discuss a data example where MSCP segments a nucleotide sequence of a human genome. Proofs are given in the Appendix 6.

2 The change point model and MOSUM on the triangle

The model $\mathcal{M}$ We fix $T \in \mathbb{N}\{0, 1\}$ and call $\{1, \ldots, T\}$ the observation regime. We consider a subset $C \subset \{1, \ldots, T - 1\}$ of cardinality $|C|$ with ordered elements $c_1 < c_2 \ldots < c_{|C|}$. We call $c_u$ the $u$-th change point and $C$ the set of change points. $C$ is treated fixed but unknown. Set $c_0 := 0$ and $c_{|C|+1} := T$, and the minimal distance of adjacent change points $\delta_{C} := \min_{u=1,\ldots,|C|+1} (c_u - c_{u-1}).$

Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space and $p > 0$. We consider a triangular scheme: let $(Z_{u,i})_{u,i}$ for $u \in \{1, 2, \ldots, |C|+1\}$ and $i = 1, 2, \ldots$ be independent RVs in $\mathcal{L}^{2+p}(\Omega, \mathcal{A}, \mathbb{P})$ with $\mathbb{E}[Z_{u,i}] = 0$ and $\text{Var}(Z_{u,i}) = 1$, and let $(Z_{u,i})_{i=1,2,\ldots}$ be an i.i.d. sequence for each $u$. Further let $\mu_1, \mu_2, \ldots, \mu_{|C|+1} \in \mathbb{R}$ with $\mu_u \neq \mu_{u+1}$, and $\sigma_1^2, \sigma_2^2, \ldots, \sigma_{|C|+1}^2$ positive. For $n \in \{1, 2, \ldots\}$ set

$$X_i := \sum_{u=1}^{|C|+1} (\mu_u + \sigma_u Z_{u,i}) \cdot 1\{nc_{u-1}+1,\ldots,nc_u\}(i), \quad i \in \{1, 2, \ldots, nT\}.$$  

Dependence of $X_i$ on $n$ is suppressed for simplicity. The process $X := (X_i)_{i=1,2,\ldots,nT}$ has $|C|$ change points. In the section from $nc_{u-1} + 1$ to $nc_u$ it has expectation $\mu_u$ and variance $\sigma_u^2$, which are then changing to $\mu_{u+1}$ and $\sigma_{u+1}^2$. The case $n = 1$ is considered the real-time
scenario. Throughout, asymptotics are studied letting $n \to \infty$, i.e., the observation regime and all change points increase linearly. Consequently, the relative change point location $nc_u/nT$ is constant over $n$. The formulation of $n$ in (2) facilitates to track scaling $n^\vartheta$ for different $\vartheta$ in the course of the article. Given $T$, the set of processes $X$ constitutes the model $\mathcal{M}$.

For all $u$, we set the first two moments $\mu_u^{(1)} := \mu_u$ and $\mu_u^{(2)} := \sigma_u^2 + \mu_u^2$ and the centered moments $\mu_u^{(1)} := 0$ and $\mu_u^{(2)} := \sigma_u^2$. If $C = \emptyset$, we set $\mu := \mu^{(1)} := \mu_1^{(1)}$, $\mu^{(2)} := \mu_1^{(2)}$, and $\sigma^2 := \sigma_1^2 := \mu_1^{(2)}$.

Local estimators and MOSUM For $(t, h) \in \Delta_\delta$, see Section 1 we consider indices interpreted a left and right window $I^{(n)}_t := \{\lfloor nt \rfloor - \lfloor nh \rfloor + 1, \ldots, \lfloor nt \rfloor\}$ and $I^{(n)}_h := \{\lfloor nt \rfloor + 1, \ldots, \lfloor nt \rfloor + \lfloor nh \rfloor\}$. We set local estimators for $\mu^{(k)}$ and $\mu^{(k)}$ for $j \in \{t, h\}$ via

$$\hat{\mu}^{(k)}_j := \frac{1}{nh} \sum_{i \in I^{(n)}_j} X_i^k$$

and abbreviate $\hat{\mu}_j := \hat{\mu}^{(1)}_j$ and $\hat{\sigma}_j^2 := \hat{\mu}^{(2)}_j$. Dependence on $t$ and $h$ is suppressed to avoid overload. Then we define

$$D^{(n)}_{t, h} := \frac{\hat{\mu}_t - \hat{\mu}_h}{\langle \hat{\sigma}_t^2 - \hat{\sigma}_h^2 \rangle / \langle nh \rangle^{1/2}},$$

noting that $D^{(1)}_{t, h} = D_{t, h}$ from (1). We set $D^{(n)}_{t, h} := 0$ if the denominator vanishes. For each $n$ the statistics are càdlàg step-functions in both directions $t$ and $h$, with discontinuities in horizontal and vertical slices at values $k/n$, see Figure 5.

![Figure 5: Process construction. The floor-functions (A) in $I^{(n)}_j$, $j \in \{t, h\}$ for both $\lfloor nt \rfloor$ and $\lfloor nh \rfloor$ appear within a factor-$n$-enlarged triangle (B), which is rescaled to $\Delta_\delta$ (C).](image)

Processes are considered in function space $(\mathcal{D}_R[\mathcal{K}], \| \cdot \|_{\infty})$. For a convex subset $\mathcal{K}$ of $\mathbb{R}^k$ with $k \in \{1, 2\}$ let $\mathcal{D}_R[\mathcal{K}]$ denote the set of $\mathbb{R}$-valued functions on $\mathcal{K}$, which in case of $k = 1$ are càdlàg, and in case of $k = 2$ càdlàg with respect to each component when the other component is fixed. $\mathcal{D}_R[\mathcal{K}]$ is endowed with uniform distance $\| \cdot \|_{\infty}$. As asymptotics yield almost surely (a.s.) continuous limits, there is no need to evoke Skorokhod topology.

3 Asymptotics of the estimators

Strong consistency We state strong convergence of the estimators. We define their limits pointwise for $(t, h) \in \Delta_\delta$. They are weightings of the theoretical moments depending on the change point locations relative to the window. Consider the left window $(t - h, t]$ and assume that the change points that are overlapped are $C_t \subset C$, i.e., we find $t - h < c_{t,1} < c_{t,2} < \cdots < c_{t,\lfloor C_t \rfloor} \leq t$, and $c_{t,\varrho}$ denotes the $\varrho$-th smallest change point in $C_t$. Set $c_{t,0} := t - h$ and...
Lemma 3.1. Let \( p > 1 \) set the distance \( d_{\ell,u} := c_{\ell,u} - c_{\ell,u-1} \) between adjacent change points. The moments related to these sections are abbreviated by \( \mu_{\ell,u}^{(k)} \) and \( \sigma_{\ell,u}^{2} \), see Figure 6 left. From this we set

\[
\tilde{\mu}_{\ell}^{(k)} := \sum_{u=1}^{\lvert C_{\ell} \rvert + 1} \frac{d_{\ell,u}}{h} \cdot \mu_{\ell,u}^{(k)} \quad \text{and} \quad \tilde{\sigma}_{\ell}^{2} := \sum_{u=1}^{\lvert C_{\ell} \rvert + 1} \frac{d_{\ell,u}}{h} \cdot [\sigma_{\ell,u}^{2} + (\tilde{\mu}_{\ell} - \mu_{\ell,u})^{2}],
\]

(5)

with \( \tilde{\mu}_{\ell} := \tilde{\mu}_{\ell}^{(1)} \). The limits weight the moments with the time \( d_{\ell,u} \) spent in a section relative to the window length \( h = \sum_{u} d_{\ell,u} \). Further we set

\[
\tilde{\sigma}_{\ell}^{2} := \sum_{u=1}^{\lvert C_{\ell} \rvert + 1} \frac{d_{\ell,u}}{h} \cdot \sigma_{\ell,u}^{2}, \quad \text{and} \quad e_{\ell}^{2} := \sum_{u=1}^{\lvert C_{\ell} \rvert + 1} \frac{d_{\ell,u}}{h} (\tilde{\mu}_{\ell} - \mu_{\ell,u})^{2},
\]

(6)

and note that \( \tilde{\sigma}_{\ell}^{2} = \tilde{\sigma}_{\ell}^{2} + e_{\ell}^{2} \), i.e., \( \tilde{\sigma}_{\ell}^{2} \) weights not only the variances \( \sigma_{\ell,u}^{2} \) as in \( \tilde{\sigma}_{\ell}^{2} \), but also the additional error term \( e_{\ell}^{2} \). The error accounts for the violation of change points in the calculation of the mean in the computation of the empirical variance - the mean is calculated from all data in the windows 'violating' unknown change points. It is \( e_{\ell}^{2} \geq 0 \). If no change points are overlapped \( C_{\ell} = 0 \), then the limits simplify to single theoretical moments \( \tilde{\mu}_{\ell}^{(k)} = \mu_{\ell,1}^{(k)} \) and \( \tilde{\sigma}_{\ell}^{2} = \tilde{\sigma}_{\ell}^{2} = \sigma_{\ell,1}^{2} \) and also \( e_{\ell}^{2} = 0 \). Analogously we define the limits \( \tilde{\mu}_{r}^{(k)} \) and \( \tilde{\sigma}_{r}^{2} \) for the right window \((t, t+h]\) by considering the change points \( C_{r} \) whose ordered elements \( c_{r,u} \) fulfill \( t < c_{r,1} < c_{r,2} < \cdots < c_{r,\lvert C_{r} \rvert} \leq t+h \). If \( C = 0 \), then for all \((t, h) \in \Delta_{\delta} \) we obtain the population parameters \( \mu_{j} = \mu, \tilde{\sigma}_{j}^{2} = \sigma^{2} \) etc. In the following we state strong consistency of the estimators.

Lemma 3.1. Let \( X \in \mathcal{M} \). For \( j \in \{\ell, r\} \) it holds in \((\mathcal{D}_{\mathbb{R}[\Delta_{\delta}]}, \| \cdot \|_{\infty})\) a.s. as \( n \to \infty \)

\[
n^{v} \cdot (\tilde{\mu}_{j}^{(k)}(h) - \mu_{j}^{(k)}(h)) \to (0)_{(t,h)} \quad \text{for} \quad v \in \begin{cases} (0, 1/2), & \text{if} \ k = 1 \\ (0, 1/2) \cap (0, p/(p+2)), & \text{if} \ k = 2, \end{cases}
\]

(8)

and \( n^{v} \cdot (\tilde{\sigma}_{j}^{2} - \sigma_{j}^{2})(h) \to (0)_{(t,h)} \) with \( v \in (0, 1/2) \cap (0, p/(p+2)) \).

The result follows from the Marcinkiewicz-Zygmund SLLN, noting that \( \mathbb{E}[\|X_{j}\|^{2+p}] < \infty \). See \( v \) in Figure 6 right: for \( k = 1 \) we cannot reach \( 1/2 \) noting CLT and LIL, and for \( k = 2 \) the rate improves with the existence of higher moments via \( p \), but only for \( p < 2 \) while \( p/(2+p) < 1/2 \), i.e., \( p \geq 2 \) brings no improvement. Any rate for \( k = 2 \) is valid for \( k = 1 \).

The systematic component of \((D_{t,h}^{(\alpha)}(h)_{(t,h) \in \Delta_{\delta}}\) is the centering \((d_{t,h}^{(\alpha)}(h)_{(t,h) \in \Delta_{\delta}}\) given by

\[
d_{t,h}^{(\alpha)} := \frac{\tilde{\mu}_{\ell} - \mu_{\ell}}{([\tilde{\sigma}_{\ell}^{2} + \sigma_{\ell}^{2}]/[nh])^{1/2}}.
\]

(7)
We note that \( n^{-1/2} \cdot d_{t,h}^{(n)} = d_{t,h}^{(1)} \) and conclude from Lemma 3.1.

**Corollary 3.2.** Let \( X \in \mathcal{M} \). For \( \nu \in (-\infty, 1/2) \cap (-\infty, p/(p+2)] \) it holds in \((D_R[\Delta_\delta], \| \cdot \|)\) a.s. as \( n \to \infty \)

\[
n^\nu \cdot \left( \frac{1}{\sqrt{n}} D_{t,h}^{(n)} - d_{t,h}^{(1)} \right)_{(t,h)} \longrightarrow (0)_{(t,h)}.
\]

**Figure 7** shows \( d_{t,h}^{(1)} \) relating to \( D_{t,h}^{(1)} \). It has an analogous shape, but smoother as noise is canceled out. We will construct the zigzag-path from \( D_{t,h}^{(n)} \) and deduce properties from a non-random path based on \( d_{t,h}^{(n)} \).

**Considering a single change point** We describe \((d_{t,h}^{(n)})_t\) for fixed \( h \) as a function of \( t \). Let \( c_u \in C \). We assume \( h \) small such that only \( c_u \) is overlapped by the double-window when it is near \( c_u \). We formalize the shape of \((d_{t,h}^{(n)})_t\) in Lemma 3.3, which will be used to show that the zigzag-path is pushed towards \( c_u \) when it comes close to it.

If \( c_{u+1} - c_u \geq h \) and \( c_u - c_{u-1} > h \), then set \( \tau_h := [c_{u-1} + h, c_{u+1} - h] \), else set \( \tau_h := \emptyset \). For \( h \) small it is \( \tau_h \neq \emptyset \). In this case \((d_{t,h}^{(n)})_t \in \tau_h \) is continuous, it vanishes if \( |t - c_u| \geq h \) because no \( c \in C \) is overlapped, and it systematically deviates from zero if \( |t - c_u| < h \) with extreme value \( t = c_u \), see Figure 5 in which the blue and red lines indicate a uniform bound for the derivative. We set \( v_{t,h} := [(\tilde{\sigma}_r^2 + \tilde{\sigma}_l^2)/(\tilde{\sigma}_r^2 + \tilde{\sigma}_l^2)]^{1/2} \), recalling \( \tilde{\sigma}_j^2 = \sigma_j^2 + c_j^2 \), see [5] and [6], and denote \( \land \) the minimum and \( \lor \) the maximum.

**Lemma 3.3.** Let \( X \in \mathcal{M}, c_u \in C \) and \( h \) be fixed such that \( \tau_h \neq \emptyset \). Then both \((d_{t,h}^{(n)})_t \in \tau_h \) and

\[(v_{t,h} \cdot d_{t,h}^{(n)})_t \in \tau_h\]

are continuous and zero for \( t \in \tau_h \setminus [c_u - h, c_u + h] \). For \( t \in [c_u - h, c_u + h] \), four cases are differentiated regarding \( d_{t,h}^{(n)} \): for \( \mu_{u+1} > \mu_u \) we find that

\[
d_{t,h}^{(n)} \text{ is } \begin{cases} \text{strictly increasing for } t \in [c_u - h, c_u], \\ \text{strictly decreasing for } t \in (c_u, c_u + h]. \end{cases}
\]

For \( \mu_{u+1} < \mu_u \) it is 'increasing' and 'decreasing' replaced. Further, for \( t \in (c_u - h, c_u) \cup \)
\((c_u, c_u + h)\) it is

\[
\sqrt{n_h} \cdot \kappa_u \leq \left| \frac{\partial}{\partial t} d_{t,h}^{(n)} \right| \leq \sqrt{n_h} \cdot \kappa_b,
\]

with constants

\[
\kappa_u := |\mu_{u+1} - \mu_u| \cdot \frac{2(\sigma_{u+1}^2 \lor \sigma_u^2)}{[2(\sigma_{u+1}^2 \lor \sigma_u^2) + (\mu_{u+1} - \mu_u)^2/4]^{3/2}},
\]

\[
\kappa_b := |\mu_{u+1} - \mu_u| \cdot \frac{2(\sigma_{u+1}^2 \lor \sigma_u^2) + (\mu_{u+1} - \mu_u)^2}{[2(\sigma_{u+1}^2 \lor \sigma_u^2)]^{3/2}}.
\]

Regarding \(v_{t,h} \cdot d_{t,h}^{(n)}\), six cases are differentiated: for \(\mu_{u+1} > \mu_u\) and \(\mu_{u+1} < \mu_u\), the expressions hold true, but with ‘convex’ and ‘concave’ as well as ‘increasing’ and ‘decreasing’ replaced. It is \(v_{t,h} \cdot |d_{t,h}^{(n)}| \geq |d_{t,h}^{(n)}|\) with equality at \(t = c_u\).

Both \((v_{t,h} \cdot d_{t,h}^{(n)})_t\) and \((d_{t,h}^{(n)})_t\) are depicted in Figure 8D and E.

Figure 8: Construction of \((d_{t,h}^{(n)})_{t\in\mathbb{T}_h}\) (for \(c = c_1\)). A: \(\tilde{\mu}_t = \tilde{\mu}_t\), B: \(\tilde{\sigma}_t^2 + \tilde{\gamma}_t^2\), C: \(e_t^2 + e_t^2\), D: \(v_{t,h} \cdot d_{t,h}^{(n)}\), E: \(d_{t,h}^{(n)}\) and lines with slopes \(\mp(nh)^{1/2} \cdot \kappa_u\) (blue) and \(\mp(nh)^{1/2} \cdot \kappa_b\) (red).

Figure 8 shows the statistics that factor into \((d_{t,h}^{(n)})_{t\in\mathbb{T}_h}\). It is \(\mu_2 > \mu_1\) and \(\sigma_2^2 > \mu_2^2\). The function \(\tilde{\mu}_t - \tilde{\mu}_t\) has the shape of a hat (A). It is positive as \(\mu_2 > \mu_1\). The function \(\tilde{\sigma}_t^2 + \tilde{\gamma}_t^2\) is linear (B). It is increasing as \(\sigma_2^2 > \sigma_1^2\). The error \(e_t^2 + e_t^2\) describes two parabolas and vanishes outside the \(h\)-neighborhood of \(c := c_u\) and also at \(c\) (C). The error is maximal at \(c \mp h/2\) which are the points where either the right or the left window is divided in half and thus equally sharing the left and right population. Panel D shows \(v_{t,h} \cdot d_{t,h}^{(n)}\) which is \((nh)^{1/2}\) times (A) divided by the square-root of (B). It is non-negative as \(\mu_2 > \mu_1\) and concave on \([c - h, c]\) and convex on \((c, c + h]\) as \(\sigma_2^2 > \sigma_1^2\). Panel D shows \(d_{t,h}^{(n)}\) which is \((nh)^{1/2}\) times (A) divided by the square-root of the sum of (B) and (C). As \(\mu_2 > \mu_1\), it is strictly increasing in \([c - h, c]\) and strictly decreasing on \((c, c + h]\). It is \(v_{t,h} \cdot d_{t,h}^{(n)} \geq d_{t,h}^{(n)}\), as \(v_{t,h} \geq 1\). Note that in Lemma 3.3 it is 0 < \(\kappa_u < \kappa_b < \infty\), and \(\kappa_u, \kappa_b\) depend only on the population parameters, such that the bounds
for the derivative in \( \frac{\partial f}{\partial y} \) depend on \( y \) but not on \( t \). The upper bound \((nh)^{1/2} \cdot \kappa_b\) is a Lipschitz constant for \((a^{(n)}_{t,h})_t\). Figure 8E shows lines with slopes \((nh)^{1/2} \cdot \kappa_a\) (blue) and \((nh)^{1/2} \cdot \kappa_b\) (red), i.e., the slope of \((a^{(n)}_{t,h})_t\) varies in between for all \( t \in (c_u - h, c_u) \cup (c_u, c_u + h) \).

**Weak convergence** We state weak convergence of \((D^{(n)}_{t,h})_{(t,h) \in \Delta^\delta}\) if \( C = \emptyset \). Let \((W_t)_{t \geq 0}\) be a standard Brownian motion. Then define \((L_{t,h})_{(t,h) \in \Delta^\delta}\) via \(L_{t,h} := (2h)^{-1/2} \cdot [(W_{t+h} - W_t) - (W_t - W_{t-h})]\). The process \((L_{t,h})_{(t,h)}\) is continuous with \(L_{t,h} \sim N(0,1)\).

**Proposition 3.4.** Let \( X \in \mathcal{M} \) with \( C = \emptyset \). In \((D_{\mathbb{R}[\Delta^\delta],\| \cdot \|_\infty})\) it holds as \( n \to \infty\) that \((D^{(n)}_{t,h})_{(t,h)} \xrightarrow{d} (L_{t,h})_{(t,h)}\).

\( L_{t,h} \) preserves the double-window structure of \( D^{(n)}_{t,h}\), zero mean aligns with \( C = \emptyset \) and unit variance results from scaling of \( D^{(n)}_{t,h}\). The proof applies Donsker’s theorem.

We mention that the upper limit theorems could be extended to dependent data structures, for which in \( D^{(n)}_{t,h}\) from (4) the variance estimator first needed to address that dependence. As a consequence, MSCP formulated in the following section would apply to those scenarios as well. For recent developments of change point detection methods w.r.t. dependent data see e.g., (Baranowski et al., 2019a; Dette et al., 2020), supplement (Fryzlewicz, 2018a) of (Fryzlewicz, 2018b), or the overview article (Aue and Horváth, 2013).

### 4 Change point detection via MSCP

**Segmentation of the triangle** We specify regions of \( \Delta^\delta\). Each \( c_u \in C \) has an area of attraction \( A_u := \{(t,h) \in \Delta^\delta : c_u - 1 \leq t - h < c_u \leq t + h < c_u + 1\} \), see Figure 2 (blue).

For all \((t,h) \in A_u\) the associated double window overlaps \( c_u \) but no other change point. It is \(|d^{(n)}_{t,h}| > 0\) for \((t,h) \in A_u\). We show that path that starts in \( A_u\) is systematically pushed towards \( c_u\), such that its endpoint becomes a proper estimate for \( c_u\). We also set \( B_u := \{(t,h) \in A_u : |t - c_u| \leq h - \delta + 1\} \), for whose elements we find the ‘\( t\)-distance’ to \( c_u\) smaller than the ‘\( h\)-distance’ (plus 1) to the bottom of \( \Delta^\delta\). For any \( t_0 \in [\delta, T - \delta] \) we define its cone \( K_{t_0} := \{(t,h) \in \Delta^\delta : t - h < t_0 \leq t + h\} \). The cones \( K_{c_u}\) of the \( c_u \in C\) are shown in Figure 2 (orange). \( K_{c_u}\) consists of all \((t,h) \in \Delta^\delta\) for which the double window overlaps \( c_u\) and possibly neighboring change points. It holds \( B_u \subset A_u \subset K_{c_u} \subset \Delta^\delta\). Finally, we consider the remainder \( R := \Delta^\delta \setminus \bigcup_{u=1}^{C} K_{c_u}\), that consists of \((t,h) \in \Delta^\delta\) for which no \( c_u\) is overlapped by the double window, such that \(d^{(n)}_{t,h} = 0\) for all \((t,h) \in R\). In Figure 2 \( R\) consists of four subtriangles of \( \Delta^\delta\) as there are three change points involved.

**The zigzag-downpath** We construct a path on \( \Delta^\delta \cap \mathbb{N}^2 \) w.r.t. either \( D^{(n)}_{t,h}\) or \( d^{(n)}_{t,h}\).

**Algorithm 4.1. (zigzag-downpath)**

**Input:** \( X \in \mathcal{M} \) and a starting value \((t_s, h_s) \in \Delta^\delta \cap \mathbb{N}^2\). Set \( f^{(n)}_{t,h} \in \{d^{(n)}_{t,h}, D^{(n)}_{t,h}\}\).

**Construction:** set \( t_s^{(n)}(0) := \min (\arg \max_{t \in \{t_{s-1}, t_s, t_{s+1}\} \cap \Delta^\delta} n^{-1/2} \cdot |f^{(n)}_{t,h}|-) \). For \( k = 1, 2, \ldots, h_s - \delta \) iteratively define

\[
t_s^{(n)}(k) := \min \left( \arg \max_{t \in \{t_s^{(n)}(k-1), t_s^{(n)}(k-1), t_s^{(n)}(k-1)+1\}} n^{-1/2} \cdot |f^{(n)}_{t,h,k-1}|- \right) .
\]
Let Lemma 4.2. of ensures uniqueness. For \( f \) and accordingly the paths we write restriction to \( \Delta_\delta \) by unity if \( |s-c_u| \leq k \), \( t_s(k) - t_s(k-1) = \begin{cases} 0, & \text{if } |t_s-c_u| \leq k, \\ 1, & \text{if } t_s < c_u - k, \\ -1, & \text{if } t_s > c_u + k, \end{cases} \) (11) and accordingly \( t_e \) and \( \hat{t}_e^{(n)} \), noting that for \( f_{t,h}^{(n)} = D_{t,h}^{(n)} \) the path is independent of \( n \), as \( n^{-1/2} \cdot |d_{t,h}^{(n)}| = |d_{t,h}^{(1)}| \).

Path behavior w.r.t \( d_{t,h}^{(n)} \) A path starting in \( A_u \) systematically tends towards \( c_u \):

**Lemma 4.2.** Let \( X \in M \), \( c_u \in C \) and \( f_{t,h}^{(n)} = d_{t,h}^{(n)} \). If \( (t_s, h_s) \in A_u \), then for all \( k \in \{1, 2, \ldots, h_s - \delta\} \) it holds

\[
|s-c_u| \leq k, \\
t_s < c_u - k, \\
t_s > c_u + k,
\]

If \( (t_s, h_s) \in B_u \), then \( t_e = c_u \). If \( (t_s, h_s) \in A_u \backslash B_u \), then \( |t_e - c_u| \leq \delta - 1 \).

**Lemma 4.2** states that in each step \( t_s(k) \) either increases by unity if \( t_s < c_u \) or it decreases by unity if \( t_s > c_u \), and if it reaches \( c_u \) then it stays. Thus, the path \( (t_s(k), h_s - k) \) is perfectly zigzagging towards the vertical line at \( c_u \). If it reaches this line then it moves vertically downwards to the lower edge of \( \Delta_\delta \), i.e., \( t_e = c_u \). The proof exploits the shape of \( (d_{t,h}^{(n)})_t \) stated in Lemma 3.3.

Path behavior w.r.t \( D_{t,h}^{(n)} \) Regarding \( D_{t,h}^{(n)} \), a path starting in \( A_u \) converges towards the path w.r.t. \( d_{t,h}^{(n)} \):

**Proposition 4.3.** Let \( X \in M \), \( c_u \in C \) and \( f_{t,h}^{(n)} = D_{t,h}^{(n)} \). Let \( (t_s, h_s) \in A_u \), then for all \( k \in \{0, 1, 2, \ldots, h_s - \delta\} \) it holds that \( \hat{t}_s^{(n)}(k) \rightarrow t_s(k) \) a.s. as \( n \rightarrow \infty \) for all \( k \), while \( t_s(k) \) is given in (12).

Consequently, starting in \( A_u \) yields detection of \( c_u \) up to a distance of \( \delta - 1 \). More precisely, for \( (t_s, h_s) \in B_u \) we find \( \lim_{n \rightarrow \infty} \hat{t}_e^{(n)} = c_u \) a.s., and for \( (t_s, h_s) \in A_u \backslash B_u \) it is \( \lim_{n \rightarrow \infty} |\hat{t}_e^{(n)} - c_u| \leq \delta - 1 \) a.s. The proof exploits proximity of \( D_{t,h}^{(n)} \) and \( d_{t,h}^{(n)} \).

Starting points We call a set \( S \subset \mathbb{N}^2 \cap \Delta_\delta \) a sufficient set of starting points if \( S \cap A_u \neq \emptyset \) for all \( u = 1, \ldots, |C| \), i.e., for all \( c_u \in C \) there lies a starting point in \( A_u \). For a mesh size \( g \in \mathbb{N} \) set a grid \( S_g := \{g \cdot \mathbb{N}^2 \cap \Delta_\delta \}, \) see Figure 3A (pink dots).

**Lemma 4.4.** Let \( X \in M \) and \( \delta < [\delta_C/2] \). A grid \( S_g \) is sufficient if \( g \leq [\delta_C/2] \).
The proof applies geometry. When a path enters $A_u$, then $c_u$ is detected up to a distance of $\delta - 1$, see Proposition 4.3. Thus, when considering all paths starting in $S$, then all $c \in C$ will be detected. We need to avoid multiple detections of any $c$ as well as false positives.

**Change point detection** Set $\min \emptyset := \infty$, and for a set $F$ denote $U(F)$ a uniformly sampled element.

**Algorithm 4.5. (MSCP)**

*Input:* $X \in \mathcal{M}$, a set of starting values $S$,

and constants $\kappa > 0$, and $\beta \in [1/2 - v, 1/2]$ for $v \in (0, 1/2) \cap (0, p/(p + 2)]$.

*Initialize:* counter $m = 1$, estimates $\hat{C}^{(n)} = \emptyset$ and starting values $S^{(n)} = S$.

*While $S \neq \emptyset$, loop:*

1. Choose the starting value $(\hat{t}^{(n)}_{s,m}, \hat{h}^{(n)}_{s,m}) := U(\arg\max_{(t,h) \subset S^{(n)}} ((nh)^{-1/2} \cdot |D^{(n)}_{t,h}|)).$

2. Run the zigzag-path w.r.t. $(\hat{t}^{(n)}_{s,m}, \hat{h}^{(n)}_{s,m})$ and $f^{(n)}_{t,h} = D^{(n)}_{t,h}$,

   call the path $(\hat{t}^{(n)}_{s,m}(k), \hat{h}^{(n)}_{s,m} - k)_{k=0,1,2,\ldots,\hat{h}^{(n)}_{s,m} - \delta}$ and the endpoint $(\hat{t}^{(n)}_{e,m}, \delta)$,

   and compute the minimal distance $\mathcal{D}^{(n)} := \min_{m} [\hat{t}^{(n)}_{e,m} - \hat{c}^{(n)}] : \hat{c}^{(n)} \in \hat{C}^{(n)}$.

3. Check for multiple detections (3a.) and false positives (3b.)

   3a. If $\mathcal{D}^{(n)}_{m} \leq 2(\delta - 1)$, then set $S^{(n)} = S^{(n)} \setminus K^{(n)}_{\hat{t}^{(n)}_{e,m}}$ and Goto 1.

   3b. If $\max_{k=0,1,\ldots,\hat{h}^{(n)}_{s,m} - \delta} |D^{(n)}_{\hat{t}^{(n)}_{e,m}(k), \hat{h}^{(n)}_{s,m} - k}| < \kappa \cdot n^\beta$, then break the loop.

4. Update $m = m + 1$, then set $\hat{C}^{(n)} = \hat{C}^{(n)}_{m - 1} \cup \{\hat{t}^{(n)}_{e,m - 1}\}$ and $S^{(n)} = S^{(n)}_{m - 1} \setminus K^{(n)}_{\hat{t}^{(n)}_{e,m - 1}}$, and Goto 1.

**Output:** $\hat{C}^{(n)} := \hat{C}^{(n)}_{m}$.

The parameters $\kappa$ and $\beta$ yield a threshold $\kappa \cdot n^\beta$ (in 3b.), and if chosen large then this supports breaking the algorithm. Recall $E[|X_1|^{2+p}] < \infty$. $\beta$ is related to $p$ via $v$, see Figure 6 right. Small $p$ forces $\beta$ close to $1/2$ and a larger $p$ allows for smaller $\beta$. The parameter $\beta$ is redundant if $n = 1$. In 1., the maximum is a.s. unique if the RVs of $X$ do not have point masses. In general, uniform sampling yields well-definedness. It is $\mathcal{D}^{(n)}_{1} = \infty$.

Algorithm 4.5 (MSCP) is depicted in Figure 3. Successively, $C$ is estimated. We comment on the steps.

1.: Among all current starting points the one maximizing $(nh)^{-1/2} \cdot |D^{(n)}_{t,h}|$ is chosen, uniformly in case of non-uniqueness. This ensures the starting point to lie outside the remainder $R$, a.s. for $n$ large, in case of remaining undetected change points.

2.: The zigzag-path is constructed and $\hat{t}^{(n)}_{e,m}$ functions as a change point candidate. The minimal distance $\mathcal{D}^{(n)}_{m}$ of $\hat{t}^{(n)}_{e,m}$ and all previous estimates is computed.

3.: Two aspects are addressed:

   3a.: ’Avoiding multiple detections’: if $\mathcal{D}^{(n)}_{m} \leq 2(\delta - 1)$, then $\hat{t}^{(n)}_{e,m}$ is rejected, all starting points from its cone are cut, and estimation restarts. The rational is that if the paths of both $\hat{t}^{(n)}_{e,m}$ and the closest estimate passed $A_u$, then they also systematically led to $c_u$ up to $\delta - 1$. Hence, rejection of such $\hat{t}^{(n)}_{e,m}$ avoids multiple detections.

12
Theorem 4.6. Let $\hat{c}_{\ell,m}^{(n)}$ be redundant for theoretical consistency, but may be used in practice.

Note that we allow an error up to $\delta - 1$, and if the distance of $\hat{c}_{\ell,m}^{(n)}$ to $c_u$ is positive, then a slightly shifted cone was cut. This gives rise for 3a. at first. Also note that if all $\epsilon \in C$ are detected and near cones are cut (again mentioning 3a.), then what is eventually left over is the remainder $R$. Thus, on the next path the objection function will no longer take extreme values which support breaking in 3b., and the previous estimates are returned. Finally note that the breaking criterion 3c. applies if the minimal distance between change points was exceeded, also accounting for estimation precision only up to $\delta - 1$. Asymptotically, this will not occur, so 3c. is obsolete for theoretical consistency, but may be used in practice.

We denote $\hat{c}_{u}^{(n)}$ the $u$-th smallest element of $\hat{C}^{(n)}$. For well-definedness, if $|\hat{C}^{(n)}| < |C|$ set $\hat{c}_{u}^{(n)} := 0$ for all $u = |\hat{C}^{(n)}| + 1, \ldots, |C|$. The algorithm succeeds:

**Theorem 4.6.** Let $|\delta_C/2| > \delta$. In Algorithm 4.5 let the input $S$ be sufficient. Then for the output $\hat{C}^{(n)}$ it holds a.s. as $n \to \infty$ that

$$|\hat{C}^{(n)}| \to |C| \quad \text{and} \quad \limsup_{n \to \infty} |\hat{c}_{u}^{(n)} - c_u| \leq \delta - 1 \quad \text{for all} \ u = 1, \ldots, |C|.$$

Location estimation is correct up to $\delta - 1$. The choice $\delta = 1$ states strong consistency. Recall that $|\delta_C/2| > \delta$ is needed for sufficiency of $S = S_g$ in Lemma 4.4. Also it implies $\delta_C > 2(\delta - 1)$, i.e., neighboring change points should be separated by more than the worst errors of their estimators. The proof applies Proposition 4.3 for the estimation of $C$, and Corollary 3.2 for correct breaking.

5 Practical aspects

**Parameter choice** We give recommendations for the choice of $S$, $\kappa$ and $\delta$ in practice, where $n = 1$. Then, the threshold is $\kappa \cdot n^\beta = \kappa$, i.e., $\beta$ is redundant.

1. **Choice of $\kappa$:** $\kappa$ can be chosen as the rejection threshold of a test for $H_0 : C = \emptyset$: For $H_0$ Proposition 3.4 implies weak convergence of $\sup_{t,h \in \Delta} |D_{t,h}^{(n)}| \to \sup_{t,h \in \Delta} |L_{t,h}|$ as $n \to \infty$, and we choose $\kappa$ as the $(1 - \alpha)$-quantile of the limit distribution (e.g., significance level $\alpha = 0.01$), which can be derived in simulations, compare e.g., Messer et al. (2014). If $|\hat{C}^{(1)}| > 0$, then $H_0$ is rejected at level $\leq \alpha$, as the maximum w.r.t. a path is bounded by the supreme over $\Delta_g$.

2. **Choice of $S$:** Lemma 4.4 states sufficiency of $S = S_g$ if $g \leq |\delta_C/2|$ (in case $\delta < |\delta_C/2|$). Thus, if $\delta_C$ is known it is reasonable to choose $g = |\delta_C/2|$ as $g$ large reduces computational complexity. If $\delta_C$ is unknown, we set $g = \delta$ or even smaller if computational complexity allows for.

3. **Choice of $\delta$:** We recommend $\delta = 20$: by construction, asymptotic considerations imply an increase of the double windows. For $n \to \infty$, weakly $D_{t,h}^{(n)} \to L_{t,h} \sim N(0,1)$ for all
MSCP TL and MM

$(t, h) \in \Delta_\delta$, but for $n = 1$ small $h$ means less observations and the approximation via $N(0, 1)$ is harder to justify. Practically, for tiny $h$ there is high variability in $\mu_j$ and $\sigma_j^2$ and thus $D_{t,h}^{(1)}$ is likely to take extreme values even if no change is involved, resulting in false positives. Thus, $\delta$ should be bounded from below. Figure 9 shows simulations for the probability that $\mathbb{P}(|\hat{C}_1^{(1)}| > 0)$ if $C = \emptyset$ (rejection probability under $H_0$), for both $\alpha = 0.05$ and $\alpha = 0.01$, depending on $\delta$. Over a variety of distributions, including normal, exponential, gamma, binomial and Poisson, for $\delta \geq 20$ the $\alpha$-level is kept throughout.

![Figure 9: Rejection probability under $C = \emptyset$ (1000 simulations) depending on $\delta \in \{2, 4, \ldots, 50\}$. $T = 1000$, left: $\alpha = 5\%$, right: $\alpha = 1\%$. Six distributions color coded: $N(0, 1)$ (magenta), $\text{Pois}(1)$ (orange), $\exp(1)$ (green), $\text{b}(10, 1/2)$ (black), $\text{gamma}(0.5, 2)$ (blue) and $\text{gamma}(2, 2)$ (red). Right: Legend.]

In the following we use $\delta = g = 20$ and $\kappa$ is derived in simulations for $\alpha = 0.01$.

**Simulation studies** We evaluate the performance of MSCP in simulation studies. Throughout it is $T = 1000$ and $|C| = 5$. We consider different scenarios of locations $c_u \in C$ and parameters $\mu_u$ and $\sigma_u$, given in Table 1. Scenarios 1–3 differ in $C$: in 1 it is $c_{u+1} - c_u = 200$, and in 2 it is $c_{u+1} - c_u = 100$ throughout, and 3 describes a multiscale setup where $c_1 = 200$ is well separated and for the others it is $c_{u+1} - c_u = 50$. For each choice of $C$ we then consider different effects by altering both changes in $\mu_u$ and $\sigma_u$. E.g., 1a states certain $\mu_u$ which are halved in 1c. Also, in 1a it is $\sigma_u = 1$ constant, while in 1b the $\sigma_u$ alter between 1 and 2. For each scenario we consider different distributions: A. $N(\mu, \sigma^2)$ (normal), B. $\text{gamma}(s, \lambda)$ (gamma), C. $\text{Poi}(\lambda)$ (Poisson), D. $\text{b}(10, p)$ (binomial, $n = 10$ fix), and E. is a combination of the previous that changes between the six sections using A.,B.,C.,D.,A.,B. (mix). Note that for $\text{Poi}(\lambda)$ and $\text{b}(10, p)$ the formulation of $\sigma_u$ can be disregarded as it follows from $\mu_u$, and also if distributions coincide due to equality of $\mu_u$ as e.g., in 1a and 1b, they are presented only once.

For each scenario we run 1000 simulations, i.e., there are $1000 \cdot |C| = 5000$ change points in total. W.r.t. a critical value $V$ we classify an estimate $\hat{c}$ as correct if $m_{\hat{c}} := \min_{c \in C} |\hat{c} - c| \leq V$. Let $\hat{C}_V$ denote the set of all correct estimates w.r.t. $V$. Then $\hat{C}_T$ is the set of all estimates. We compute $|\hat{C}_T|$, and for $V \in \{10, 5, 2\}$ both $|\hat{C}_V|$ and the mean absolute deviation $M_V := |\hat{C}_V|^{-1} \sum_{\hat{c} \in \hat{C}_V} m_{\hat{c}}$. The results, depending on the scenarios and distributions, are reported in Tables 2.

We summarize the results. Scenarios 1a and 1b are well-posed in the sense that results are strong and yield precise estimates: $|\hat{C}_T|$ is close to 5000 over all scenarios and distributions. Notably, in 1a it is exactly $|\hat{C}_{10}| = 5000$ throughout, and $|\hat{C}_5| = 5000$ for three distributions. Also, the strict measure $|\hat{C}_2|$ reveals precise results. In 1b we consider varying $\sigma_u$, but the estimates remain similarly precise. The setups 2a and 2b are again well-posed: 2a and 2b are pretty similar to 1a and 1b. In setups 3a and 3b performance is slightly reduced as compared
| Scenario | Distribution | $C$ | $\mu_\alpha$ | $\sigma_\alpha$ |
|----------|--------------|----|-------------|-------------|
| 1a       | A. (normal)  | 100, 300, 500, 700, 900 | 1,4,1,8,1,4 | 1,1,1,1,1,1 |
| 1b       | B. (gamma)   | 100, 300, 500, 700, 900 | 1,4,1,8,1,4 | 1,2,1,2,1,2 |
| 1c       | C. (Poisson) | 100, 300, 500, 700, 900 | 0.5, 2,0.5, 4,0.5, 2 | 1,1,1,1,1,1 |
| 2a       | B. (gamma)   | 300, 400, 500, 600, 700 | 1,4,1,8,1,4 | 1,1,1,1,1,1 |
| 2b       | B. (gamma)   | 300, 400, 500, 600, 700 | 1,4,1,8,1,4 | 1,2,1,2,1,2 |
| 2c       | B. (gamma)   | 300, 400, 500, 600, 700 | 0.5, 2,0.5, 4,0.5, 2 | 1,1,1,1,1,1 |
| 3a       | B. (gamma)   | 200, 500, 550, 600, 750 | 1,4,1,8,1,4 | 1,1,1,1,1,1 |
| 3b       | B. (gamma)   | 200, 500, 550, 600, 750 | 1,4,1,8,1,4 | 1,2,1,2,1,2 |
| 3c       | B. (gamma)   | 200, 500, 550, 600, 750 | 0.5, 2,0.5, 4,0.5, 2 | 1,1,1,1,1,1 |
| 3d       | B. (gamma)   | 200, 500, 550, 600, 750 | 0.5, 2,0.5, 4,0.5, 2 | 1,2,1,2,1,2 |
| 3e       | B. (gamma)   | 200, 500, 550, 600, 750 | 1,2,4,8,4,2   | 1,1,1,1,1,1 |

Table 1: Scenarios of $C$, $\mu$ and $\sigma$ considered in simulations.

| Scenario | Distribution | $|C_T|$ | $|C_{10}|$, $M_{10}$ | $|C_5|$, $M_5$ | $|C_2|$, $M_2$ |
|----------|--------------|-------|----------------------|----------------|----------------|
| 1a       | A. (normal)  | 0005  | 5000, 0.1            | 5000, 0.1      | 4994, 0.1      |
| 1a       | B. (gamma)   | 0002  | 5000, 0.1            | 5000, 0.1      | 4993, 0.1      |
| 1a       | C. (Poisson) | 0005  | 5000, 0.4            | 4993, 0.3      | 4905, 0.3      |
| 1a       | D. (binomial)| 0005  | 5000, 0.2            | 4998, 0.2      | 4954, 0.2      |
| 1a       | E. (mix)     | 0001  | 5000, 0.1            | 5000, 0.1      | 4993, 0.1      |
| 1b       | A. (normal)  | 0001  | 4998, 0.3            | 4993, 0.3      | 4906, 0.2      |
| 1b       | B. (gamma)   | 0000  | 5000, 0.3            | 4997, 0.3      | 4909, 0.2      |
| 1b       | E. (mix)     | 0004  | 5000, 0.3            | 4999, 0.3      | 4906, 0.2      |
| 1c       | A. (normal)  | 0051  | 4935, 0.5            | 4912, 0.5      | 4698, 0.4      |
| 1c       | B. (gamma)   | 0053  | 4932, 0.4            | 4925, 0.4      | 4823, 0.3      |
| 1c       | C. (Poisson) | 0064  | 4626, 0.6            | 4600, 0.6      | 4370, 0.5      |
| 1c       | D. (binomial)| 0091  | 4883, 0.6            | 4858, 0.5      | 4642, 0.4      |
| 1c       | E. (mix)     | 0093  | 4929, 0.6            | 4903, 0.5      | 4707, 0.4      |

Table 2: Simulation results for scenarios 1a–c.

| Scenario | Distribution | $|C_T|$ | $|C_{10}|$, $M_{10}$ | $|C_5|$, $M_5$ | $|C_2|$, $M_2$ |
|----------|--------------|-------|----------------------|----------------|----------------|
| 2a       | A. (normal)  | 0002  | 5000, 0.1            | 5000, 0.1      | 4993, 0.1      |
| 2a       | B. (gamma)   | 0002  | 5000, 0.1            | 5000, 0.1      | 4988, 0.1      |
| 2a       | C. (Poisson) | 0005  | 5000, 0.3            | 4998, 0.3      | 4908, 0.3      |
| 2a       | D. (binomial)| 0002  | 5000, 0.2            | 5000, 0.2      | 4958, 0.2      |
| 2a       | E. (mix)     | 0002  | 5000, 0.1            | 5000, 0.1      | 4989, 0.1      |
| 2b       | A. (normal)  | 0002  | 4998, 0.3            | 4995, 0.3      | 4914, 0.2      |
| 2b       | B. (gamma)   | 0003  | 5000, 0.3            | 4998, 0.3      | 4928, 0.2      |
| 2b       | E. (mix)     | 0008  | 5000, 0.3            | 4998, 0.3      | 4935, 0.2      |
| 2c       | A. (normal)  | 0084  | 4873, 0.5            | 4855, 0.5      | 4663, 0.4      |
| 2c       | B. (gamma)   | 0066  | 4870, 0.4            | 4863, 0.3      | 4766, 0.3      |
| 2c       | C. (Poisson) | 0053  | 4541, 0.7            | 4520, 0.6      | 4285, 0.5      |
| 2c       | D. (binomial)| 0047  | 4841, 0.5            | 4828, 0.5      | 4647, 0.4      |
| 2c       | E. (mix)     | 0026  | 4920, 0.5            | 4902, 0.5      | 4720, 0.4      |

Table 3: Simulation results for scenarios 2a–c.
to 2a and 2b due to closer distances. Nevertheless, we obtain about 5000 estimates at precise location for 3a and 3b. In scenarios 1c, 2c and 3c we have jump sizes halved as compared to 1a-3a and 1b-3b. Here, deficiencies start to arise: $|\hat{C}_T|$ reduces to about 4750 to 4950 for distributions A., B., D. and E., while for C., it drops to about 4400 to 4800. In scenario 3d we reduced the effects w.r.t. both smaller jump sizes and higher variances, resulting in higher performance losses as we obtain only around 3000 estimates. But note that e.g., for $c_6$ it is $\mu_6 - \mu_5 = 1.5$ with $\sigma_6 = 2$ which is reasonably hard to be detected, also taking into account the smaller spacings between $c_u$. Despite the reduced number, the precision for detected estimates is kept high. Scenario 3e is another example with different jump sizes resulting in a good overall outcome, but slightly weaker than 3a as effects are reduced similarly to 3c. Overall, MSCP yields reliable results over different scenarios of change points and effects: both the number and the locations are reliably estimated. Notably, performance is also kept over different distributions and regardless of changes in variance.

### Comparison with state of the art methods

We present simulation studies which reveal competitiveness of mscp with existing change point detection methods available on CRAN. For all competitors we consider the scenarios 1c, 2c and 3c, see Table 1 and all distributions A.–E. from the previous paragraph. The reason for the choice of theses scenarios is that for MSCP deficiencies in detection started to arise. Again, for each such setup we run 1000 simulations which yields 5000 change points in total.

Note that between the three scenarios the locations of $C$ vary while $\mu_u = 0.5, 2, 0.5, 4, 0.5, 2$ and $\sigma_u = 1, \ldots, 1$ remain unchanged. In particular, due to constant $\sigma_u$, results should not be

| Scenario | Distribution | $|\hat{C}_T|$ | $|\hat{C}_{10}|, M_{10}$ | $|\hat{C}_5|, M_5$ | $|\hat{C}_2|, M_2$ |
|----------|--------------|-------------|----------------|----------------|----------------|
| 3a       | A. (normal)  | 5004        | 4990, 0.3      | 4910, 0.2      | 4846, 0.1      |
| 3a       | B. (gamma)   | 5002        | 4990, 0.3      | 4879, 0.2      | 4828, 0.1      |
| 3a       | C. (Poisson) | 4942        | 4876, 0.9      | 4580, 0.5      | 4334, 0.3      |
| 3a       | D. (binomial)| 5005        | 4965, 0.6      | 4787, 0.3      | 4652, 0.2      |
| 3a       | E. (mix)     | 5000        | 4991, 0.3      | 4901, 0.1      | 4857, 0.1      |
| 3b       | A. (normal)  | 4988        | 4928, 0.8      | 4657, 0.4      | 4478, 0.3      |
| 3b       | B. (gamma)   | 4995        | 4939, 0.8      | 4689, 0.4      | 4482, 0.2      |
| 3b       | E. (mix)     | 5001        | 4947, 0.6      | 4789, 0.3      | 4638, 0.2      |
| 3c       | A. (normal)  | 4814        | 4703, 1.3      | 4286, 0.7      | 3936, 0.4      |
| 3c       | B. (gamma)   | 4820        | 4749, 1.2      | 4334, 0.5      | 4095, 0.3      |
| 3c       | C. (Poisson) | 4387        | 4249, 1.5      | 3845, 0.8      | 3480, 0.5      |
| 3c       | D. (binomial)| 4750        | 4644, 1.5      | 4146, 0.7      | 3809, 0.4      |
| 3c       | E. (mix)     | 4856        | 4756, 1.3      | 4334, 0.7      | 3990, 0.4      |
| 3d       | A. (normal)  | 3093        | 2946, 1.6      | 2620, 0.8      | 2380, 0.5      |
| 3d       | B. (gamma)   | 2962        | 2842, 1.8      | 2472, 0.8      | 2193, 0.4      |
| 3d       | E. (mix)     | 2896        | 2804, 1.4      | 2548, 0.7      | 2296, 0.4      |
| 3e       | A. (normal)  | 4680        | 4566, 1.1      | 4257, 0.5      | 3984, 0.3      |
| 3e       | B. (gamma)   | 4735        | 4620, 1.0      | 4302, 0.5      | 4046, 0.3      |
| 3e       | C. (Poisson) | 3411        | 3085, 2.3      | 2584, 1.2      | 2154, 0.6      |
| 3e       | D. (binomial)| 4129        | 3883, 1.8      | 3346, 0.8      | 2959, 0.4      |
| 3e       | E. (mix)     | 4220        | 4014, 1.8      | 3452, 0.8      | 3070, 0.4      |

Table 4: Simulation results for scenarios 3a-e.
affected through a method’s capability of handling changes in variance, which holds true for MSCP. For that also keep in mind that, in fact, \( \sigma_u \equiv 1 \) only holds true for A. B. and parts of E., while for C. and D. \( \sigma_u \) always changes with \( \mu_u \).

As competitors of mscp we consider mosum [Meier et al. (2021)] with individual windows \( G=50, 100 \) and 200, respectively, wbs [Baranowski and Fryzlewicz (2019)], not [Baranowski et al. (2019b)] with contrast=pcwsConstMeanVar and pcwsConstMeanHT, respectively, changepoint [Killick et al. (2016)] considering cpt.meanvar with method=PELT and test.stat='Normal' for A. D. and E., 'Gamma' for B., and 'Poisson' for C., stepR [Pein et al. (2020)] with stepFit, cumSeg [Muggeo (2020)] using jumpoints, and FDRSeg [Li and Sieling (2017)] using fdrseg with sd estimating the global standard deviation. Any other tuning-parameters are kept default.

The results are presented in Table 5 (scenario 1c), Table 6 (scn. 2c) and Table 7 (scn. 3c). We first recall that overall, mscp yields a total number of estimates \(|\hat{C}_T|\) of about 4750 to 4950 for distributions A., B., D. and E., and about 4400 to 4800 for C. Further, recall high estimation precision regarding different levels of tolerance, i.e., values of \(|\hat{C}_V|\) and \(M_V\) for \( V = \{10, 5, 2\} \). In total, also the competitors often show a similar performance in many scenarios, but also somewhat weaken in certain setups. For mosum we nicely see, e.g., in Table 5 A., that performance depends on the selection of the individual window. For example we find \( G=50 \) to slightly overestimate, and \( G=200 \) to underestimate the total number of change points, while the middle window \( G=100 \) performs best, also noting slightly reduced location precision for small windows as compared to mscp. Note that overall the problem of a fixed window becomes even more evident when change point distances are reduced, see Table 6 and 7, particularly because large windows lose sensitivity. Of course, window selection is a major challenge in practice as change point locations are unknown, and mscp circumvents it by nature. Apart from this, mosum yields quite stable results among different distributions, but similar to mscp it loses power for distribution C. For wbs we find good performance for distribution A., w.r.t. both the number and precision of estimates over all scenarios, while it tends to overestimate for other distributions, most extreme for C. and B. We overall find good performance for the method not. However, we mention that contrast=pcwsConstMeanVar tends to overestimate particularly for distribution B., noting that the contrast assumes normality. Interestingly, contrast=pcwsConstMeanHT, which assumes heavy-tailed additive errors, overall performs nicely, while showing somewhat reduced estimation precision, as compared to mscp. For changepoint we find great performance for A. and C. where parametric assumptions are met, but tremendous overestimation if not, see D. and E. Although parametric assumptions are met in B., it also overestimates. stepR shows reliable results in A., while it tends to overestimate the number of change points if normality is violated. This tendency is less drastic in E. where the distribution is normal in at least two sections, or D., but more problematic for distribution B. Over all setups cumSeg perform nicely and only very slightly overestimates the number of change points, but shows somewhat reduced estimation precision as compared to mscp. FDRSeg also shows good performance among almost all setups presented, w.r.t. both different scenarios and distributions. However, we mention overestimation in C., and a slight reduction in estimation precision as compared to mscp.

In summary, the results presented reveal serious competitiveness of mscp with state of the art change point detection methods. Particularly, mscp overcomes the selection of a tuning bandwidth as compared to single window MOSUM techniques. Also, the consideration of sce-
narios 1c–3c reveals comparably good performance regarding different change point distances. Besides that, a comparably stable performance over different distributions supports the value of the non-parametric nature of mscp for practice.

| Scn | Distr. | Method | |C| | |C||M0| | |C||M0| | |C||M0| |
|---|---|---|---|---|
| 1c | A. (normal) | mscp | 4951 | 4935, 0.5 | 4912, 0.3 | 4698, 0.4 |
| | | mosum G = 50 | 5057 | 4969, 0.9 | 4809, 0.7 | 4378, 0.4 |
| | | mosum G = 100 | 5062 | 4969, 0.8 | 4852, 0.7 | 4430, 0.4 |
| | | mosum G = 200 | 4962 | 3972, 0.4 | 3905, 0.3 | 3757, 0.2 |
| | | wbs | 5030 | 4983, 0.7 | 4890, 0.6 | 4320, 0.4 |
| | | not MeanVar | 5007 | 4986, 0.7 | 4886, 0.6 | 4517, 0.4 |
| | | not MeanHT | 5010 | 4974, 0.8 | 4846, 0.6 | 4438, 0.4 |
| | | changepoint | 5035 | 5062, 0.8 | 4893, 0.6 | 4323, 0.4 |
| | | stepR | 5037 | 4978, 1.4 | 4863, 1.2 | 4443, 1 |
| | | cumSeg | 5019 | 4935, 1.5 | 4782, 1.4 | 3953, 0.9 |
| | | FDRSeg | 5000 | 4979, 1.4 | 4867, 1.2 | 4446, 1 |
| B. (gamma) | mscp | 4953 | 4932, 0.4 | 4925, 0.4 | 4382, 0.3 |
| | | mosum G = 50 | 5082 | 4873, 1 | 4658, 0.7 | 4223, 0.4 |
| | | mosum G = 100 | 5014 | 4889, 0.9 | 4696, 0.6 | 4286, 0.4 |
| | | mosum G = 200 | 4905 | 3948, 0.4 | 3988, 0.3 | 3748, 0.2 |
| | | wbs | 29906 | 6827, 2.1 | 5692, 0.9 | 4846, 0.4 |
| | | not MeanVar | 15799 | 5076, 1.9 | 4442, 0.8 | 3708, 0.2 |
| | | not MeanHT | 5020 | 4991, 0.6 | 4928, 0.5 | 4611, 0.4 |
| | | changepoint | 14442 | 5529, 2.6 | 4397, 1.3 | 3390, 0.6 |
| | | stepR | 30957 | 6568, 2.7 | 5408, 1.5 | 4504, 1.1 |
| | | cumSeg | 5033 | 4911, 1.5 | 4734, 1.3 | 4016, 0.9 |
| | | FDRSeg | 6230 | 4922, 1.5 | 4760, 1.3 | 4350, 1 |
| C. (Poisson) | mscp | 4640 | 4626, 0.6 | 4600, 0.6 | 4370, 0.5 |
| | | mosum G = 50 | 4812 | 4679, 1.1 | 4512, 0.9 | 4002, 0.5 |
| | | mosum G = 100 | 4847 | 4752, 1.1 | 4561, 0.8 | 4048, 0.5 |
| | | mosum G = 200 | 3800 | 3886, 0.5 | 3844, 0.4 | 3672, 0.2 |
| | | wbs | 13155 | 5447, 1.6 | 4948, 0.9 | 4226, 0.5 |
| | | not MeanVar | 5231 | 4991, 1 | 4755, 0.8 | 4276, 0.4 |
| | | not MeanHT | 5690 | 4906, 1.1 | 4679, 0.8 | 4182, 0.5 |
| | | changepoint | 5000 | 4964, 0.9 | 4834, 0.8 | 4302, 0.4 |
| | | stepR | 12518 | 5253, 2.1 | 4743, 1.5 | 3953, 1.1 |
| | | cumSeg | 5021 | 4896, 1.8 | 4637, 1.5 | 3700, 1 |
| | | FDRSeg | 5253 | 4921, 1.7 | 4685, 1.4 | 4030, 1.1 |
| D. (binomial) | mscp | 4891 | 4883, 0.6 | 4858, 0.5 | 4642, 0.4 |
| | | mosum G = 50 | 4887 | 4739, 0.9 | 4585, 0.7 | 4165, 0.4 |
| | | mosum G = 100 | 4846 | 4794, 1 | 4622, 0.7 | 4189, 0.4 |
| | | mosum G = 200 | 3911 | 3886, 0.4 | 3854, 0.3 | 3715, 0.2 |
| | | wbs | 6472 | 5035, 1 | 4816, 0.7 | 4415, 0.4 |
| | | not MeanVar | 5184 | 4972, 0.9 | 4798, 0.7 | 4399, 0.4 |
| | | not MeanHT | 5055 | 4952, 1 | 4764, 0.7 | 4334, 0.4 |
| | | changepoint | 191707 | 21855, 4.9 | 12235, 2.5 | 6215, 1 |
| | | stepR | 6369 | 4987, 1.6 | 4756, 1.4 | 4208, 1.1 |
| | | cumSeg | 5009 | 4928, 1.6 | 4726, 1.4 | 3934, 0.9 |
| | | FDRSeg | 5001 | 4953, 1.6 | 4772, 1.3 | 4263, 1.1 |
| E. (mix) | mscp | 4936 | 4929, 0.6 | 4903, 0.5 | 4707, 0.4 |
| | | mosum G = 50 | 5040 | 4971, 0.9 | 4848, 0.7 | 4414, 0.4 |
| | | mosum G = 100 | 4996 | 4969, 0.8 | 4861, 0.7 | 4472, 0.4 |
| | | mosum G = 200 | 4001 | 3973, 0.4 | 3912, 0.3 | 3750, 0.2 |
| | | wbs | 6406 | 5056, 0.8 | 4926, 0.6 | 4547, 0.4 |
| | | not MeanVar | 5180 | 4995, 0.8 | 4876, 0.6 | 4504, 0.4 |
| | | not MeanHT | 5050 | 4987, 0.8 | 4894, 0.6 | 4519, 0.4 |
| | | changepoint | 76448 | 11544, 3.9 | 7751, 1.9 | 5740, 0.8 |
| | | stepR | 6288 | 5020, 1.4 | 4905, 1.2 | 4481, 1 |
| | | cumSeg | 5007 | 4938, 1.5 | 4819, 1.3 | 4034, 0.9 |
| | | FDRSeg | 5068 | 4984, 1.3 | 4909, 1.2 | 4499, 1 |

Table 5: Comparative simulation study w.r.t. scenario 1c.

Data example The complete sequence of a human genome (T2T-CHM13) was recently presented for the first time in Nurk et al. (2022). We analyze the nucleotide sequence of
| Scen. | Distrib. | Method | $|C_T|$ | $|C_{10}|$, $M_0$ | $|C_{20}|$, $M_0$ | $|C_{50}|$, $M_0$ | $|C_{70}|$, $M_2$ |
|-------|----------|--------|--------|----------------|----------------|----------------|----------------|
| 2c    | A. (normal) | mscp  | 4754 | 4873 | 0.2 | 4855 | 0.6 | 4693 | 0.4 |
|       | mscG     | 5067 | 4964 | 0.9 | 4798 | 0.7 | 4348 | 0.4 |
|       | mscG     | 5015 | 4992 | 0.4 | 4935 | 0.3 | 4749 | 0.2 |
|       | mscG     | 3999 | 1505 | 2.3 | 1266 | 1.3 | 994 | 0.6 |
|       | mscG     | 5021 | 4985 | 0.7 | 4872 | 0.6 | 4539 | 0.4 |
|       | mscG     | 5019 | 4997 | 0.4 | 4896 | 0.6 | 4532 | 0.4 |
|       | mscG     | 5008 | 4977 | 0.8 | 4949 | 0.6 | 4384 | 0.4 |
|       | mscG     | 5022 | 4968 | 0.8 | 4887 | 0.6 | 4318 | 0.4 |
|       | stepR    | 5046 | 4979 | 1.4 | 4859 | 1.2 | 4290 | 1  |
|       | cumSeg   | 5063 | 4890 | 1.7 | 4633 | 1.4 | 3750 | 0.9 |
|       | FDRSeg   | 5000 | 4979 | 1.4 | 4860 | 1.2 | 4422 | 1  |
|       | mscp     | 4706 | 4870 | 0.4 | 4863 | 0.3 | 4766 | 0.3 |
|       | mscp     | 5073 | 4881 | 1  | 4695 | 0.6 | 4232 | 0.3 |
|       | mscp     | 5048 | 4938 | 0.5 | 4850 | 0.3 | 4646 | 0.2 |
|       | mscp     | 3873 | 1475 | 1.9 | 1269 | 1  | 1060 | 0.4 |
|       | mscp     | 35098 | 6790 | 2.1 | 5698 | 0.9 | 4359 | 0.4 |
|       | mscp     | 15200 | 2989 | 2  | 4217 | 0.9 | 3614 | 0.3 |
|       | mscp     | 5041 | 4985 | 0.6 | 4917 | 0.5 | 4602 | 0.4 |
|       | mscp     | 14300 | 5495 | 2.5 | 4407 | 1.2 | 3120 | 0.4 |
|       | FDRSeg   | 5063 | 4850 | 1.7 | 4603 | 1.4 | 3792 | 0.9 |
|       | FDRSeg   | 7008 | 4965 | 1.5 | 4763 | 1.3 | 4331 | 1  |
|       | mscp     | 4553 | 4541 | 0.7 | 4520 | 0.6 | 4286 | 0.5 |
|       | mscp     | 4820 | 4682 | 1.1 | 4482 | 0.8 | 3984 | 0.5 |
|       | mscp     | 4741 | 4681 | 0.5 | 4609 | 0.4 | 4396 | 0.2 |
|       | mscp     | 4394 | 1354 | 1.9 | 1193 | 1.2 | 3794 | 0.6 |
|       | mscp     | 10456 | 5351 | 1.5 | 4907 | 0.9 | 2270 | 0.4 |
|       | mscp     | 5215 | 4930 | 1.1 | 4729 | 0.8 | 4228 | 0.5 |
|       | mscp     | 5094 | 4863 | 1.2 | 4662 | 0.8 | 4316 | 0.5 |
|       | mscp     | 5000 | 4966 | 0.9 | 4818 | 0.7 | 4361 | 0.4 |
|       | mscp     | 9851 | 5222 | 2.1 | 4718 | 1.5 | 3916 | 1.1 |
|       | mscp     | 5039 | 4794 | 1.9 | 4487 | 1.6 | 3488 | 1  |
|       | mscp     | 5742 | 4996 | 1.8 | 4689 | 1.4 | 3902 | 1.1 |
|       | mscp     | 4847 | 4841 | 0.5 | 4828 | 0.5 | 4477 | 0.4 |
|       | mscp     | 4855 | 4728 | 0.9 | 4600 | 0.7 | 4173 | 0.4 |
|       | mscp     | 4799 | 4757 | 0.4 | 4711 | 0.3 | 4538 | 0.2 |
|       | mscp     | 3415 | 1389 | 1.8 | 1247 | 1.1 | 1049 | 0.6 |
|       | mscp     | 5984 | 5004 | 0.9 | 4837 | 0.7 | 4430 | 0.4 |
|       | mscp     | 5108 | 4957 | 0.9 | 4807 | 0.7 | 4426 | 0.4 |
|       | mscp     | 5076 | 4949 | 1  | 4775 | 0.7 | 4349 | 0.4 |
|       | mscp     | 195010 | 21834 | 4.9 | 12190 | 2.5 | 6114 | 1  |
|       | mscp     | 5094 | 4988 | 1.6 | 4786 | 1.3 | 4219 | 1  |
|       | mscp     | 5030 | 4864 | 1.7 | 4615 | 1.4 | 3748 | 0.9 |
|       | mscp     | 5034 | 4953 | 1.5 | 4803 | 1.3 | 4245 | 1  |
|       | mscp     | 4926 | 4920 | 0.5 | 4902 | 0.6 | 4720 | 0.4 |
|       | mscp     | 5042 | 4963 | 0.8 | 4859 | 0.7 | 4469 | 0.4 |
|       | mscp     | 5026 | 4991 | 0.4 | 4950 | 0.3 | 4794 | 0.2 |
|       | mscp     | 3822 | 1551 | 2  | 1357 | 1.2 | 1110 | 0.6 |
|       | mscp     | 5580 | 5059 | 0.8 | 4954 | 0.6 | 4600 | 0.4 |
|       | mscp     | 5115 | 4990 | 0.7 | 4898 | 0.6 | 4542 | 0.4 |
|       | mscp     | 5022 | 4984 | 0.8 | 4883 | 0.7 | 4514 | 0.4 |
|       | mscp     | 40166 | 11549 | 3.9 | 7748 | 1.9 | 6500 | 0.8 |
|       | mscp     | 5863 | 5042 | 1.4 | 4920 | 1.2 | 4484 | 1  |
|       | mscp     | 5053 | 4895 | 1.7 | 4682 | 1.4 | 3835 | 0.9 |
|       | mscp     | 5044 | 4996 | 1.3 | 4917 | 1.2 | 3510 | 1  |

Table 6: Comparative simulation study w.r.t. scenario 2c.

the X-chromosome (Homo sapiens isolate CHM13 chromosome X, GenBank: CP068255.2). It consists of about $b = 1542 \cdot 10^5$ base pairs. We segment the sequence into segments of length $d = 2 \cdot 10^5$, within which we compute the frequency of the base cytosine. This yields $T = b/d = 771$ data points to which we apply MSCP, see Figure[10].

Nine change points were estimated, $\hat{C} = \{20, 80, 124, 181, 383, 479, 635, 699, 739\}$, yielding parameter estimates $\hat{\mu}_c \approx (45, 40, 41, 37, 41, 36, 39, 40, 38, 44) \cdot 10^5$ and
\[ \sigma_u \approx (4.5, 2.3, 2.5, 1.5, 3.2, 1.0, 2.6, 2.1, 1.5, 5.7) \cdot 10^3 \]

The segmentation closely aligns with visual inspection. We note that model assumptions are fairly met: first, an increase in mean is accompanied with an increase in variance, and vice versa, which aligns with the assumption that the theoretical variance may also change when change in expectation occurs. Second, in each section, serial correlation proves to be moderate (e.g., for lag one it is between 0.3 (section 6) and 0.7 (last section 10), and it rapidly decreases for higher lags).
representation of the data as frequency counts is supported by the non-parametric assumptions on the distributions. Finally, we mention that the result is stable under variation of $d$. Being an interesting result in itself, the example shows that MSCP can be helpful for the segmentation of genomic data: yielding segments with relatively stable first order moments, it can be considered a preprocessing step in data analysis.

6 Appendix

Proof of Lemma 3.1 W.l.o.g. let $j = \ell$. For $\hat{\mu}_\ell$ first let $C = \emptyset$ where we need to show
\[ n^v \cdot (\hat{\mu}_\ell - \mu^{(k)})_{(t,h)} \to (0)_{(t,h)} \text{ a.s. as } n \to \infty. \]
We show some uniform convergence w.r.t $[0,T]$ and then extend considerations to $\Delta_\delta$. It holds a.s. as $n \to \infty$
\[ n^v \cdot \sup_{t \in [0,T]} \left| \frac{1}{n} \sum_{i=1}^{nt} X^k_i - t\mu^{(k)} \right| \to 0, \] for $v \in \begin{cases} (-\infty, 1/2), & \text{if } k = 1, \\ (-\infty, 1/2) \cap (-\infty, p/(p + 2)], & \text{if } k = 2. \end{cases}$ (13)

For this, first apply Marcinkiewicz-Zygmund (MZ)-SLLN to $(X_i)_i$ and $(X^2_i)_i$: For $(X_i)_i$ note that for any $\alpha \in (0, 2)$ it is $\mathbb{E}[|X_i|^\alpha] < \infty$, and thus for $k = 1$ a.s. as $n \to \infty$
\[ n^{(\alpha - 1)/\alpha} \cdot \left( \frac{1}{n} \sum_{i=1}^{n} X^k_i - \mu^{(k)} \right) = n^{-1/\alpha} \cdot \left( \sum_{i=1}^{n} X^k_i - n\mu^{(k)} \right) \to 0. \] (14)

For the squares $(X^2_i)_i$, we choose $\alpha \in (0, (2 + p)/2]$ if $p < 2$ and $\alpha \in (0, 2)$ if $p \geq 2$. Then $\mathbb{E}[|X^2_i|^\alpha] < \infty$ and thus (14) applies for $k = 2$. As $(\alpha - 1)/\alpha \leq p/(2 + p)$ if $p < 2$, and
(α − 1)/α < 1/2 if p ≥ 2, we obtain for k = 1, 2 a.s. as n → ∞

\[ n^v \cdot \left( \frac{1}{n} \sum_{i=1}^{n} X_i^k - \mu^{(k)} \right) \to 0, \quad \text{for } v \in \begin{cases} (-\infty, 1/2), & \text{if } k = 1, \\ (-\infty, 1/2) \cap (-\infty, p/(p + 2)], & \text{if } k = 2. \end{cases} \tag{15} \]

For uniform convergence in \((13)\) we discretize time and apply \((15)\), a.s. as \(n \to \infty\)

\[ n^v \cdot \max_{m \in \{1, \ldots, nT\}} \left| \frac{1}{n} \sum_{i=1}^{m} X_i^k - \frac{m}{n} \mu^{(k)} \right| \leq n^v \cdot \sup_{0 \leq t \leq T} \left| \frac{1}{n} \sum_{i=1}^{[nt]} X_i^k - \frac{[nt]}{n} \mu^{(k)} \right| + r_n \]

\[ = n^v \cdot \max_{m \in \{1, \ldots, nT\}} \left| \frac{1}{n} \sum_{i=1}^{m} X_i^k - \frac{m}{n} \mu^{(m)} \right| + r_n \to 0, \]

as \(r_n := n^v \cdot \sup_{t \in [0, T]} |t - [nt]/n| \mu^{(k)} \leq n^v \cdot (1/n) \mu^{(k)} \to 0\) as \(n \to \infty\), and for \(\varepsilon > 0\)

\[ n^v \cdot \max_{m \leq m_0} \frac{1}{n} \sum_{i=1}^{m} X_i^k - m \mu^{(k)} \leq T^{-v} \max_{m_0 < m \leq nT} \left( \frac{m}{nT} \right)^{1-v} n^v \cdot \frac{1}{n} \sum_{i=1}^{m} X_i^k - m \mu^{(k)} \leq \varepsilon \]

a.s. for \(n\) large enough, as the maximum in the first summand is independent from \(n\), and for the second summand note \((m/nT)^{1-v} \leq 1\) and \((15)\) yields that for almost every realization we find \(m_0 \in \mathbb{N}\) such that \(n^v \cdot [(1/m) \sum_{i=1}^{m} X_i^k - \mu^{(k)}] < T^{-1} \varepsilon/2\) for all \(m > m_0\). Thus, \((13)\) holds. Now we consider \(\Delta_\delta\). As \(0 \leq t - h \leq T\) for all \((t, h) \in \Delta_\delta\), \((13)\) yields a.s. as \(n \to \infty\)

\[ n^v \cdot \sup_{(t, h) \in \Delta_\delta} \left| \frac{1}{n} \sum_{i=1}^{[nt]} X_i^k - t \mu^{(k)} \right| \to 0 \quad \text{and} \]

\[ n^v \cdot \sup_{(t, h) \in \Delta_\delta} \left| \frac{1}{n} \sum_{i=1}^{[n(t-h)]} X_i^k - (t - h) \mu^{(k)} \right| \to 0, \]

We include the factor \(1/h \geq 2/T > 0\) and obtain in \((\mathcal{D}_\mathbb{R}[\Delta_\delta], \| \cdot \|_\infty)\) a.s. as \(n \to \infty\)

\[ n^v \cdot (\hat{\mu}_t^{(k)} - \mu^{(k)})_{(t, h)} = n^v \cdot \left[ \left( \frac{1}{nh} \sum_{i=1}^{[nt]} X_i^k - \frac{t}{h} \mu^{(k)} \right) - \left( \frac{1}{nh} \sum_{i=1}^{[n(t-h)]} X_i^k - \frac{t - h}{h} \mu^{(k)} \right) \right]_{(t, h)} \to 0_{(t, h)}. \tag{16} \]

Now let \(C \neq \emptyset\). We segment \((t - h, t]\) according to \(C\) and find \(\hat{\mu}_t^{(k)} - \mu^{(k)}\) as

\[ \sum_{u=1}^{\lfloor C \rfloor + 1} \left( \left\lfloor \frac{n \ell_{t,u}}{nh} \right\rfloor - \left\lfloor n \ell_{t,u-1}/nh \right\rfloor \right) \cdot \frac{1}{\left\lfloor n \ell_{t,u}/nh \right\rfloor - \left\lfloor n \ell_{t,u-1}/nh \right\rfloor} \sum_{i=\left\lfloor n \ell_{t,u-1}/nh \right\rfloor + 1}^{\left\lfloor n \ell_{t,u}/nh \right\rfloor} X_i^k - \frac{d_{t,u}}{h} \mu^{(k)}_{\ell_{t,u}}. \tag{17} \]

The \(u\)-th summand refers to a subsection that relates to the error sequence \((Z_{u,i})_{i=1,2,\ldots}\) and
in which \( X_i \) equals \( \mu_u + \sigma_u \cdot Z_{u,i} \). From (16) we conclude a.s. as \( n \to \infty \)

\[
\nu \cdot \sup_{1 \leq d \leq T} \left| \frac{1}{[nd]} \sum_{i=[n(t-d)]+1}^{[nt]} (\mu_u + \sigma_u \cdot Z_{u,i})^k - \mu_u^k \right| \to 0, \tag{18}
\]

which states uniform convergence w.r.t all subintervals of varying length \( d \). When including the factor \( d/h \geq 1 \) and summing over all \( u \), the expression still vanishes a.s. as \( n \to \infty \), and it states an upper bound for \( \nu \cdot \sup_{(t,h) \in \Delta_k} |\hat{\mu}^{(k)} - \tilde{\mu}^{(k)}| \).

Regarding \( \hat{\sigma}^2 \) we decompose in the \( u \)-th subsection \( (X_i - \hat{\mu})^2 = (X_i - \mu_{\ell,u})^2 - 2(X_i - \mu_{\ell,u})(\hat{\mu}_u - \mu_{\ell,u}) + (\hat{\mu}_u - \mu_{\ell,u})^2 \). Averages in the subsection tend to \( \sigma_{\ell,u}^2 + 0 + (\hat{\mu}_u - \mu_{\ell,u})^2 \), and summation over subsections yields a.s. as \( n \to \infty \) that \( \nu \cdot \sup_{(t,h) \in \Delta_k} |\hat{\sigma}^2 - \tilde{\sigma}^2| \to 0 \), as before using MZ-SLLN and discretization arguments.

**Proof of Lemma 3.3** Continuity is inherited from the limits in [5] and [6]. W.l.o.g. consider \( c_1 = c \). As \( v_{t,h} \geq 1 \) with equality at \( c \) it is \( v_{t,h} \cdot |d_{t,h}^{(n)}| \geq |d_{t,h}^{(n)}| \cdot \tilde{\mu}_r - \hat{\mu}_\ell \) has the shape of a hat: it is zero outside the \( h \)-neighborhood of \( c \), it is \( \mu_2 - \mu_1 \) at \( c \), and it is linearly interpolated in between. \( \tilde{\sigma}_h^2 + \hat{\sigma}_h^2 \) is \( 2\sigma_1^2 \) left of \( c - h \), and \( 2\sigma_2^2 \) right of \( c + h \), and linearly interpolated in between. Thus, in the \( h \)-neighborhood of \( c \) the root \( \left( \tilde{\sigma}_h^2 + \hat{\sigma}_h^2 \right)^{1/2} \) is constant if \( \sigma_2^2 = \sigma_1^2 \), it is strictly convex if \( \sigma_2^2 > \sigma_1^2 \) and strictly concave if \( \sigma_2^2 < \sigma_1^2 \). As the numerator is piecewise linear, i.e., of order \( t \), and the denominator of order \( t^{1/2} \), the statements about the curvatures of \( v_{t,h} \cdot d_{t,h}^{(n)} \) hold true. We now turn to \( d_{t,h}^{(n)} \). We represent \( \tilde{\sigma}_h^2 \) through \( \hat{\sigma}_h^2 \) plus errors

\[
\tilde{\sigma}_h^2 = \sum_{u=1}^{2} \frac{d_u}{h} \cdot \tilde{\sigma}_u^2 + (\tilde{\mu}_u - \mu_u^2) = \tilde{\sigma}_h^2 + \frac{d_1 d_2}{h^2} \cdot (\mu_2 - \mu_1)^2,
\]

for \( j \in \{t, r\} \), i.e., we find the error as \( e_j^2 = |d_1 d_2/h^2|(\mu_2 - \mu_1)^2 \). We abbreviated \( d_u := d_{j,u} \).

Note that \( e_2^2 = e_t^2 \), and set \( x := d_1/h \) and \( d_2/h = (1 - h)/h = (1 - x) \) with \( x \in [0, 1] \), which yields a representation through the proportions of the window left and right of \( c \). Note that the error \( e_t^2 = x(1 - x) \cdot (\mu_2 - \mu_1)^2 \) is quadratic and maximal for \( x = 1/2 \), thus taking the value \( (\mu_2 - \mu_1)^2/4 \), which is plausible as half of the window refers to the left and the other half to the right population. Within both \( [c - h, c] \) and \( (c, c + h) \) it holds that \( \tilde{\mu}_c - \hat{\mu}_c \) is linear in \( x \) and the denominator of \( d_{t,h}^{(n)} \) is now the root of first function \( \tilde{\sigma}_h^2 + \hat{\sigma}_h^2 \) which is linear in \( x \) plus second the quadratic error \( e_t^2 \). W.l.o.g. we consider \( t \in [c - h, c] \) where the right window contains \( c \). Then we find \( d_{t,h}^{(n)} \) as a function \( f \) of \( x \) as

\[
f(x) = \sqrt{nh} \cdot \frac{(\mu_2 - \mu_1) \cdot (1 - x) \cdot x}{\sqrt{[(\sigma_2^2 - \sigma_1^2) \cdot x + 2\sigma_1^2] + [(\mu_2 - \mu_1)^2 \cdot x \cdot (1 - x)]}}, \quad x \in [0, 1],
\]

for all \( h \in (\delta, T/2] \), which yields the derivative w.r.t. \( x \)

\[
f'(x) = \sqrt{nh} \cdot (\mu_2 - \mu_1) \cdot \frac{2^{-1}[((\sigma_2^2 - \sigma_1^2) + (\mu_2 - \mu_1)^2) \cdot x + 2\sigma_1^2]}{[(\sigma_2^2 - \sigma_1^2) \cdot x + 2\sigma_1^2 + (\mu_2 - \mu_1)^2 \cdot x \cdot (1 - x)]^{3/2}}.
\]

In [19] we see that the fraction is positive and thus \( f'(x) > 0 \) if \( \mu_2 > \mu_1 \) and \( f'(x) < 0 \) if \( \mu_2 < \mu_1 \), which gives (5). To bound \( |f'(x)| \) we find the numerator \( \geq 2^{-1}[\sigma_2^2 + 3\sigma_1^2 + (\mu_2 - \mu_1)^2] \land 2\sigma_1^2 \)
and for the denominator we get $(\sigma_2^2 - \sigma_1^2)x + 2\sigma_1^2 \leq 2(\sigma_2^2 \vee \sigma_1^2)$ and $x(1-x) \leq 1/4$, such that

$$|f'(x)| \geq \sqrt{n}h \cdot |\mu_2 - \mu_1| \cdot \frac{2^{-1}[\sigma_2^2 + 3\sigma_1^2 + (\mu_2 - \mu_1)^2]}{[2(\sigma_2^2 \vee \sigma_1^2) + (\mu_2 - \mu_1)^2/4]^{3/2}} \geq \sqrt{n}h \cdot |\mu_2 - \mu_1| \cdot \frac{2(\sigma_2^2 \vee \sigma_1^2)}{[2(\sigma_2^2 \vee \sigma_1^2) + (\mu_2 - \mu_1)^2/4]^{3/2}} > 0.$$  

(19)

In the second inequality we omitted $(\mu_2 - \mu_1)^2 \geq 0$ and used that $2\sigma_1^2 < 2^{-1}[\sigma_2^2 + 3\sigma_1^2]$ iff $\sigma_1^2 < \sigma_2^2$. This is the lower bound in (19). For the upper bound we find the numerator $\leq 2^{-1}[\sigma_2^2 + 3\sigma_1^2 + (\mu_2 - \mu_1)^2] \vee 2\sigma_1^2 \leq 2(\sigma_2^2 \vee \sigma_1^2) + (\mu_2 - \mu_1)^2$, and for the denominator we mention $(\sigma_2^2 - \sigma_1^2)x + 2\sigma_1^2 \geq 2(\sigma_2^2 \wedge \sigma_1^2)$ such that

$$|f'(x)| \leq \sqrt{n}h \cdot |\mu_2 - \mu_1| \cdot \frac{2(\sigma_2^2 \vee \sigma_1^2) + (\mu_2 - \mu_1)^2}{[2(\sigma_2^2 \wedge \sigma_1^2)]^{3/2}},$$

which completes (19).

\[ \square \]

**Proof of Proposition 3.4** Donsker’s theorem yields in $(\mathcal{D}_R[0,T], \| \cdot \|_\infty)$ as $n \to \infty$

$$\left[ \frac{1}{\sigma \sqrt{n}} \cdot \sum_{i=1}^{[nt]} (X_i - \mu) \right]_{\cdot} \overset{d}{\rightarrow} (W_t)_{\cdot}, \quad (20)$$

Define a continuous map $\varphi$ from $(\mathcal{D}_R[0,T], \| \cdot \|_\infty)$ to $(\mathcal{D}_R[\Delta s], \| \cdot \|_\infty)$ via

$$\varphi : (f(t))_{\cdot} \rightarrow \left( \frac{[f(t+h) - f(t)] - [f(t) - f(t-h)]}{\sqrt{2h}} \right)_{\cdot},$$

and apply $\varphi$ on (20). The continuous mapping yields in $(\mathcal{D}_R[\Delta s], \| \cdot \|_\infty)$ as $n \to \infty$

$$\left( \frac{1}{(2\sigma^2 nh)^{1/2}} \left[ \sum_{i=[nt]}^{[nt+h]} X_i - \sum_{i=[nt-h]}^{[nt]} X_i \right] \right)_{\cdot} \overset{d}{\rightarrow} (L_{t,\cdot})_{\cdot},$$

where the constant $\mu$ vanishes. Now by replacing $2\sigma^2$ by $\hat{\sigma}_1^2 + \hat{\sigma}_2^2$ and using Lemma 3.1 and Slutsky’s theorem weak convergence of $(D_{t,h}^{(n)})_{(t,h)\in\Delta s}$ follows.

\[ \square \]

**Proof of Lemma 4.2** Within $A_u$ the slope of $(n^{-1/2} \cdot |d_{t,h}^{(n)}|)_{\cdot}$ is positive left of $c_u$ and negative right of it, see Lemma 3.3. Thus, $|t_s(k) - c_u| = |t_s - c_u| - (k+1) \Leftrightarrow |t_s - c_u| \geq k + 1$ and (12) holds true. The path ends after $h - \delta$ steps, i.e., $t_s = c_u \Leftrightarrow |t_s - c_u| \leq h - \delta + 1 \Leftrightarrow (t_s, h_s) \in B_u$. Further, $(t_s, h_s) \in A_u$ implies $|t_s - c_u| \leq h_s$ (double window overlaps $c_u$), and thus $(t_s, h_s) \in A_u \setminus B_u$ implies $|t_s - c_u| \in \{h_s - \delta + 2, \ldots, h_s\}$, hence $|t_s - c_u| = |t_s - c_u| - (h_s - \delta + 1) \in \{1, \ldots, \delta\}$.

\[ \square \]

**Proof of Proposition 4.3** From Corollary 3.2 we obtain a.s. as $n \to \infty$

$$\sup_{(t,h)\in\Delta s} \left| n^{-1/2} \cdot D_{t,h}^{(n)} - |d_{t,h}^{(1)}| \right| \longrightarrow 0.$$  

(21)

On $A_u$ (excluding $t = c_u$) a lower bound for the derivative of $(|d_{t,h}^{(1)}|)_{\cdot}$ is given through $\kappa_u \delta^{1/2} > 0$, see Lemma 3.3. Thus, for any $\varepsilon > 0$ it is

$$|d_{t,h}^{(1)} - |d_{t,\cdot,h}^{(1)}| \geq \kappa_u \delta^{1/2} \varepsilon, \quad \text{if } t \leq c_u, \quad \text{and} \quad |d_{t,h}^{(1)} - |d_{t+\varepsilon,h}^{(1)}| \geq \kappa_u \delta^{1/2} \varepsilon, \quad \text{if } t > c_u.$$  

(22)
provided that \((t, h), (t - \varepsilon, h), (t + \varepsilon, h)\) lie in \(A_u\). Consider the initializing step of the path, \(k = 0\). Both \(\hat{t}_s^{(n)}(k)\) and \(t_s(k)\) take a value in \(\{t_s - 1, t_s, t_s + 1\} \cap \Delta_s\). As \(\hat{t}_s^{(n)}(k)\) is a maximizer defined via \(|D_{t,h}^{(n)}|\) we obtain \(|n^{-1/2} \cdot D_{\hat{t}_s^{(n)}(k), h-k}^{(n)}| \geq |n^{-1/2} \cdot D_{t_s(k), h-k}^{(n)}| \) a.s. as \(n \to \infty\), where the convergence follows from \([21]\), i.e.,

\[
|n^{-1/2} \cdot D_{\hat{t}_s^{(n)}(k), h-k}^{(n)}| \geq |d_{t_s(k), h-k}^{(1)}| + o_a.s.(1). \tag{23}
\]

From this we bound

\[
|d_{t_s(k), h-k}^{(1)}| - |d_{\hat{t}_s^{(n)}(k), h-k}^{(1)}| + o_a.s.(1)
\leq |n^{-1/2} \cdot D_{\hat{t}_s^{(n)}(k), h-k}^{(n)}| - |d_{\hat{t}_s^{(n)}(k), h-k}^{(1)}| + o_a.s.(1)
\leq \left| n^{-1/2} \cdot D_{\hat{t}_s^{(n)}(k), h-k}^{(n)} - d_{\hat{t}_s^{(n)}(k), h-k}^{(1)} \right| + o_a.s.(1)
\leq \sup_{(t,h) \in \Delta_s} \left| n^{-1/2} \cdot D_{t,h}^{(n)} - d_{t,h}^{(1)} \right| + o_a.s.(1) \xrightarrow{a.s.} 0. \tag{24}
\]

In the first inequality we used \([23]\), in the second the triangle inequality, and the convergence follows from \([21]\). The estimator \(\hat{t}_s^{(n)}(k)\) is now tied to the nonrandom function via \(d_{\hat{t}_s^{(n)}(k), h-k}^{(1)}\), and \([22]\) brings us from the function to the estimator. Let \(\varepsilon > 0\). Then almost everywhere it holds that if \(|t_s(k) - \hat{t}_s^{(n)}(k)| > \varepsilon\) i.o., thus \(|d_{t_s(k), h-k}^{(1)} - d_{\hat{t}_s^{(n)}(k), h-k}^{(1)}| > \varepsilon \kappa_\delta \delta^{1/2}\) i.o. But as the latter only occurs finitely often it is \(t_s(k) = \hat{t}_s^{(n)}(k)\) a.s. for \(n\) large. Iteratively, this extends for all \(k = 1, 2, ..., h - \delta\), i.e., eventually the paths w.r.t. \(D_{t,h}^{(n)}\) and \(d_{t,h}^{(1)}\) coincide a.s.\]

**Proof of Lemma 4.4** From \(\delta < \lfloor \delta_C / 2 \rfloor\) conclude that \(A_u\) contains a square \(S\) with horizontal and vertical edges of length \(\lfloor \delta_C / 2 \rfloor\): choose the center of \(S\) as \((c_u, \delta_C)\). In fact, the right corners of \(S\) may only be adjacent to \(A_u\). Nevertheless, \(S \cap S_{\lfloor \delta_C / 2 \rfloor} \neq \emptyset\), and thus \(S \cap S_g \neq \emptyset\) for \(g \leq \lfloor \delta_C / 2 \rfloor\). \[
\]

**Proof of Theorem 4.6** For \((t_s, h_s) \in S\) set \(\hat{t}_s^{(n)} := \hat{t}_s^{(n)}\), making the relation to \((t_s, h_s) \in S\) explicit. Let \(C \neq \emptyset\). Then a.s. for \(n\) large enough it is

\[
\max_{(t_s, h_s) \in S \setminus R} \left( \min\{|\hat{t}_s^{(n)} - c_u| : c_u \in C\} \right) \leq \delta - 1, \tag{25}
\]

i.e., all potential paths that start in the ‘upper part’ \(S \setminus R\) of \(\Delta_s\) end close to a \(c_u \in C\). \([25]\) holds true, as for any \((t_s, h_s) \in S \setminus R\) the path eventually enters an \(A_u\) and then Proposition 4.3 states a.s. \(|\hat{t}_s^{(n)} - c_u| \leq \delta - 1\) for \(n\) large. The maximum follows from \(|S|\) being finite. Further, a.s. for \(n\) large it is

\[
\min_{(t_s, h_s) \in \bigcup_{u=1, \ldots, |C|} A_u} |(nh)^{-1/2} \cdot |D_{t,h}^{(n)}|| > \max_{(t_s, h_s) \in R} |(nh)^{-1/2} \cdot |D_{t,h}^{(n)}||, \tag{26}
\]

i.e., as long as the algorithm does not break, starting points are first chosen from \(S \setminus R \supset \bigcup_u A_u\) up until all of them are cut out, and after that they are chosen from \(R\). \([26]\) follows from Corollary 3.2 and Lemma 3.3 noting that \(d_{t,h}^{(1)} = 0\) if \((t, h) \in R\), and \(|d_{t,h}^{(1)}| > 0\) if \((t, h) \in A_u\). Combining \([25]\) and \([26]\), it follows that a.s. for \(n\) large at first all \(c \in C\) are estimated
up to a distance $\delta - 1$ from starting values within $S \setminus R$, given the algorithm does not break. Possibly step 3a is applied in between. We need to show that the breaking criterion 3b applies appropriately: First, for $m \leq |C|$ we show that it does not apply. A path starting in $S \setminus R$ must pass an $A_u$ such that a.s. for $n$ large

$$\min_{(t_s, h_s) \in S \setminus R} \left[ \max_{k=0, \ldots, h_s - \delta} \left| D^{(n)}_{t_s} (k), h_s - k \right| \right] \geq \min_{(t, h) \in \bigcup A_u} |D^{(n)}_{t, h}| > n^\beta,$$

i.e., the maximum w.r.t. any path starting in $S \setminus R$ exceeds the minimum w.r.t. to all $A_u$, and the second inequality holds as $D^{(n)}_{t, h} = n^{1/2}d^{(1)}_{t, h} + o_a.s.(n^{1/2})$ uniformly on $\Delta_\delta$ (Corollary 3.2) and $|d^{(1)}_{t, h}| > 0$ within all $A_u$ (Lemma 3.3), noting that $\beta < 1/2$. Thus, a.s. for $n$ large there is no break.

Second, for $m = |C| + 1$ the algorithm now breaks. All $c \in C$ are estimated. Thus, possibly after applying criterion 3a, all remaining starting points lie in $R$. This also covers $C = \emptyset$. Note that a path that starts in $R$ remains in $R$. It holds a.s. for $n$ large

$$\max_{(t_s, h_s) \in R} \left[ \max_{k=0, \ldots, h_s - \delta} \left| D^{(n)}_{t_s} (k), h_s - k \right| \right] \leq \max_{(t, h) \in R} |D^{(n)}_{t, h}| < n^{1/2 - v},$$

i.e., the maximum w.r.t. all paths starting in $R$ deceases the maximum w.r.t the entire $R$. For the second inequality we note that within $R$ it is $d^{(1)}_{t, h} = 0$ and thus $D^{(n)}_{t, h} = o_a.s.(n^{1/2 - v})$ uniformly on $R$ (Corollary 3.2). As $\beta > 1/2 - v$, a.s. for $n$ large the algorithm breaks.

Finally note that 3c. is asymptotically redundant: a.s. for $n$ large, for $m \leq |C|$ we stated correct estimation up to the error, so neighboring estimates will have at least distance $\delta_C - 2(\delta - 1)$, and for $m = |C| + 1$ the algorithm already broke in 3b.

\[\square\]

References

Antoch, J. and Hušková, M. (1999). Estimators of changes. In Asymptotics, nonparametrics, and times series, volume 158 of Statist. Textbooks Monogr., pages 533–577. Dekker, New York.

Aston, J. A. D. and Kirch, C. (2012). Evaluating stationarity via change-point alternatives with applications to fmri data. Ann. Appl. Stat., 6(4):1906–1948.

Aue, A. and Horváth, L. (2013). Structural breaks in time series. J. Time Ser. Anal., 34(1):1–16.

Baranowski, R., Chen, Y., and Fryzlewicz, P. (2019a). Narrowest-over-threshold detection of multiple change-points and change-point-like features. J. R. Stat. Soc., B: Stat. Methodol., 81:649–672.

Baranowski, R., Chen, Y., and Fryzlewicz, P. (2019b). not: Narrowest-Over-Threshold Change-Point Detection. R package version 1.2.

Baranowski, R. and Fryzlewicz, P. (2019). wbs: Wild Binary Segmentation for Multiple Change-Point Detection. R package version 1.4.

Berkes, I., Horváth, L., Kokoszka, P., and Shao, Q.-M. (2006). On discriminating between long-range dependence and changes in mean. Ann. Statist., 34(3):1140–1165.
Broder, B. (2017). Change-point analysis in nonstationary stochastic models. CRC Press, Boca Raton, FL.

Chen, J. and Gupta, A. K. (2000). Parametric statistical change point analysis. Birkhäuser Boston, Inc., Boston, MA.

Cho, H. and Kirch, C. (2022). Two-stage data segmentation permitting multiscale change points, heavy tails and dependence. Ann Inst Stat Math, 74(4):653–684.

Chu, C.-S. J., Hornik, K., and Kuan, C.-M. (1995). MOSUM tests for parameter constancy. Biometrika, 82(3):603–617.

Csorgő, M. and Horváth, L. (1997). Limit theorems in change-point analysis. Wiley Series in Probability and Statistics. John Wiley & Sons, Ltd., Chichester. With a foreword by David Kendall.

Dehling, H., Rooch, A., and Taqqu, M. S. (2017). Power of change-point tests for long-range dependent data. Electron. J. Stat., 11(1):2168–2198.

Dette, H., Eckle, T., and Vetter, M. (2020). Multiscale change point detection for dependent data. Scand. J. Stat., 47(4):1243–1274.

Döring, M. (2010). Multiple change-point estimation with U-statistics. J. Statist. Plann. Inference, 140(7):2003–2017.

Eichinger, B. and Kirch, C. (2018). A mosum procedure for the estimation of multiple random change points. Bernoulli, 24(1):526–564.

Fang, X., Li, J., and Siegmund, D. (2020). Segmentation and estimation of change-point models: false positive control and confidence regions. Ann. Statist., 48(3):1615–1647.

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

Fryzlewicz, P. (2018a). Supplement to "tail-greedy bottom-up data decompositions and fast multiple change-point detection".

Fryzlewicz, P. (2018b). Tail-greedy bottom-up data decompositions and fast multiple change-point detection. Ann. Statist., 46(6B):3390–3421.

Gombay, E. and Horváth, L. (1994). An application of the maximum likelihood test to the change-point problem. Stochastic Process. Appl., 50(1):161–171.

Gombay, E. and Horváth, L. (2002). Rates of convergence for U-statistic processes and their bootstrapped versions. volume 102, pages 247–272. Silver jubilee issue.

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

Hinkley, D. V. (1971). Inference about the change-point from cumulative sum tests. Biometrika, 58(3):509–523.

Holmes, M., Kojadinovic, I., and Quessy, J.-F. (2013). Nonparametric tests for change-point detection à la Gombay and Horváth. J. Multiv. Anal., 115:16–32.

Horváth, L. and Hušková, M. (2005). Testing for changes using permutations of U-statistics. J. Statist. Plann. Inference, 128(2):351–371.
Horváth, L. and Shao, Q.-M. (2007). Limit theorems for permutations of empirical processes with applications to change point analysis. *Stochastic Process. Appl.*, 117(12):1870–1888.

Hušková, M. and Slabý, A. (2001). Permutation tests for multiple changes. *Kybernetika*, 37(5):605–622.

Kass-Hout, T., Xu, Z., Memmurray, P., Park, S., Buckeridge, D., Brownstein, J., Finelli, L., and Groseclose, S. (2012). Application of change point analysis to daily influenza-like illness emergency department visits. *JAMIA*, 19:1075–81.

Killick, R., Eckley, I. A., Ewans, K., and Jonathan, P. (2010). Detection of changes in variance of oceanographic time-series using changepoint analysis. *Ocean Eng.*, 37(13):1120–1126.

Killick, R., Haynes, K., and Eckley, I. A. (2016). *changepoint*: An R package for changepoint analysis. R package version 2.2.2.

Lavielle, M. and Moulines, E. (2000). Least-squares estimation of an unknown number of shifts in a time series. *J. Time Ser. Anal.*, 21(1):33–59.

Levajković, T. and Messer, M. (2021). *mscp*: Multiscale Change Point Detection via Gradual Bandwidth Adjustment in Moving Sum Processes. R package version 1.0.

Li, H. and Sieling, H. (2017). *FDRSeg*: FDR-Control in Multiscale Change-Point Segmentation. R package version 1.0-3.

Matteson, D. S. and James, N. A. (2014). A nonparametric approach for multiple change point analysis of multivariate data. *J. Am. Stat. Assoc.*, 109(505):334–345.

Meier, A., Kirch, C., and Cho, H. (2021). *mosum*: A package for moving sums in change-point analysis. *J. Stat. Softw.*, 97(8):1–42.

Messer, M. (2022). Bivariate change point detection: joint detection of changes in expectation and variance. *Scand. J. Stat.*, 49:886–916.

Messer, M., Kirchner, M., Schiemann, J., Roeper, J., Neininger, R., and Schneider, G. (2014). A multiple filter test for the detection of rate changes in renewal processes with varying variance. *Ann. Appl. Stat.*, 8(4):2027–2067.

Muggeo, V. M. (2020). *cumSeg*: Change Point Detection in Genomic Sequences. R package version 1.3.

Nurk, S., Koren, S., Rhie, A., Rautiainen, M., Bzikadze, A., Mikheenko, A., Vollger, M., Altemose, N., Uralsky, L., Gershman, A., Aganezov, S., Hoyt, S., Diekhans, M., Logsdon, G., Alonge, M., Antonarakis, S., Borchers, M., Bouffard, G., Brooks, S., Caldas, G., Chen, N., Cheng, H., Chin, C., Chow, W., de Lima, L., Dishuck, P., Durbin, R., Dvorkina, T., Fiddes, I., Formenti, G., Fulton, R., Fungtammasan, A., Garrison, E., Grady, P., Graves-Lindsay, T., Hall, I., N.F., H., Hartley, G., Haukness, M., Howe, K., Hunkapiller, M., Jain, C., Jain, M., Jarvis, E., Kerpedjiev, P., Kirsch, M., Kolmogorov, M., Koralch, J., Kremitzki, M., Li, H., Maduro, V., Marshall, T., McCartney, A., McDaniel, J., Miller, D., Mullkin, J., Myers, E., Olson, N., Paten, B., Polusko, P., Pevzner, P., Porubsky, D., Potapova, T., Rogaev, E., Rosenfeld, J., Salzberg, S., Schneider, V., Sedlazeck, F., Shafin, K., Shew, C., Shumate, A., Sims, Y., Smrt, A., Soto, D., Sölvé, I., Storer, J., Streets, A., Sullivan, B., Thibaud-Nissen, F., Torrance, J., Wagner, J., Walenz, B., Wenger, A., Wood, J., Xiao, C., Yan, S., Young, A., Zarate, S., Surti, U., McCoy, R., Dennis, M., Alexandrov, I., Gerton, J., O’Neill, R., Timp, W., Zook, J., Schatz, M., Eichler, E., Miga, K., and Phillippy, A. (2022). The complete sequence of a human genome. *Science*, 376:44–53.
Page, E. S. (1954). Continuous inspection schemes. Biometrika, 41(1-2):100–115.

Pein, F., Hotz, T., Sieling, H., and Aspelmeier, T. (2020). stepR: Multiscale change-point inference. R package version 2.1-1.

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

Reeves, J., Chen, J., Wang, X. L., Lund, R., and Lu, Q. Q. (01 Jun. 2007). A review and comparison of changepoint detection techniques for climate data. JAMC, 46(6):900–915.

Rybach, D., Gollan, C., Schluter, R., and Ney, H. (2009). Audio segmentation for speech recognition using segment features. In 2009 IEEE International Conference on Acoustics, Speech and Signal Processing, pages 4197–4200.

Spokoiny, V. (2009). Multiscale local change point detection with applications to value-at-risk. Ann. Statist., 37(3):1405–1436.

Steinebach, J. and Eastwood, V. R. (1995). On extreme value asymptotics for increments of renewal processes. volume 45, pages 301–312. Extreme value theory and applications (Villeneuve d’Ascq, 1992).

Contact information

Tijana Levajković
email: tijana.levajkovic@tuwien.ac.at
Michael Messer
email: michael.messer@tuwien.ac.at

Vienna University of Technology
Institute of Statistics and Mathematical Methods in Economics
Wiedner Hauptstraße 8-10/105
1040 Vienna, Austria