Exact tests for offline changepoint detection in multichannel binary and count data with application to networks

Shyamal K. De and Soumendu Sundar Mukherjee

ABSTRACT
We consider offline detection of a single changepoint in binary and count time-series. We compare exact tests based on the cumulative sum (CUSUM) and the likelihood ratio (LR) statistics, and a new proposal that combines exact two-sample conditional tests with multiplicity correction, against standard asymptotic tests based on the Brownian bridge approximation to the CUSUM statistic. We see empirically that the exact tests are much more powerful in situations where normal approximations driving asymptotic tests are not trustworthy: (i) small sample settings; (ii) sparse parametric settings; (iii) time-series with changepoint near the boundary. We also compare these exact tests against their bootstrap-based variants which do not have exact type I error control guarantees. Further, we consider a multichannel version of the problem, where channels can have different changepoints. Controlling the False Discovery Rate (FDR), we simultaneously detect changes in multiple channels. This ‘local’ approach is shown to be more advantageous than multivariate global testing approaches when the number of channels with changepoints is much smaller than the total number of channels. As a natural application, we consider network-valued time-series and use our approach with (a) edges as binary channels and (b) node-degrees or other local subgraph statistics as count channels. The local testing approach is seen to be much more informative than global network changepoint algorithms.

1. Introduction

Changepoint analysis is an important problem in statistics with roots in statistical quality control [1–3]. The goal of changepoint analysis is to decide if there are distributional changes in a given time-series (the detection part), and estimate the change if any (the estimation part). There is a huge body of literature on the univariate changepoint problem. An excellent treatment can be found in the book [4].
Some notable works on the multivariate version of the problem are [5–8] in parametric settings, and [9–11] in non-parametric settings.

We should mention that there are two types of changepoint problems: (a) offline, where the whole time-series is available to the statistician; (b) online, where data is still arriving at the time of analysis. We will be concerned with the offline problem in this article.

Although a lot of work has been done on changepoint detection for continuous time-series data, results for discrete data are lacking, especially in ‘small sample’ settings where the length of the time-series is relatively small. In this article, our main goal is to develop methods for offline changepoint detection for binary and count data that have good performance in small sample settings.

We adapt well-known conditional two-sample tests for binary and count data to the changepoint setup using a multiple testing approach. We also consider exact tests based on natural statistics such as the CUSUM statistic and the LR statistic. We conduct a comprehensive small-sample power analysis of these tests and compare them against the large sample CUSUM test based on a Brownian bridge approximation [4]. We find that, in small sample scenarios, and in cases where the true changepoint lies near the boundary, the exact tests are significantly more powerful than large sample tests.

Although these methods are developed for single changepoint problems, they seem to work well when multiple changepoints are present, especially if there is one strong change. We report some empirical findings in this direction in the supplementary article [12].

We then consider multichannel binary or count time-series. Using a False Discovery Rate (FDR) controlling mechanism, we simultaneously test for changepoints in all the channels. This ‘local’ approach vastly outperforms the ‘global’ approach of using some statistic of all channels together (e.g. a vector CUSUM statistic), when the number of channels with changepoints is much smaller than the total number of channels.

As an application of this approach, we consider local vs. global testing in network-valued time-series. Although there has been a recent surge of interest in network changepoints [13–20], we note that the existing works are focused on large sample asymptotics and use global statistics for detection or estimation. If we use edges (resp. node degrees or some other local subgraph statistics such as local triangle counts) as separate channels, then we have multichannel binary (resp. count) data. We compare the proposed local approach against a standard CUSUM-based global approach in real-world networks. We see that, in addition to picking up strong global changes, the local approach can identify relatively weak and rare changes.

The rest of the paper is organized as follows. In Section 2, we describe the problem set-up precisely and detail our methodology. In Section 3, we discuss our multiple testing based local approach for multichannel changepoint detection. In Section 4, we report our simulations: In Section 4.1, we perform a comprehensive power analysis of the various proposed methods against existing approaches. In Section 4.3, we compare the local testing approach vs. multivariate CUSUM-based global testing approaches in multichannel problems. More detailed results are provided in the supplementary article [12]. Then, in Section 5, we apply our methodology on two real-life examples: a time-series of US senate voting pattern networks, and another time-series of phone-call networks. We conclude the paper with a discussion in Section 6.
2. Set-up and methodology

Suppose that we have time-indexed independent variables $X_1, \ldots, X_T$, with $E(X_i) = \pi_i$. We want to test if the $\pi_i$ have changed over time. The single changepoint testing problem is:

$$H_0 : \pi_i = \pi \quad \text{for all } 1 \leq i \leq T \text{ (no change)}$$

vs.

$$H_1 : \exists 1 \leq \tau \leq T - 1 \text{ such that } \pi_i = \pi_1 \mathbb{I}_{[i \leq \tau]} + \pi_2 \mathbb{I}_{[i > \tau]} \quad \text{(at least one change)},$$

where $\mathbb{I}_A$ stands for the indicator of event $A$.

We are interested in the situation where the $X_i$'s are binary or counts. The binary case is obviously modelled by an independent Bernoulli time-series, whereas we model count data using the Poisson distribution. Keeping that in mind, let us now discuss some natural test statistics for the testing problem (1). All the tests considered below use the cumulative sums $S_t = \sum_{i=1}^t X_i, t = 1, \ldots, T$. We begin by deriving the likelihood ratio statistic.

The Likelihood Ratio (LR) statistic.

Binary data: Note that under $H_1$, the likelihood of the data is

$$L(\pi_1, \pi_2, \tau) = \pi_1^{S_\tau} (1 - \pi_1)^{\tau - S_\tau} \times \pi_2^{S_t - S_\tau} (1 - \pi_2)^{T - \tau - (S_T - S_\tau)}.$$

The maximizers of $L$ for a fixed $\tau$ are $\hat{\pi}_1 = \frac{S_\tau}{\tau}$ and $\hat{\pi}_2 = \frac{S_T - S_\tau}{T - \tau}$. Thus the profile log-likelihood for $\tau$ is

$$\ell_{PL}(\tau) = -\tau H(\hat{\pi}_1) - (T - \tau) H(\hat{\pi}_2),$$

where $H(x) = -x \log x - (1 - x) \log(1 - x)$ is the entropy of a Ber$(x)$ variable. Define

$$T_b = \min_{1 \leq t \leq T - 1} \left[ t H\left( \frac{S_t}{t} \right) + (T - t) H\left( \frac{S_T - S_t}{T - t} \right) \right].$$

Then the log-LR statistic is

$$T_{LR}^{(b)} = -2(\ell_0 - \ell_1) = -2(-TH(S_T/T) + T_b).$$

We would reject $H_0$ for large values of this statistic.

Count data: Recall that we are modelling counts using the Poisson distribution. Under $H_1$, the likelihood of the data is

$$L(\pi_1, \pi_2, \tau) \propto e^{-\tau \pi_1} \pi_1^{S_\tau} e^{-(T - \tau) \pi_2} \pi_2^{S_T - S_\tau}.$$

The maximizers of $L$ for a fixed $\tau$ are $\hat{\pi}_1 = \frac{S_\tau}{\tau}$ and $\hat{\pi}_2 = \frac{S_T - S_\tau}{T - \tau}$. Therefore the profile log-likelihood for $\tau$ is

$$\ell_{PL}(\tau) = -\tau G(\hat{\pi}_1) - (T - \tau) G(\hat{\pi}_2),$$

where $G(x) = x(1 - \log x)$. Define

$$T_c = \min_{1 \leq t \leq T - 1} \left[ t G\left( \frac{S_t}{t} \right) + (T - t) G\left( \frac{S_T - S_t}{T - t} \right) \right].$$
Then the log-LR statistic is
\[ T_{LR}^{(c)} = -2(\ell_0 - \ell_1) = -2(-TG(S_T/T) + T_c). \] (3)

We would reject \( H_0 \) for large values of this statistic.

**CUSUM statistic.**

A well-known and often-used statistic in changepoint problems is the so-called CUSUM statistic. For \( 0 < a < b < 1 \), suppose \( aT \) and \( bT \) are known upper and lower bounds on the locations of the potential changepoints. For \( 0 \leq \delta \leq 1 \), the CUSUM statistic is defined as
\[ T_{CUSUM}^{(\delta)} = \max_{aT \leq t \leq bT} \left[ \frac{t}{T} \left( 1 - \frac{t}{T} \right) \right]^\delta \left| \frac{S_t - S_T - S_t}{T - t} \right|. \] (4)

This can be used with both binary and count data.

### 2.1. Asymptotic tests

First, we will consider an asymptotic test based on the CUSUM statistic (4). The asymptotic null distribution can be calculated using a Brownian bridge approximation. For details see, e.g. [4].

**Proposition 2.1:** Let \( B_0(t) \) denote a standard Brownian bridge. Under \( H_0 \), \( \pi_i = \pi \) for all \( i \), and
\[ \frac{\sqrt{T} T_{CUSUM}^{(\delta)}}{\sigma} \xrightarrow{T \to \infty} M_{ab}^{(\delta)}, \]
where \( M_{ab}^{(\delta)} = \max_{a \leq t \leq b} \left\{ \frac{|B_0(t)|}{(t(1-t))^{1-\delta}} \right\} \) and \( \sigma^2 = \text{Var}(X_1) \). (Note that \( \sigma^2 \) equals \( \pi(1 - \pi) \) for Bernoulli data and \( \pi \) for Poisson data.)

**Corollary 2.2:** Let \( \hat{\pi} = \frac{1}{T} \sum_{s=1}^{T} X_s \xrightarrow{a.s.} \pi \) under \( H_0 \). Then, under \( H_0 \),
\[ \frac{\sqrt{T} T_{CUSUM}^{(\delta)}}{\hat{\sigma}} \xrightarrow{T \to \infty} M_{ab}^{(\delta)}, \]
where \( \hat{\sigma}^2 \) equals \( \hat{\pi}(1 - \hat{\pi}) \) for Bernoulli data and \( \hat{\pi} \) for Poisson data.

Using this result, we can perform an asymptotic test for \( H_0 \) in the ‘large sample’ regime where \( T \) is large. This test would be good when \( \pi \) is not too small (so that the underlying normal approximations to the partial sums \( \sum_{s=1}^{t} X_s \) go through). It is well-known that in this asymptotic framework the choice \( \delta = 1/2 \) is the best for estimation (See, e.g. [4, Chapter 3]), while \( \delta = 1 \) is the best for minimizing type-1 error, \( \delta = 0 \) for minimizing type-2 error. However, in the small sample situations, explored in this paper, we do not see such a clear-cut distinction (see Section 4).
2.2. Conditional tests

Our exact tests are based on the following simple lemma.

**Lemma 2.3:** Suppose $X_1, \ldots, X_T$ are independent $\text{Bin}(n, \pi)$ (or $\text{Poisson}(\pi)$). Let $S_i = \sum_{j=1}^i X_j$. Then the joint distribution of $(S_1, \ldots, S_{T-1})$ given $S_T$ does not depend on $\pi$.

**Proof:** Since $S_T$ is sufficient for $\pi$, the distribution of $(X_1, \ldots, X_T)$ given $S_T$ does not depend on $\pi$. Therefore we can perform an exact conditional test using $S_T$. More elaborately, we can simulate the exact null distribution of $T$ given $S_T$ via Monte Carlo and then estimate the $p$-value $P_{H_0}(T \geq T_{obs} | S_T)$, where $T_{obs}$ is the observed value of $T$. We also note here that instead of the statistic $T^{(b)}_{LR}$ (or $T^{(c)}_{LR}$), one can equivalently use $T^{(b)}$ (resp. $T^{(c)}$), which is what we do in our implementation.

**Approach 2.1:** Let $T$ be any one among the statistics $T^{(b)}_{LR}$, $T^{(c)}_{LR}$, or $T^{(d)}_{\text{CUSUM}}$. As $T$ is a function of the partial sums $(S_i)_{i=1}^T$ only, by Lemma 2.3, the distribution of $T$ given $S_T$ does not depend on $\pi$ under $H_0$. Therefore we can perform an exact conditional test using $T$. More elaborately, we can simulate the exact null distribution of $T$ given $S_T$ via Monte Carlo and then estimate the $p$-value $P_{H_0}(T \geq T_{obs} | S_T)$, where $T_{obs}$ is the observed value of $T$. We also note here that instead of the statistic $T^{(b)}_{LR}$ (or $T^{(c)}_{LR}$), one can equivalently use $T^{(b)}$ (resp. $T^{(c)}$), which is what we do in our implementation.

**Approach 2.2:** Note that we can decompose $H_1$ as a disjoint union of the following $(T-1)$ hypotheses:

$$H_{1i} : \tau = i, \quad 1 \leq i \leq T - 1,$$

and test these separately against $H_0$, and, finally, rejecting $H_0$ if one of these $T - 1$ hypotheses gets rejected. As testing $H_{1i}$ against $H_0$ is basically two-sample testing, we can use the well-known conditional tests for comparing two binomial/Poisson populations (see, for example, Section 4.5 of [21]).

**Binary data:** Suppose $X_i \sim \text{Ber}(\pi_i)$. Note that if we use $S_i = \sum_{j=1}^i X_j$ as a test statistic for testing $H_0$ against $H_{1i}$, then, under $H_0$,

$$S_i \mid S_T \sim \text{Hypergeometric}(i, S_T, T).$$

Therefore, we get a $p$-value $p_i$ from this conditional distribution as

$$p_i = \sum_{q : f(q;i,S_T,T) \leq f(S_i;i,S_T,T)} f(q;i,S_T,T),$$

where $f(q;i,S_T,T)$ is the PMF of the Hypergeometric($i, S_T, T$) distribution.

**Count data:** For count data $X_i \sim \text{Poisson}(\pi_i)$, we can use the same procedure as above using the observation that, under $H_0$,

$$S_i \mid S_T \sim \text{Binomial} \left( S_T, \frac{i}{T} \right).$$

In this case, we get a $p$-value $p_i$ from the above conditional distribution as

$$p_i = \sum_{q : g(q+S_T,i/T) \leq g(S_i+S_T,i/T)} g(q;S_T,i/T),$$

where $g(q;S_T,i/T)$ is the PMF of the Binomial($S_T, i/T$) distribution.
Once we get hold of the individual \( p \)-values \( p_1, \ldots, p_{T-1} \), we can use some \textit{familywise error rate} (FWER) controlling test procedure to decide about \( H_0 \). It follows from Lemma 2.3 that the joint distribution of \( (p_1, \ldots, p_{T-1}) \) given \( S_T \) does not depend on \( \pi \) under \( H_0 \). Thus we can exactly simulate the null distribution of \( p_{(1)} \) given \( S_T \) via Monte Carlo to estimate the \( p \)-value \( P_{H_0}(p_{(1)} \leq p_{(1),\text{obs}} | S_T) \), where \( p_{(1),\text{obs}} \) is the observed value of \( p_{(1)} \).

### 2.3. Permutation tests

Since the \( X_i \)'s are exchangeable under the null model of no changepoints, it is natural to have recourse to permutation tests. In the Bernoulli case, however, note that \( S_T \) is permutation invariant and permutations of the data \( (X_1, \ldots, X_T) \) give all possible samples with a fixed \( S_T \). It follows that tests based on the conditional distribution of the data given \( S_T \) are the same as permutation tests.

With count data, however, permutation tests are different from conditional tests. One advantage permutation tests have over the conditional tests considered earlier is that we do not need the Poisson assumption – the permutation methodology gives us exact level \( \alpha \) tests under any distribution. However, when the underlying data is actually Poisson, permutation tests are more conservative than conditional tests – see Figure 3. We also see this in our local changepoint tests for network data, where the degree of a node is often well approximated by a Poisson distribution (see Section 5.2).

### 2.4. Bootstrap tests

One of the anonymous reviewers pointed out that bootstrap-based procedures have been used earlier in the changepoint literature (see, e.g., [22–24]). For continuous data, the Residual Bootstrap (RB) approach of [22] works well, but in our discrete set-up, RB-based tests have very poor performance. Instead, we use two other bootstrap procedures: (i) Direct Bootstrap (DB) and (ii) Parametric Bootstrap (PB). In both of these procedures, we first obtain \( B \) bootstrap samples \( \{(X_t^{(b)})_{t=1}^T : 1 \leq b \leq B\} \). In case of DB, for each \( b \), \( (X_t^{(b)})_{t=1}^T \) is obtained by resampling (with replacement) directly from the observed sample \( (X_t)_{t=1}^T \). On the other hand, in case of PB, we first estimate the parameter \( \pi \) (under the null hypothesis of no change), by its maximum likelihood estimate \( \hat{\pi}_{\text{MLE}} \), and then generate the bootstrap samples \( (X_t^{(b)})_{t=1}^T \) from the Bernoulli (for binary data) or Poisson (for count data) distribution with parameter \( \hat{\pi}_{\text{MLE}} \).

For each bootstrap sample \( (X_t^{(b)})_{t=1}^T \), we compute our test statistic of interest, and thus obtain \( B \) samples from its (approximate) null distribution, which is then used to compute \( p \)-values.

We note here that for Bernoulli data, the DB and PB tests are identical. Indeed, under the null hypothesis of no change, \( \hat{\pi}_{\text{MLE}} = \frac{1}{T} \sum_{t=1}^T X_t \). On the other hand, since there are \( T\hat{\pi}_{\text{MLE}} \) many 1’s in our sample, the probability that a with-replacement draw from it results in a 1 is exactly \( \hat{\pi}_{\text{MLE}} \). This implies that for any \( 1 \leq b \leq B \), the DB sample \( (X_t^{(b)})_{t=1}^T \) is an i.i.d. sample from the Bernoulli(\( \hat{\pi}_{\text{MLE}} \)) distribution, and hence a PB sample.

For count data, nevertheless, the two procedures are different. In fact, under DB, one cannot observe a value that is within the support of the underlying Poisson distribution but has not been realized in the sample \( (X_t)_{t=1}^T \). PB does not suffer from this constraint.
In Figures 4 and 5, we compare the bootstrap based-tests against the exact conditional tests derived earlier in terms of type I error control. We see that under the null hypothesis, in small sample settings, the $p$-values of the exact tests stochastically dominate those of the bootstrap-based tests. In moderate to large sample scenarios, unsurprisingly, the $p$-values behave more-or-less in the same manner for both types of tests. In terms of power, tests based on PB dominate the exact conditional tests in most cases, likely because of the inherent inability of the DB procedure itself to produce values not already in the observed sample. See Section 4.2 for more details.

### 2.5. Changepoint estimation

While we are interested in changepoint detection, the testing methods give bona-fide estimators of the underlying changepoint. For example, the likelihood ratio statistics are based on maximizing the profile log-likelihood $\ell_{PL}(\tau)$ and the maximizer gives an estimate of $\tau$. Similarly, for the CUSUM statistic, the maximizer in the definition gives one estimate. As for the conditional testing approach, the minimizing index of the individual $p$-values gives an estimate of the changepoint. One can show that, under a single changepoint model, these estimates are consistent, because all these objective functions are based on the cumulative average $S_t/t$, and one can use the fact that a properly rescaled version of this process converges to a Brownian motion under the null hypothesis of no changepoints. For example, an analysis of the CUSUM estimator along these lines can be found in [4]. In Section 5, we obtain channel-specific estimates of changepoints in this way and plot their histograms (see Figures 10, 12, and 13).

A statistically valid procedure for simultaneous detection and estimation may be obtained by using an even-odd sample splitting: separate out the observations with even and odd time indexes, use the even ones for testing, and based on the decision, use the odd one for further estimation. However, as with any sample splitting method, this method will suffer a loss in power in small sample scenarios.

### 3. The multichannel case: global vs. local testing

Suppose we observe an $m$-variate ($m > 1$) independent time-series

$$X_1, \ldots, X_\tau \overset{i.i.d.}{\sim} F_1, \quad X_{\tau+1}, \ldots, X_T \overset{i.i.d.}{\sim} F_2,$$

where $F_1$ and $F_2$ are $m$-variate distributions. We would like to test the global null $H_0$ : ‘no change in the $m$-variate time-series’, i.e. $H_0 : \tau = T$. Since permutations of $X_1, \ldots, X_T$ are equally likely under $H_0$, a natural approach for testing $H_0$ is to adopt a permutation test using the global CUSUM statistic

$$C^{(\delta)} = \max_{1 \leq t \leq T-1} \left[ \frac{t}{T} \left( 1 - \frac{t}{T} \right) \right]^{\delta} \left\| \frac{1}{t} \sum_{i=1}^{t} X_i - \frac{1}{T-t} \sum_{i=t+1}^{T} X_i \right\|$$

for $\delta \in [0, 1]$,
where $\| \cdot \|$ denotes any suitable norm on $\mathbb{R}^m$. Permuting the time-series multiple times, we construct a randomized size-$\alpha$ test that rejects $H_0$ for a large value of observed $C(\delta)$. However, this test cannot determine which channels were responsible for the global change.

Now suppose that $X_i = (X_{1,i}, \ldots, X_{m,i})$ for $i = 1, \ldots, T$, and the time-series for the $j$th channel is

$$X_{j,1}, \ldots, X_{j,\tau} \overset{i.i.d.}{\sim} F_{j,1}, \quad X_{j,\tau+1, \ldots, X_{j,T}} \overset{i.i.d.}{\sim} F_{j,2} \quad \text{for } j = 1, 2, \ldots, m.$$ 

In this article, $F_{j,1} \equiv \text{Ber}(p_{j,1})$ or $\text{Pois}(\lambda_{j,1})$ and $F_{j,2} \equiv \text{Ber}(p_{j,2})$ or $\text{Pois}(\lambda_{j,2})$ depending on whether we deal with binary or count data. The global null $H_0$ is true implies that $\bigcap_{j=1}^m H_{0,j}$ is true where $H_{0,j} : \text{‘no change in the } j \text{th channel’}$. A local approach for testing $H_0$ is to compute $p$-values corresponding to $H_{0,j}$ for $j = 1, \ldots, m$, and apply some suitable multiple testing procedure controlling FWER or FDR. Note that FDR equals FWER under $\bigcap_{j=1}^m H_{0,j}$. Since FDR controlling methods are known to be more powerful than traditional FWER controlling methods such as the Bonferroni and Holm procedures [25] when $m$ is large, we use some popular methods for FDR control such as the Benjamini-Hochberg (BH) step-up procedure [26], the adaptive Benjamini-Hochberg (ABH) procedure [27], and the adaptive Storey-Taylor-Siegmund (STS) procedure [28]. Although these methods require independence or positive regression dependence of the $p$-values to control FDR theoretically, we do not make any such assumptions. As long as a method controls FDR at level $\alpha$, it serves our purpose. Therefore, we refrain from recommending any particular FDR controlling mechanism.

Let $\mathcal{R} = \{1 \leq j \leq m : H_{0,j} \text{ is rejected}\}$ be the rejection set obtained from some FDR controlling procedure. The global null $H_0$ is rejected if and only if $\mathcal{R}$ is nonempty. This test for $H_0$ is referred to as a local test and denoted by $\phi := \mathbb{I}_{(\mathcal{R} \neq \emptyset)}$.

**Remark 3.1:** If FDR is controlled at level $\alpha$, then $\phi$ is a valid level-$\alpha$ test for the global null $H_0$ since $P_{H_0}(H_0 \text{ rejected}) = P_{H_0}(\mathcal{R} \neq \emptyset) = P_{H_0}(\bigcup_{j=1}^m H_{0,j} \text{ rejected}) = \text{FWER} = \text{FDR} \leq \alpha$.

**Remark 3.2:** The local testing approach enjoys a few advantages over the global testing approach. First, local testing is much more informative in the sense that channels responsible for the global change, if any, are also determined. Second, under the rare signal regime where signals are available only in a few out of a large number of channels, global tests may fail to detect a change whereas local tests are more likely to detect the change as they scrutinize all channels. These points are empirically demonstrated in the simulations of Section 4.3.

**Remark 3.3:** Although we have formulated the local testing approach for a single global changepoint so as to compare it to the global testing approach, it is clear that the former applies to situations where individual channels have different changepoints. This advantage of the local testing approach over the global testing approach will be clear in Sections 4.3 and 5, where we plot histograms of detected local changepoints.
4. Simulations

4.1. Exact vs. asymptotic tests in a single channel

We first compare the proposed exact level-$\alpha$ tests (discussed in Section 2.2) against the asymptotic level-$\alpha$ tests (discussed in Section 2.1) in single channel set-ups. Exact conditional tests based on the $T_{LR}(b)$ and $T_{LR}(c)$ test statistics are referred to as ‘LR’ tests, whereas those based on the CUSUM statistics $T_{CUSUM}(0.5)$ and $T_{CUSUM}(1)$ are referred to as ‘CU.5’ and ‘CU1’ tests, respectively. On the other hand, exact conditional tests based on the $p^{(1)} = \min_{1 \leq i \leq T-1} p_i$ statistic are referred to as ‘minP’ tests. We estimate the $p$-values of these tests using $M = 50,000$ Monte Carlo samples from their respective null distributions given $S_T$.

More elaborately, let $T$ denote any one of the test statistics $T_{LR}(b)$, $T_{LR}(c)$, $T_{CUSUM}(0.5)$, $T_{CUSUM}(1)$, or $p^{(1)}$. Note that larger values of $T$ provide evidence against $H_0$. We generate $M = 50,000$ Monte Carlo samples, say $T_1, \ldots, T_M$, from the null distribution of $T$ given $S_T$, and then estimate the $p$-value $P_{H_0}(T \geq T_{obs} | S_T)$ by $\frac{1}{M} \sum_{i=1}^{M} I(T_i \geq T_{obs})$, where $T_{obs}$ is the observed value of $T$.

Asymptotic tests based on Brownian bridge approximations (see Corollary 2.2) are considered for $\delta = 0.5$ and $\delta = 1$, and are referred to as ‘BB.5’ and ‘BB1’ tests, respectively.

Additionally, we consider two tests based on the functions `cpt.mean` and `cpt.meanvar` from the R package `changepoint`. Both of these functions are designed to estimate the number of changepoints in a univariate time-series. `cpt.mean` assumes that changes occur only in the mean of the series, whereas `cpt.meanvar` can handle changes in both the mean or the variance. The `cpt.mean` test (resp. the `cpt.meanvar` test) detects a change (i.e. rejects $H_0$) if the number of estimated changepoints by the `cpt.mean` (resp. `cpt.meanvar`) function is at least one. In the `cpt.mean` function, we use the ‘BinSeg’ method (i.e. binary segmentation [29]) along with the ‘CUSUM’ statistic, whereas in the `cpt.meanvar` function, we use the ‘BinSeg’ method along with the ‘Poisson’ statistic. In both of these functions the penalty values (‘pen.value’) are chosen manually such that the attained type I errors approximately equal $\alpha = 0.1$.

Figure 1 considers the Bernoulli case. We see that the exact conditional tests perform well in both sparse and dense situations and always outperform the `cpt.mean` test. The asymptotic tests (especially BB1) also provide reasonable power if the sample size $T$ is large and the changepoint $\tau$ is near the middle (Figure 1(a,c)). However, if the changepoint is closer to the boundary (Figure 1(b,d)), then the exact conditional tests minP and LR perform significantly better than the asymptotic tests.

Figure 2 considers the Poisson case. The proposed exact tests perform well even when the sample size is as small as $T = 10$, and, in this case, they uniformly outperform the asymptotic tests and the `cpt.meanvar` test. If the changepoint is close to the boundary, then the exact tests (especially minP and LR) yield much higher power than their competitors (see Figure 2(b,d)). For large sample sizes (e.g. $T = 50$), when the Brownian bridge approximations kick in, asymptotic tests become comparable to the exact tests in terms of performance.

In Figure 3, we compare permutation tests (discussed in Section 2.3) against the exact conditional tests for Poisson time-series. Permutation tests using the $T_{LR}(c)$, $T_{CUSUM}(0.5)$, $T_{CUSUM}(1)$ and $p^{(1)}$ statistics are referred to as ‘perm-LR’, ‘perm-CU.5’, ‘perm-CU1’ and
Figure 1. Comparison of change detection probabilities of exact conditional tests, asymptotic tests and the cptmean test with $\alpha = 0.1$ in the time-series $X_1, \ldots, X_\tau \sim \text{Ber}(\pi_1)$, $X_{\tau+1}, \ldots, X_T \sim \text{Ber}(\pi_2)$.

4.2. Exact vs bootstrap tests in a single channel

In this section, we compare the empirical distributions of $p$-values as well as the attained probabilities of change detection (i.e. power and attained level) for the exact conditional and bootstrap-based tests in single channel set-ups. We use $B = 1000$ bootstrap samples to compute $p$-values for the bootstrap-based tests and estimate the attained probabilities of change detection using 1000 Monte Carlo replications.

In Figure 4, we compare the empirical survival functions of the $p$-values of the exact conditional and PB-based tests for Bernoulli time-series using the $p_{(1)}, \mathcal{T}_{LR}^{(b)}, \mathcal{T}_{CUM}^{(0.5)}$ and $\mathcal{T}_{CUM}^{(1)}$ statistics. The empirical survival function $G_{\min P}(x)$ equals $1 - \hat{F}_{\min P}(x)$, where
Figure 2. Comparison of change detection probabilities of exact conditional tests, asymptotic tests and the \( \text{cptmeanvar} \) test with \( \alpha = 0.1 \) in the time-series: \( X_1, \ldots, X_T \overset{i.i.d.}{\sim} \text{Pois}(\lambda_1) \), \( X_{T+1}, \ldots, X_T \overset{i.i.d.}{\sim} \text{Pois}(\lambda_2) \).

\( \hat{F}_{\text{minP}}(x) \) denotes the empirical cumulative distribution function of the \( p \)-values of either the exact conditional or PB-based test using the \( p_{(1)} \) statistic. The empirical survival functions \( G_{LR}, G_{CU.5} \) and \( G_{CU1} \) corresponding to the \( T_{LR}^{(b)}, T_{\text{CUSUM}}^{(0.5)} \) and \( T_{\text{CUSUM}}^{(1)} \) statistics, respectively, are defined in a similar way. We observe from Figure 4 that the \( p \)-values of the exact conditional tests are stochastically larger than those of the PB-based tests. Also, the magnitude of their difference diminishes as the sample size \( T \) increases, as seen from the second row of Figure 4. We note here that unlike the PB-based tests, the exact conditional tests enjoy theoretical guarantees of type I error control. In fact, we empirically observe that the PB-based tests tend to yield slightly higher attained level than the desired level of significance when the sample size is small.

In Figure 5, we compare the empirical survival functions of the \( p \)-values of the exact conditional, DB-based and PB-based tests for Poisson time-series using the \( p_{(1)}, T_{LR}^{(c)}, T_{\text{CUSUM}}^{(0.5)} \) and \( T_{\text{CUSUM}}^{(1)} \) statistics. The empirical survival functions \( G_{\text{minP}}, G_{LR}, G_{CU.5} \) and \( G_{CU1} \) of
Figure 3. Comparison of change detection probabilities of exact conditional tests and permutation tests.

Figure 4. Comparison of the empirical survival functions of the p-values of the exact conditional and PB-based tests for Bernoulli time-series using the $p_{(1)}$, $T_{LR}^{(b)}$, $T_{CUSUM}^{(0.5)}$ and $T_{CUSUM}^{(1)}$ statistics. The solid red lines and long-dashed blue lines correspond to the exact conditional and PB-based tests, respectively. The p-values are obtained under the no-changepoint scenario where $X_1, \ldots, X_T \overset{i.i.d.}{\sim} \text{Ber}(0.5)$ with $T = 10$ in the first row and $T = 30$ in the second.

$p$-values are defined similarly to the binary case. Note that the $p$-values of the exact conditional tests are again stochastically larger than those of the PB-based tests, which in turn are stochastically larger than those of the DB-based tests. As the sample size $T$ increases, the magnitudes of the differences between these empirical survival functions diminish, as seen from the second row of Figure 5.

Table 1 provides a power comparison of the exact conditional and PB-based tests for independent Bernoulli time-series. PB-based tests using the $p_{(1)}$, $T_{LR}^{(b)}$, $T_{CUSUM}^{(0.5)}$ and $T_{CUSUM}^{(1)}$ statistics are referred to as ‘PB-minP’, ‘PB-LR’, ‘PB-CU.5’ and ‘PB-CU1’ tests,
Figure 5. Comparison of the empirical survival functions of the $p$-values of the exact conditional, DB-based and PB-based tests for Poisson time-series using the $P_{1(1)}$, $T_{LR}^{(c)}$, $T_{CUSUM}^{(0.5)}$ and $T_{CUSUM}^{(1)}$ statistics. The solid red lines, long-dashed blue lines and dashed green lines correspond to the exact conditional, DB-based and PB-based tests, respectively. The $p$-values are obtained under the no-changepoint scenario where $X_1, \ldots, X_T \sim_{i.i.d.} \text{Pois}(2)$ with $T = 10$ in the first row and $T = 30$ in the second.

Table 1. Power comparison of the exact conditional and PB-based tests for independent univariate Bernoulli time-series: $X_1, \ldots, X_T \sim_{i.i.d.} \text{Ber}(\pi_1)$, $X_{T+1}, \ldots, X_T \sim_{i.i.d.} \text{Ber}(\pi_2)$; $\alpha = 0.1$.

| $\pi_1$ | minP | PB-minP | LR | PB-LR | CU.5 | PB-CU.5 | CU1 | PB-CU1 |
|-------|-------|---------|----|-------|------|---------|-----|--------|
| 0.05  | 0.829 | 0.835   | 0.829 | 0.831 | 0.829 | 0.845   | 0.834 | 0.831  |
| 0.25  | 0.529 | 0.542   | 0.529 | 0.531 | 0.529 | 0.553   | 0.533 | 0.540  |
| 0.45  | 0.259 | 0.288   | 0.259 | 0.273 | 0.259 | 0.309   | 0.270 | 0.280  |
| 0.65  | 0.093 | 0.132   | 0.093 | 0.129 | 0.093 | 0.170   | 0.125 | 0.104  |

$T = 10$, $\tau = 5$, $\pi_2 = 0.85$

$T = 30$, $\tau = 15$, $\pi_2 = 0.65$

respectively. Although, in terms of power, the exact conditional and PB-based tests are quite similar, the latter slightly dominate the former in most cases. This is likely the case because PB-based tests, having stochastically smaller null $p$-values than the exact conditional tests (see Figure 4), do not guarantee type I error control.

In Table 2, we provide a power comparison of the exact conditional, DB-based and PB-based tests for independent Poisson time-series. Similar to the PB-based tests, the DB-based tests using the $P_{1(1)}$, $T_{LR}^{(c)}$, $T_{CUSUM}^{(0.5)}$ and $T_{CUSUM}^{(1)}$ statistics are referred to as ‘DB-minP’, ‘DB-LR’, ‘DB-CU.5’ and ‘DB-CU1’ tests, respectively. As in the binary case, PB-based tests slightly dominate the exact conditional tests in terms of power in most cases. Note, however, that DB-based tests perform significantly worse than both the exact conditional and PB-based tests, possibly due to the fact that DB samples from an observed Poisson time-series cannot contain a value that is within the support of the underlying Poisson distribution but has not been realized in the observed sample.
Table 2. Power comparison of the exact conditional, DB-based and PB-based tests for independent univariate Poisson time-series: $X_1, \ldots, X_\tau \overset{i.i.d.}{\sim} \text{Pois}(\lambda_1), X_{\tau+1}, \ldots, X_T \overset{i.i.d.}{\sim} \text{Pois}(\lambda_2); \alpha = 0.1.$

| $\lambda_1$ | minP | DB-minP | PB-minP | LR | DB-LR | PB-LR | CU.5 | DB-CU.5 | PB-CU.5 | CU1 | DB-CU1 | PB-CU1 |
|-------------|------|---------|---------|----|-------|-------|------|---------|---------|-----|--------|--------|
| 2           | 0.866| 0.689   | 0.870   | 0.844| 0.685 | 0.860 | 0.870 | 0.667   | 0.865  | 0.909| 0.799  | 0.907  |
| 3           | 0.545| 0.437   | 0.556   | 0.541| 0.430 | 0.549 | 0.552 | 0.450   | 0.562  | 0.617| 0.536  | 0.629  |
| 4           | 0.278| 0.242   | 0.284   | 0.275| 0.241 | 0.281 | 0.278 | 0.251   | 0.284  | 0.334| 0.312  | 0.341  |
| 5           | 0.132| 0.134   | 0.136   | 0.123| 0.130 | 0.136 | 0.136 | 0.137   | 0.136  | 0.138| 0.153  | 0.144  |

- $T = 10, \tau = 5, \lambda_2 = 6$
- $T = 30, \tau = 15, \lambda_2 = 2$
- $T = 40, \tau = 20, \lambda_2 = 1.5$

Figure 6. Comparison of $P(gCD)$ of global and local tests in $m = 1000$ independent Bernoulli series: $X_{j,1}, \ldots, X_{j,\tau} \overset{i.i.d.}{\sim} \text{Ber}(\pi_1 = 0.05), X_{j,\tau+1}, \ldots, X_{j,T} \overset{i.i.d.}{\sim} \text{Ber}(\pi_2 = 0.25)$ where $T = 200$. Tests are conducted at level $\alpha = 0.1$ and $n_{cp}$ channels undergo change at time-point $\tau$.

Figure 7. Comparison of $P(gCD)$ of global and local tests in $m = 200$ independent Poisson series: $X_{j,1}, \ldots, X_{j,\tau} \overset{i.i.d.}{\sim} \text{Pois}(\lambda_1 = 0.25), X_{j,\tau+1}, \ldots, X_{j,T} \overset{i.i.d.}{\sim} \text{Pois}(\lambda_2 = 1.5)$ where $T = 20$. Tests are conducted at level $\alpha = 0.1$ and $n_{cp}$ channels undergo change at time-point $\tau$.

4.3. Global vs. local testing in multiple channels

Global testing of $H_0$ is done by permutation tests using $C^{(5)}$ with Euclidean norm as discussed in Section 3. The $m$-variate time-series $X_1, \ldots, X_T$ is permuted $B = 1000$ times to
Figure 8. A single changepoint setup with $T = 200$ for multichannel Bernoulli data with potentially different changepoints in individual channels. $n_{cp} = 90$ out of $m = 1000$ channels have changes. There are 3 different changepoints, namely 60, 100 and 140, each appearing in 30 channels. We plot the distribution of changepoint locations across channels for each local test.

Figure 9. A single changepoint setup with $T = 50$ for multichannel Poisson data with potentially different changepoints in individual channels. $n_{cp} = 45$ out of $m = 200$ channels have changes. There are 3 different changepoints, namely 12, 24 and 36, each appearing in 15 channels. We plot the distribution of changepoint locations across channels for each local test.

obtain a randomized size-α test. For power comparisons, two tests ‘gCU.5’ and ‘gCU1’ are considered that are obtained using the global CUSUM statistics $C^{(0.5)}$ and $C^{(1)}$ respectively.

To test each channel for possible changepoints, we consider three exact conditional tests, namely minP, LR and CU1. After computing $p$-values from these tests, we employ the BH procedure to obtain $R = \{1 \leq j \leq m : H_{0j}$ is rejected$. \}$ Henceforth, we refer to these local tests as ‘minP-BH’, ‘LR-BH’ and ‘CU1-BH’ respectively. Figures 6 and 7 compare probabilities of global change detection (gCD), i.e. probabilities of rejecting $H_0$ (this is $P(R \neq \emptyset)$ for local tests) for global and local tests. Figure 6 considers Bernoulli channels while Figure 7 deals with Poisson channels. We find that the local tests are significantly more powerful than the global tests in the rare signal regime where $n_{cp}$ is small or moderate. Also, the power advantage is more and continues over a longer range of $n_{cp}$ when the changepoint is near the boundary. The local and global tests have comparable power for large $n_{cp}$, as expected.

In Figures 8 and 9 we consider situations where individual channels have different changepoints. Note that a global test for single change cannot be applied to such scenarios. We find that all three local tests detect these changepoint locations quite well.

In a simulation study presented in the supplementary article [12], we consider two additional FDR controlling methods, namely the adaptive Benjamini-Hochberg (ABH) and
the adaptive Storey-Taylor-Siegmund (STS) methods. Performances of these methods are comparable to that of the vanilla BH method as the simulation study is done under the rare signal regime.

5. Real data

Now we analyze two datasets that present themselves naturally as networks. These give real examples of multichannel binary and count data with potential changepoints.

5.1. US senate rollcall data

From the US senate rollcall dataset [30], we construct networks where nodes represent US senate seats. Each epoch represents a proposed bill on which votes were taken. An edge between two seats is formed if they voted similarly on that bill. We have $n = 100$ nodes. We consider $T = 50$ time-points between August 10, 1994 and January 24, 1995.

There are $m = \binom{n}{2} = 4950$ channels (i.e. edges). Of these, 622 channels are ignored while analyzing this data since those channels contain too many zeros or ones (more than 45). We applied the BH procedure to simultaneously test the remaining 4328 channels controlling FDR at level $\alpha = 0.05$. For each significant channel, the corresponding changepoint location is also reported (see the discussion in Section 2.5).

In Figure 10, we plot the histograms of the changepoint locations of the significant channels. Note particularly the peak near time-point 24 (which corresponds to December 1, 1994). There is a historically well-documented change near December 1994, which saw the end of the conservative coalition (see, e.g. [31]). Interestingly, the global method also detected a changepoint at $t = 24$. Changepoints at nearly the same location were found earlier in [14,15]. However, the local methods have the advantage of identifying the channels that underwent a change. The number of significant channels, $\hat{n}_{cp}$, is reported below each histogram. A number of channels had extremely small $p$-values. For example, Figure 11(a) depicts the time-series of edges $(4, 5)$ and $(4, 6)$. Changes are visible to the naked eye. Seats 5 and 6 are in Arizona, while seat 4 is in Arkansas. Clearly, seat 4 went from agreeing with seats 5 and 6 to disagreeing. On the other hand, seat 3 is also from Arkansas, and no changepoints were found in the channels $(3, 5)$ and $(3, 6)$ (see Figure 11(b)).

![Figure 10](image-url)  
**Figure 10.** Distribution of detected changepoint locations in the US senate rollcall data. All three methods report a mode at $t = 24$. 

Figure 11. (a) Edges (4, 5) and (4, 6). Seats 5 and 6 are in Arizona, while seat 4 is in Arkansas. Clearly, seat 4 went from agreeing with seats 5 and 6 to disagreeing. (b) Edges (3, 5) and (3, 6). Seat 3 is also in Arkansas. No changepoints are detected in these channels.

5.2. MIT reality mining data

We use the MIT reality mining data [32] to construct a series of networks involving $n = 90$ individuals (staff and students at the university). The data consists of call logs between these individuals from 20th July 2004 to 14th June 2005. We construct $T = 48$ weekly networks, where a weighted edge between nodes $u$ and $v$ reports the number of phone calls between them during the corresponding week. There are $m = \binom{n}{2} = 4005$ channels (i.e., edges). 3945 channels are ignored while analyzing this data since those channels contain too many zeros (more than 44). The remaining 60 channels are tested for possible changepoints.

We model the weighted edges as Poisson variables and apply the exact tests $\text{minP}$, $\text{LR}$ or $\text{CU1}$ on each channel. Then we apply the BH method to simultaneously test the 60 channels controlling FDR at level $\alpha = 0.05$. Figure 12 contains the histograms of the detected changepoint locations. Note particularly the peaks near $t = 20$ and $t = 33$. For comparison, a changepoint at $t = 24$ was found by a global algorithm in [15]. The graph-based multivariate (global) change detection methods of [11] found changepoints at approximately $t = 22$ and $t = 25$ (their analyzes were on daily networks). The global algorithms consider global characteristics, and thus it is not surprising that they find changepoints

Figure 12. Changepoint locations from edge-based analysis of the MIT reality mining data.
somewhat in the middle of the predominant local changepoints near $t = 20$ and $t = 33$. It turns out that $t = 20$ is just before the start of the Winter break, and $t = 33$ is just before the start of the Spring break.

We also perform changepoint analysis with node-degrees as channels, modelled as a Poisson series. Figure 13 shows the histograms of the detected changepoint locations. Analyses of edge and degree time-series detect 44 and 46 common nodes (i.e. detected by all three tests: minP-BH, LR-BH and CU1-BH) respectively. Among these, 40 nodes are declared significant by both analyzes.

We also apply permutation tests without the Poisson assumption on edges or degrees. However, the sets of changepoints detected by these tests are strict subsets of the sets of changepoints detected by the corresponding exact conditional tests. This can be explained by the facts that (i) for Poisson data, permutation tests are more conservative than exact conditional tests (see, e.g. Figure 3), and (ii) for sparse networks like the MIT reality mining networks, both weighted edges and degrees are well-approximated by suitable Poisson distributions.

Finally, in Figure 14, we show the average networks before and after time-point 20. The 46 channels (i.e. nodes) declared to have changepoints by all three tests under the degree-based analysis are shown as green circles. A structural change is clearly visible.
6. Discussion

In this article, we have considered the problem of changepoint detection for binary and count data, and proposed exact tests that perform significantly better than Brownian-bridge based asymptotic tests in small samples. We have also considered multichannel data and used a multiple testing approach to test for changes in all channels simultaneously. This local approach outperforms the global approach of treating all channels together as a single object quite significantly in case of rare signals (i.e. when the number of channels with a changepoint is much smaller than the total number of channels).

Although the methods we propose are technically for single changes, they work quite well for multiple changepoints, especially when there is one large change. This is empirically demonstrated in the supplementary article [12].

Acknowledgments

We thank the referees for their thoughtful comments which led to an improved version of this manuscript.

Data availability statement

The US senate rollcall dataset is available in [30] and the MIT reality mining data is available in [32]. We constructed the corresponding network-valued time-series using the above mentioned publicly available datasets. The constructed network time-series that support the findings of this study are available from the corresponding author upon request.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

The research of Shyamal K. De is supported by grant MTR/2017/000503 under the MATRICS scheme provided by Science and Engineering Research Board, Government of India. Soumendu Sundar Mukherjee is supported by an INSPIRE Faculty Fellowship from the Department of Science and Technology, Government of India.

References

[1] Page ES. Continuous inspection schemes. Biometrika. 1954;41(1/2):100–115.
[2] Page ES. On problems in which a change in a parameter occurs at an unknown point. Biometrika. 1957;44(1/2):248–252.
[3] Girshick MA, Rubin H. A bayes approach to a quality control model. Ann Math Stat. 1952;23(1):114–125.
[4] Brodsky E, Darkhovsky BS. Nonparametric methods in change point problems. Vol. 243. Dordrecht: Springer; 1993.
[5] Zhang NR, Siegmund DO, Ji H, et al. Detecting simultaneous changepoints in multiple sequences. Biometrika. 2010;97(3):631–645.
[6] Siegmund D, Yakir B, Zhang NR. Detecting simultaneous variant intervals in aligned sequences. Ann Appl Stat. 2011;5(2A):645–668.
[7] Srivastava M, Worsley KJ. Likelihood ratio tests for a change in the multivariate normal mean. J Am Stat Assoc. 1986;81(393):199–204.
[8] James B, James KL, Siegmund D. Asymptotic approximations for likelihood ratio tests and confidence regions for a change-point in the mean of a multivariate normal distribution. Stat Sin. 1992;2(1):69–90.

[9] Harchaoui Z, Moulines E, Bach FR. Kernel change-point analysis. In: Advances in Neural Information Processing Systems, Vancouver, B.C., Canada; 2009. p. 609–616.

[10] Lung-Yut-Fong A, Lévy-Leduc C, Cappé O. Homogeneity and change-point detection tests for multivariate data using rank statistics. Preprint; 2011. arXiv:11071971.

[11] Chen H, Zhang N. Graph-based change-point detection. Ann Statist. 2015 Feb;43(1):139–176.

[12] Supplement to “Exact tests for change-point detection in multichannel binary and count data”. 2021.

[13] Peel L, Clauset A. Detecting change points in the large-scale structure of evolving networks. In: AAAI, Austin (TX), USA; 2015. p. 2914–2920.

[14] Roy S, Atchadé Y, Michailidis G. Change point estimation in high dimensional markov random-field models. J R Stat Soc Ser B (Stat Methodol). 2017;79(4):1187–1206.

[15] Mukherjee SS. On some inference problems for networks [dissertation]. Berkeley: University of California; 2018.

[16] Wang D, Yu Y, Rinaldo A. Optimal change point detection and localization in sparse dynamic networks. Preprint; 2018. arXiv:180909602.

[17] Padilla OHM, Yu Y, Priebe CE. Change point localization in dependent dynamic nonparametric random dot product graphs. Preprint; 2019. arXiv:191107494.

[18] Zhao Z, Chen L, Lin L. Change-point detection in dynamic networks via graphon estimation. Preprint; 2019. arXiv:190801823.

[19] Bhattacharjee M, Banerjee M, Michailidis G. Change point estimation in a dynamic stochastic block model. J Mach Learn Res. 2020;21(107):1–59.

[20] Bhattacharyya S, Chatterjee S, Mukherjee SS. Consistent detection and optimal localization of all detectable change-points in piecewise stationary arbitrarily sparse network-sequences. In preparation. 2020.

[21] Lehmann EL, Romano JP. Testing statistical hypotheses. New York: Springer Science & Business Media; 2005.

[22] Antoch J, Hušková M, Veraverbeke N. Change-point problem and bootstrap. J Nonparametr Stat. 1995;5(2):123–144.

[23] Gombay E, Horváth L. Change-points and bootstrap. Environmetrics. 1999;10(6):725–736.

[24] Gijbels I, Goderniaux A. Bootstrap test for change-points in nonparametric regression. J Nonparametr Stat. 2004;16(3-4):591–611.

[25] Holm S. A simple sequentially rejective multiple test procedure. Scand J Statist. 1979;6(2):65–70.

[26] Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. J R Stat Soc Ser B (Methodol). 1995;57(1):289–300.

[27] Hochberg Y, Benjamini Y. More powerful procedures for multiple significance testing. Stat Med. 1990;9(7):811–818.

[28] Storey JD, Taylor JE, Siegmund D. Strong control, conservative point estimation and simultaneous conservative consistency of false discovery rates: a unified approach. J R Stat Soc Ser B (Stat Methodol). 2004;66(1):187–205.

[29] Vostrikova LY. Detecting disorder in multidimensional random processes. In: Doklady Akademii Nauk. Vol. 259. Russian Academy of Sciences; 1981. p. 270–274.

[30] Lewis JB, Poole K, Rosenthal H, et al. Voteview: congressional roll-call votes database. 2020. Available from: https://voteview.com/.

[31] Moody J, Mucha PJ. Portrait of political party polarization. Netw Sci. 2013;1(1):119–121.

[32] Eagle N, Pentland AS. Reality mining: sensing complex social systems. Pers Ubiquitous Comput. 2006;10(4):255–268.