Change Point Detection Based On Directed Nearest Neighbor Graph

Yuze Zhou

2017 年 6 月 14 日

Contents

1 Abstract

Change-point models are usually used in various fields of study to detect the lack of homogeneity in a sequence of observations. In a typical situation, the observations $\{\mathbf{y_i} \quad i = 1, 2, \cdots n\}$are assumed to have distribution F_0 for i before τ and another distribution F_1 for i after τ.

Change-point models have a variety of applications to real-life data. Here I present two special examples. It is known to all that the industrial revolution has greatly changed the history of Britain and Europe, which boosted the growth of many industries. If the data of the profits of all industries in every year around the revolution could be obtained, then change point detection could be applied to the data to find out whether the revolution had significantly changed the pattern of industries. Change point detection could also be applied to text analysis of novels. For example, it has long been debated whether the author of *A Dream in Red Mansions* has changed at around the 80th chapter, and the two authors might have different preference for the characters, so we can analyzed the character-event data by change point detection to validate or falsify those guesses.

In general, the change-point problems can be represented in the following statistical formulation: A sequence of observations $\{\mathbf{y_i} \quad i = 1, 2, \cdots n\}$ index by some meaningful orderings $\{1, 2, \cdots n\}$, such as time and rankings. We are concerned with testing the null hypothesis:

$$H_0: \quad \mathbf{y_i} \sim F_0$$

against the alternative hypothesis:

$$\begin{aligned} H_\alpha: \quad & \mathbf{y_i} \sim F_0 \qquad i \leq \tau \\ & \mathbf{y_i} \sim F_1 \qquad i > \tau \end{aligned}$$

where F_0 and F_1 are two different distributions.

Change-point detection problems are most studied under the assumption that the $\mathbf{y_i}$ s from the same distribution are independently and identically sampled, which can be easily violated in many cases. However, the $i.i.d$ assumption is very important as it makes theoretical studies possible.

2 Introduction

In general, the types of distribution of $\mathbf{y_i}$ are not strictly required, however, some distance structures are required so that the observations $\mathbf{y_i}$ can be represented in the graph, where the edges in the graph connects observations that are close to each other.

Directed K - Nearest Neighbor Graph **KNNG**(*k-nearest neighbor graph*) is an algorithm widely used in statistical estimation and pattern recognition, the idea of which is simple. For a set P consisted of n observations from a metric space (usually a Euclidean space), a KNNG graph is a directed graph with P being its vertex and with a directed edge connecting p to q whenever q is among the k vertex which is closet to p.

In most literatures of machine learning, **KNNG** is often represented in an undirected version. An edge connecting p with q exists as long as p is among the nearest neighbors of q or q is among the nearest neighbors p. However, in the special case of change-point problems, a directed **KNNG** is much more powerful than an undirected **KNNG** as it maintains more information and could handle more cases.

Here I specially choose three different cases to demonstrate how a directed **KNNG** could represent the dissimilarities between observations. These cases are: the general case with no change points, the meanshift case and the variance change case. For each case, we have 30 observations in all, where the pink dots represent observations sampled before $\tau = 15$, and the purple triangles represent observations sampled after $\tau = 15$

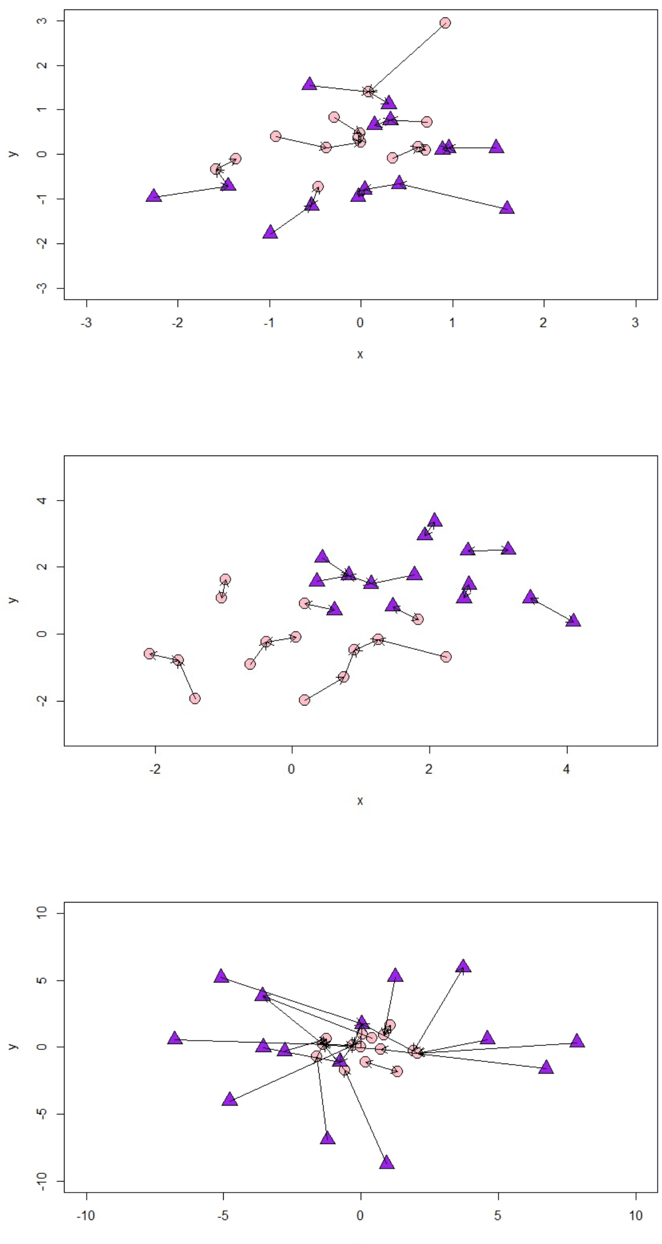

Figure 1: From above to bottom: **Case One: No Change Point**, **Case Two: Mean Shift**, **Case Three: Variance Change**

Case One: No Change Point In this case, all the dots and triangles are sampled from the distribution $N(\mathbf{0}, I_2)$, which means no change point has ever occurred. All observations and edges are scattered equally and evenly in the figure.

Case Two: Meanshift In case one, the pink dots are samples from the distribution $N(\mathbf{0}, I_2)$ and the purple triangles are samples from the distribution $N((2, 2), I_2)$, where the two distributions have the same variance but different means. Compared to case one, in case two, the number of edges connecting pink dots to purple triangles, as well those connecting purple triangles to pink dots, have both declines significantly.

Case Three: Variance Change In case three, the pink dots are sample from the distribution $N(\mathbf{0}, I_{30})$ and the purple triangles are sampled from the distribution $N(\mathbf{0}, 4 * I_{30})$, where the two distributions have the same mean but different variances. The dimensions of the observations are reduced by **MDS**(*Multi-dimensional Scaling*) method so that they can be plotted on a two-dimensional graph. Compared to case one, in case three, the number of edges connecting pink dots to purple triangles has declines significantly, however, number of those connecting purple triangles to pink dots has much increased.

In case two, a directed **KNNG** does not have any advantages over an undirected **KNNG**, as the number of edges connecting the pink dots and the purple triangles also decreases in a meanshift case. However, in case three, a directed **KNNG** much outperforms the undirected graph. The reason is that, though the number of edges connecting pink dots to purple triangles decreases and the number of those connecting purple triangles to pink dots increases, the sum of both of them, which is the number of edges connecting the pink dots and the purple triangles in an undirected **KNNG**, does not change much.

3 The New Test Statistics

3.1 The Basic Statistics

First of all, let's derive the basic statistics that are required for constructing the new statistic, for each different t, all the observations can be divided into two different groups; observations sampled before t and those sampled after t. Let G be the **KNNG** constructed, including both the vertexes and edges. We also use i or j to represent a vertex in the graph G, and (i,j) to represent an edge in G that connecting i to j, if existed. Also for any event x, let I_x be the indicator function that take value 1 if x happens and takes value 0 if x does not occur. Then we can derive the two basic statistics $G_{12}(t)$ and $G_{21}(t)$:

Definition 3.1.1 *$G_{12}(t)$ and $G_{21}(t)$*

$$G_{12}(t) = \sum_{(i,j)\in G} I_{g_i(t)}(1 - I_{g_j(t)})$$

$$G_{21}(t) = \sum_{(i,j)\in G} (1 - I_{g_i(t)})I_{g_j(t)}$$

$$I_{g_i(t)} = I_{i\leq t}$$

In another sense, $G_{12}(t)$ counts the number of the edges that connects observations sampled before and at t to those sampled after t, and contrarily, $G_{21}(t)$ counts the number of the edges that connects observations that sample after t to those sampled before and at t.

3.2 Properties of the Basic Statistics

Under the null hypothesis H_0 and the $i.i.d$ assumption, the joint distribution of $\{\mathbf{y_i} \quad i = 1, 2, \cdots n\}$ is the same under the permutation distribution.We define the null distribution of $G_{12}(t)$ and $G_{21}(t)$ to be the permutation distribution, which places $1/n!$ probability on each of the $n!$ permutations of $\{\mathbf{y_i} \quad i = 1, 2, \cdots n\}$. Let $\pi(i)$ be the time of observing $\mathbf{y_i'}$ after permutation, then for the permuted sequence, $g_j(t)$ becomes $I_{\pi(i)\leq t}$. Notice that the graph G is determined by the values of $\mathbf{y_i'}$s, not their order of appearance, and thus remains constant under permutation. When there is no further specification, we denote by **P**, **E**, **Var** the probability, expectation and variance, respectively, under the permutation null distribution.

Lemma 3.2.1 *Under the permutation null distribution, the expectations of $G_{12}(t)$ and $G_{21}(t)$ are*

$$\boldsymbol{E}(G_{12}(t)) = K * \tfrac{t(n-t)}{n-1}$$
$$\boldsymbol{E}(G_{21}(t)) = K * \tfrac{t(n-t)}{n-1}$$

Lemma 3.2.2 *Under the permutation null distribution, the variances and covariances of $G_{12}(t)$ and $G_{21}(t)$ are*

$$\begin{aligned}
\boldsymbol{Var}(G_{12}(t)) &= Var(\textstyle\sum_{i=1}^{t}\sum_{j=t+1}^{n} A_{ij}^{+}) \\
&= t*(n-t)*Var(A_{ij}^{+}) \\
&+2*t*C_{n-t}^{2}*cov(A_{ij}^{+},A_{il}^{+}) \\
&+2*(n-t)*C_{t}^{2}*cov(A_{ij}^{+},A_{lj}^{+}) \\
&+4*C_{t}^{2}*C_{n-t}^{2}*cov(A_{ij}^{+},A_{kl}^{+})
\end{aligned}$$

$$\begin{aligned}
\boldsymbol{Var}(G_{21}(t)) &= Var(\textstyle\sum_{i=t+1}^{n}\sum_{j=1}^{t} A_{ij}^{+}) \\
&= t*(n-t)*Var(A_{ij}^{+}) \\
&+2*(n-t)*C_{t}^{2}*cov(A_{ij}^{+},A_{il}^{+}) \\
&+2*t*C_{n-t}^{2}*cov(A_{ij}^{+},A_{lj}^{+}) \\
&+4*C_{t}^{2}*C_{n-t}^{2}*cov(A_{ij}^{+},A_{kl}^{+})
\end{aligned}$$

$$\begin{aligned}
\boldsymbol{cov}(G_{12}(t),G_{21}(t)) &= cov(\textstyle\sum_{i=1}^{t}\sum_{j=t+1}^{n} A_{ij}^{+},\sum_{k=t+1}^{n}\sum_{l=1}^{t} A_{kl}^{+}) \\
&= \textstyle\sum_{i=1}^{t}\sum_{j=t+1}^{n}\sum_{k=t+1}^{n}\sum_{l=1}^{t} cov(A_{ij}^{+},A_{kl}^{+}) \\
&= t*(n-t)*cov(A_{ij}^{+},A_{ji}^{+}) \\
&+2*t*C_{n-t}^{2}*cov(A_{ij}^{+},A_{ki}^{+}) \\
&+2*(n-t)*C_{t}^{2}*cov(A_{ki}^{+},A_{ij}^{+}) \\
&+4*C_{t}^{2}*C_{n-t}^{2}*cov(A_{ij}^{+},A_{kl}^{+})
\end{aligned}$$

where, in the equations above,

$$A_{ij}^{+} = \boldsymbol{I}(\boldsymbol{y}_j \quad \textit{is among the } k \textit{ nearest neighbours of} \quad \boldsymbol{y}_i)$$

Lemma 3.2.3 *Under the permutation null distribution, the variances and covariances of different pairs of A_{ij}^{+} can also be calculated, which are,*

$$\begin{aligned}
&\boldsymbol{Var}(A_{ij}^{+}) = \tfrac{K*(n-1-K)}{(n-1)^2} \\
&\boldsymbol{cov}(A_{ij}^{+},A_{ji}^{+}) = \tfrac{\sum_{i,j=1}^{n} a_{ij}^{+},a_{ji}^{+}}{n*(n-1)} - (\tfrac{K}{n-1})^2 \\
&\boldsymbol{cov}(A_{ij}^{+},A_{il}^{+}) = -\tfrac{K*(n-1-K)}{(n-1)^2*(n-2)} \\
&\boldsymbol{cov}(A_{ij}^{+},A_{ki}^{+}) = \tfrac{n*K^2-\sum_{i,j=1}^{n} a_{ij}^{+},a_{ji}^{+}}{n*(n-1)(n-2)} - (\tfrac{K}{n-1})^2 \\
&\boldsymbol{cov}(A_{ij}^{+},A_{lj}^{+}) = \tfrac{\sum_{i,j,l=1}^{n} a_{ij}^{+}*a_{lj}^{+}-K*n}{n*(n-1)*(n-2)} - (\tfrac{K}{n-1})^2 \\
&\boldsymbol{cov}(A_{ij}^{+},A_{lk}^{+}) = \\
&\tfrac{K*(n*K-K+1)}{(n-1)*(n-2)*(n-3)} - \tfrac{2*n*K^2+\sum_{i,j,l=1}^{n} a_{ij}^{+}*a_{lj}^{+}-\sum_{i,j=1}^{n} a_{ij}^{+}*a_{ji}^{+}}{n*(n-1)*(n-2)*(n-3)} - (\tfrac{K}{n-1})^2
\end{aligned}$$

In the equations above, $\sum_{i,j=1}^{n} a_{ij}^{+}a_{ji}^{+}$ is the number of pairs of edges which connect the same nodes but in different directions and $\sum_{i,j,l=1}^{n} a_{ij}^{+}a_{lj}^{+}$ is the number of pairs of edges which share the same ending nodes, including those also sharing the same starts.

3.3 Derivation of the New Test Statistic

3.3.1 Transformation of the Basic Statistic

We first derive two new statistics that are necessary for constructing the test statistic from the two basic statistics. First, we have to make clear several important symbols:

$$r_0 = \frac{\sum_{i,j=1}^{n} a_{ij}^{+}a_{ji}^{+}}{n}$$

$$r_1 = \frac{\sum_{i,j,l=1}^{n} a_{ij}^{+}a_{lj}^{+}}{n}$$

$$p = \frac{t}{n}$$

Having made those important symbols clear, we can now transform the two original statistics into two new statistics, $\hat{G_{12}}(t)$ and $\hat{G_{21}}(t)$:

Definition 3.3.1 *$\hat{G_{12}}(t)$ and $\hat{G_{21}}(t)$:*

$$\begin{pmatrix} \hat{G_{12}}(t) \\ \hat{G_{21}}(t) \end{pmatrix} = \begin{pmatrix} \frac{1}{\sqrt{(r_1-K^2)p(1-p)}} & 0 \\ 0 & \frac{1}{\sqrt{(r_0+K)p^2(1-p)^2}} \end{pmatrix} \begin{pmatrix} 1 & -1 \\ 1-p & p \end{pmatrix} \begin{pmatrix} \frac{1}{\sqrt{n}}(G_{12}(t) - E(G_{12}(t))) \\ \frac{1}{\sqrt{n}}(G_{21}(t) - E(G_{21}(t))) \end{pmatrix}$$

The transformation of the two original statistics, $G_{12}(t)$ and $G_{21}(t)$ seems to be a little bit awkward and unnatural at the first sight, and the two new statistics seems to contain little information related to the graph. However, $\hat{G_{12}}(t)$ and $\hat{G_{21}}(t)$ display very good asymptotic qualities as we will show later in the asymptotic section.

3.3.2 Derivation of the Test Statistic

Now first let us define another new statistic that is used directly for constructing the test statistic:

Definition 3.3.2 $Z_G(t)$

$$Z_G(t) = \hat{G_{12}}^2(t) + \hat{G_{21}}^2(t)$$

Theoretically, if a change point ever occurs, at least one of the two statistics, $G_{12}(t)$, $G_{21}(t)$, will be much deviated from its permutation expectation, also making $\hat{G_{12}}^2(t)$ and $\hat{G_{21}}^2(t)$ deviate from their expectation, thus the value of $Z_G(t)$ will increase. This result is also further confirmed through simulations.

The same three cases, **No Change Point**, **Mean Shift**, **Variance Change**, are chosen to demonstrate how $Z_G(t)$ can represent the dissimilarities between observations, and the figures of $Z_G(t)$ versus t are also plotted below.

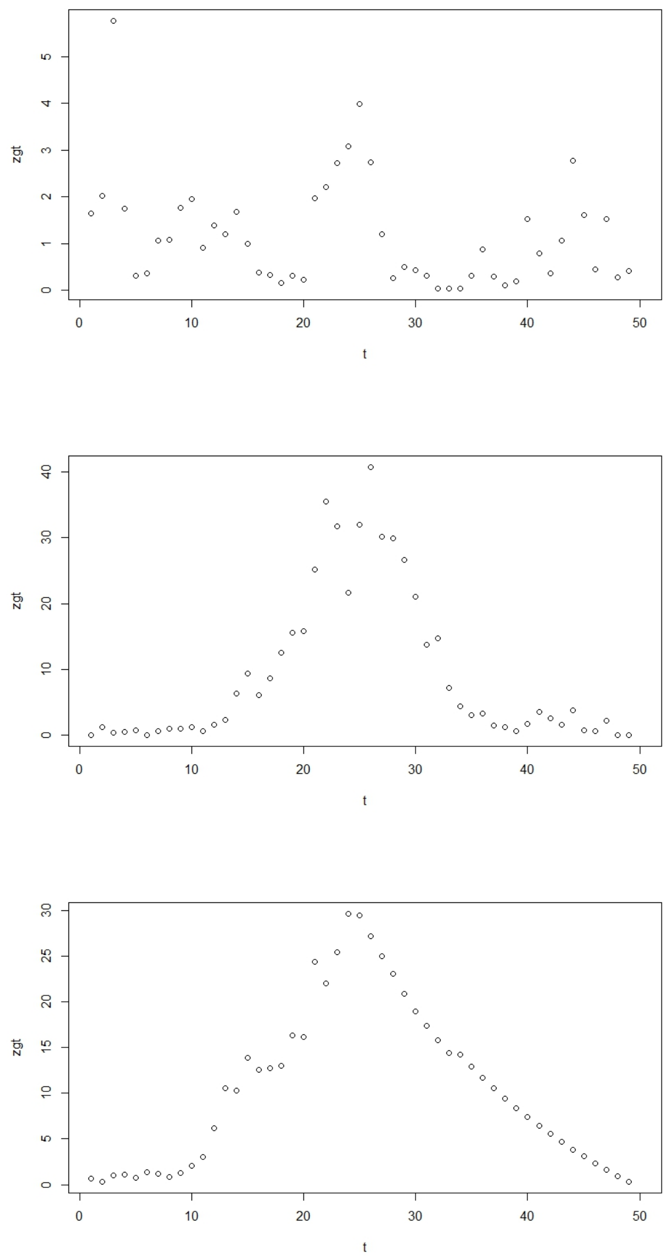

Figure 2: Figure of $Z_G(t)$, From top to bottom: **Case One: No Change Point**, **Case Two: Mean Shift**, **Case Three: Variance Change**

Case One: No Change Point In this case, all observations are sampled from the distribution $N(\mathbf{0}, I_2)$, which means no change point has ever occurred. $Z_G(t)$ has remained comparably low, less than 5, all the time.

Case Two: Meanshift In case one, the first 25 observations are sampled from the distribution $N(\mathbf{0}, I_2)$ and the next 25 observations are sampled from the distribution $N((2,2), I_2)$, where the two distributions have the same variance but different means. Compared to case one, in case two, $Z_G(t)$ has increased significantly to more than 40 at around the 25th observation.

Case Three: Variance Change In case three, the first 25 observations are sampled from the distribution $N(\mathbf{0}, I_{30})$ and the next 25 are sampled from the distribution $N(\mathbf{0}, 4 * I_{30})$, where the two distributions have the same mean but different variances. Compared to case one, in case three, $Z_G(t)$ has also increased significantly to more than 30 at around the 25th observation.

From the three examples above, we can find out that in most cases, if a change point occurs at t, the value of $Z_G(t)$ will be greatly increased at around t, which motivates us to derive the scan statistic to test H_0 versus H_α:

Definition 3.3.2 The Test Statistic

$$\max_{n_0 \leq t \leq n_1} Z_G(t)$$

where n_0 and n_1 are some pre-specified constraints for the range of t, because it is hard to judge whether a change point has occurred only within a few observations. The null hypothesis is rejected if the maximal value of $Z_G(t)$ is greater than some threshold.

4 Asymptotic

How larger does the test statistic need to be to provide sufficient evidence against the null hypothesis? In other words, we are much concerned with the tail probability of the test statistic under the null hypothesis H_0, which is,

$$P(\max_{n_0 \leq t \leq n_1} Z_G(t) > b)$$

The null distributions of $\max_{n_0 \leq t \leq n_1} Z_G(t)$ are defined as the permutation null distribution. For small n, we can sample directly from the permutation distribution to approximate the test statistic; however, when n becomes larger, permutation is computationally prohibitive. Therefore, we have to derive analytic expressions for the tail probabilities to make the method applicable. Moreover, the distribution of $Z_G(t)$ is quite complicated, thus making the exact expression of it impossible to derive, so in the rest of this chapter, we will give the analytic approximations for that tail probability.

4.1 Asymptotic Distribution

In this section, we will derive the limiting distribution of $G_{12}([nu])$ and $G_{21}([nu])$ as well as $\hat{G_{12}}([nu])$ and $\hat{G_{21}}([nu])$, where $0 < u < 1$. But before the discussion of the limiting distribution, we have to introduce some notations about the edges in the graph first:

For any edge e in the graph, let

$$K_e = \left\{e^{'} : e^{'} shares \quad a \quad node \quad with \quad e\right\}$$

$$N_e = \{nodes \quad in \quad K_e\}$$

$$L_e = \left\{e^{"} : e^{"} has \quad a \quad node \quad in \quad N_e\right\}$$

and also let

$$|K_e| = the \quad number \quad edges \quad in \quad K_e$$

$$|L_e| = the \quad number \quad edges \quad in \quad L_e$$

Then we can define an asymptotic condition for the graph

Condition 4.4.1

$$\sum_{e \in G} |K_e||L_e| \sim o(n^{\alpha}) \qquad 1 < \alpha < 1.5$$

In a specific graph, $\sum_{e \in G} |K_e||L_e|$ indicates how big the 'hub' of edges can get, and α indicates the order of the hub. In most cases, 1.5 is a comparably big upper-bound that can be easily satisfied.

Lemma 4.1.1 *Under* ***Condition 4.1.1****,*
both $(\frac{1}{\sqrt{n}}(G_{12}([nu]) - E(G_{12}([nu])))$ *and* $\frac{1}{\sqrt{n}}(G_{21}([nu]) - E(G_{12}([nu]))))$ *will converge in distribution to Gaussian processes*

Lemma 4.1.2 *Under* ***Condition 4.1.1****,*
$\hat{G_{12}}([nu])$ *and* $\hat{G_{21}}([nu])$ *will converge to two independent Gaussian processes* $\hat{G_{12}}^*(u)$ *and* $\hat{G_{21}}^*(u)$ *separately, which have the covariance structure:* $0 < u < v < 1$

$$Var\begin{pmatrix} \hat{G_{12}}^*(u) \\ \hat{G_{21}}^*(u) \\ \hat{G_{12}}^*(v) \\ \hat{G_{21}}^*(v) \end{pmatrix} = \begin{pmatrix} 1 & 0 & \frac{u(1-v)}{\sqrt{uv(1-u)(1-v)}} & 0 \\ 0 & 1 & 0 & \frac{u(1-v)}{(1-u)v} \\ \frac{u(1-v)}{\sqrt{uv(1-u)(1-v)}} & 0 & 1 & 0 \\ 0 & \frac{u(1-v)}{(1-u)v} & 0 & 1 \end{pmatrix}$$

From the results above, we can find out that $Z_G([nu])$ will also converge to a stochastic process $Z_G^*(u)$. It is also fairly easy to find out that for $0 < u < v < 1$, both $Z_G^*(u)$ and $Z_G^*(v)$ conform to χ^2 distributions, however, other than independent, they are much correlated.

4.2 Asymptotic P-value Approximation

4.2.1 Approximations of the Tail Probability

In this section, we are concerned with approximating the value of $P(\max_{n_0 \le t \le n_1} Z_G(t) > b)$ through the limiting process $Z_G^*(u)$. As $\hat{G_{12}}^*(u)$ and $\hat{G_{21}}^*(u)$ are two independent Gaussian processes with unit variances, we can derive the approximation as:

$$\begin{aligned}
& P(\max_{n_0 \le t \le n_1} Z_G(t) > b) \\
\approx\ & P(\max_{\frac{n_0}{n} \le t \le \frac{n_1}{n}} Z_G^*(t) > b) \\
=\ & P(\max_{\frac{n_0}{n} \le t \le \frac{n_1}{n}} \hat{G_{12}}^*(u)^2 + \hat{G_{21}}^*(u)^2 > b) \\
\approx\ & P(\{\max_{\frac{n_0}{n} \le t \le \frac{n_1}{n}} \hat{G_{12}}^*(u)^2 > \tfrac{b}{2}\} \bigcap \{\max_{\frac{n_0}{n} \le t \le \frac{n_1}{n}} \hat{G_{21}}^*(u)^2 > \tfrac{b}{2}\}) \\
=\ & P(\max_{\frac{n_0}{n} \le t \le \frac{n_1}{n}} \hat{G_{12}}^*(u)^2 > \tfrac{b}{2}) P(\max_{\frac{n_0}{n} \le t \le \frac{n_1}{n}} \hat{G_{21}}^*(u)^2 > \tfrac{b}{2})
\end{aligned}$$

where separately, $P(\max_{\frac{n_0}{n}\le t\le\frac{n_1}{n}} \hat{G_{12}}^*(u)^2 > \frac{b}{2})$ and $P(\max_{\frac{n_0}{n}\le t\le\frac{n_1}{n}} \hat{G_{21}}^*(u)^2 > \frac{b}{2})$ are approximated as:

$$\begin{aligned}
&P(\max_{\frac{n_0}{n}\le t\le\frac{n_1}{n}} \hat{G_{12}}^*(u)^2 > \tfrac{b}{2}) \\
&= P(\max_{\frac{n_0}{n}\le t\le\frac{n_1}{n}} |\hat{G_{12}}^*(u)| > \sqrt{\tfrac{b}{2}}) \\
&\approx 2 * P(\max_{\frac{n_0}{n}\le t\le\frac{n_1}{n}} \hat{G_{12}}^*(u) > \sqrt{\tfrac{b}{2}})
\end{aligned}$$

$$\begin{aligned}
&P(\max_{\frac{n_0}{n}\le t\le\frac{n_1}{n}} \hat{G_{21}}^*(u)^2 > \tfrac{b}{2}) \\
&= P(\max_{\frac{n_0}{n}\le t\le\frac{n_1}{n}} |\hat{G_{21}}^*(u)| > \sqrt{\tfrac{b}{2}}) \\
&\approx 2 * P(\max_{\frac{n_0}{n}\le t\le\frac{n_1}{n}} \hat{G_{21}}^*(u) > \sqrt{\tfrac{b}{2}})
\end{aligned}$$

From the derivation above, the key to approximating the tail probability $P(\max_{n_0\le t\le n_1} Z_G(t) > b)$ is the estimation of $P(\max_{\frac{n_0}{n}\le t\le\frac{n_1}{n}} \hat{G_{12}}^*(u) > \sqrt{\frac{b}{2}})$ and $P(\max_{\frac{n_0}{n}\le t\le\frac{n_1}{n}} \hat{G_{21}}^*(u) > \sqrt{\frac{b}{2}})$.

4.2.2 Asymptotic Approximations to the p-values

In this section, we are mainly concerned with approximating $P(\max_{\frac{n_0}{n}\le t\le\frac{n_1}{n}} \hat{G_{12}}^*(u) > \sqrt{\frac{b}{2}})$ and $P(\max_{\frac{n_0}{n}\le t\le\frac{n_1}{n}} \hat{G_{21}}^*(u) > \sqrt{\frac{b}{2}})$ by taking advantage of its covariance structure. Our approximation involves the function $\upsilon(x)$, which is defined as:

$$\upsilon(x) = 2x^{-2}exp\{-2\sum_{m=1}^{\infty} m^{-1}\Phi(-\tfrac{1}{2}xm^{\frac{1}{2}})\}$$

where $\Phi(x)$ is the probability distribution function of standard normal distribution. This function is closely related to the Laplace transform of the overshoot over the boundary of a random walk. A simple approximation given in Siegmund and Yakir is sufficient for numerical purposes:

$$\upsilon(x) \approx \tfrac{(2/x)(\Phi(x/2)-0.5)}{(x/2)\Phi(x/2)+\phi(x/2)}$$

where $\phi(x)$ is the probability density function of the standard normal distribution

Lemma 4.2.1 *Assume that as $n \to \infty$,*

$$\tfrac{n_0}{n} \to x_0, \tfrac{n_1}{n} \to x_1 \text{ and } \sqrt{\tfrac{b}{2n}} \to b_0$$

Then also as $n \to \infty$,

$$P(\max_{\frac{n_0}{n}\le t\le\frac{n_1}{n}} \hat{G_{12}}^*(u) > \sqrt{\tfrac{b}{2}}) \sim \sqrt{\tfrac{b}{2}}\phi(\sqrt{\tfrac{b}{2}})\int_{x_0}^{x_1} g_{12}(x)\upsilon(b_0\sqrt{2g_{12}(x)})dx$$
$$P(\max_{\frac{n_0}{n}\le t\le\frac{n_1}{n}} \hat{G_{21}}^*(u) > \sqrt{\tfrac{b}{2}}) \sim \sqrt{\tfrac{b}{2}}\phi(\sqrt{\tfrac{b}{2}})\int_{x_0}^{x_1} g_{21}(x)\upsilon(b_0\sqrt{2g_{21}(x)})dx$$

Where

$$g_{12}(x) = \frac{1}{2x(1-x)}$$
$$g_{21}(x) = \frac{1}{x(1-x)}$$

Based on **Lemma 4.2.1**, when $\sum_{e \in G} |K_e||L_e| \sim o(n^\alpha)$ where $1 < \alpha < 1.5$, we can approximate $P(\max_{n_0 \le t \le n_1} Z_G(t) > b)$ by

$$\begin{aligned} & P(\max_{n_0 \le t \le n_1} Z_G(t) > b) \\ & \approx 4 * P(\max_{\frac{n_0}{n} \le t \le \frac{n_1}{n}} \hat{G_{12}}^*(u) > \sqrt{\tfrac{b}{2}}) * P(\max_{\frac{n_0}{n} \le t \le \frac{n_1}{n}} \hat{G_{21}}^*(u) > \sqrt{\tfrac{b}{2}}) \\ + \quad & \approx 2b\phi^2(\sqrt{\tfrac{b}{2}}) \int_{x_0}^{x_1} g_{12}(x)\upsilon(b_0\sqrt{2g_{12}(x)})dx \int_{x_0}^{x_1} g_{21}(x)\upsilon(b_0\sqrt{2g_{21}(x)})dx \end{aligned}$$

Lemma 4.2.2 *This lemma specially indicates how $g_{12}(x)$ and $g_{21}(x)$ are derived.*

$$g_{12}(x) = \lim_{u \nearrow x} \frac{\partial \rho_{g_{12}}(u,x)}{\partial u}$$
$$g_{21}(x) = \lim_{u \nearrow x} \frac{\partial \rho_{g_{21}}(u,x)}{\partial u}$$

where

$$\rho_{g_{12}}(u,x) = cov(\hat{G_{12}}^*(u), \hat{G_{12}}^*(x))$$
$$\rho_{g_{21}}(u,x) = cov(\hat{G_{21}}^*(u), \hat{G_{21}}^*(x))$$

The lemma above shows that both $g_{12}(x)$ and $g_{21}(x)$ have some special meanings, they both indicate the limitation of the partial derivations of the covariances of the two independent Gaussian processes $\hat{G_{12}}^*(u)$ and $\hat{G_{21}}^*(u)$.

4.3 Skewness Correction

General speaking, convergence of both $\hat{G_{12}}(t)$ and $\hat{G_{21}}(t)$, as well as $Z_G(t)$ to Gaussian distribution is slow if $\frac{t}{n}$ comes closer to 0 or 1. As we can show in the following figures of simulation results, where the samples are sampled from the 10-dimensional standard Gaussian distribution.

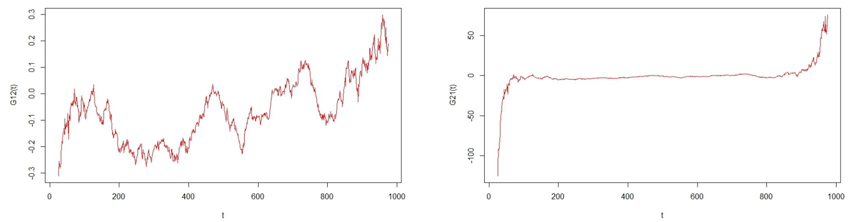

Figure 3: Simulation Results of $\mathbf{E}\hat{G_{12}}(t)^3$(left) and $\mathbf{E}\hat{G_{21}}(t)^3$(right)

Judging from the figures above, we could also conclude that both $\hat{G_{12}}(t)$ and $\hat{G_{21}}(t)$ become very left-skewed when $\frac{t}{n}$ comes close to 0 and very right-skewed when $\frac{t}{n}$ comes closer to 1, making the p-value estimation inaccurate at the two ends, which calls for skewness correction in the tail probability approximation.

We first consider the approximation of the marginal probability $P(\hat{G_{12}}(t) \in b + dx/b)$. Since $\hat{G_{12}}(t)$ is already standardized, namely, $\mathbf{E}(\hat{G_{12}}(t)) = 0$, and $\mathbf{Var}(\hat{G_{12}}(t)) = 1$, we can make full use of the cumulative generating function $\psi(\theta) = log\mathbf{E}_P(e^{\theta G_{12}(t)})$. By change of measure, $d\mathcal{Q}_\theta = e^{\theta G_{12}(t) - \psi(\theta)}dP$, we can approximate $P(\hat{G_{12}}(t) \in b + dx/b)$ by

$$\frac{1}{\sqrt{2\pi(1+\gamma\theta_b)}} exp(-\theta_b b - x\theta_b/b + \theta_b^2(1 + \gamma\theta_b/3)/2)$$

where θ_b is specially chosen to satisfy $\dot{\psi}(\theta_b) = b$. By Taylor expansion, we can approximate θ_b by

$$\theta_b \approx (-1 + \sqrt{1 + 2\gamma b})/\gamma$$

where $\gamma(t) = \mathbf{E}(G_{12}(t)^3)$

After skewness correction, we can approximate $P(\max_{\frac{n_0}{n} \leq t \leq \frac{n_1}{n}} \hat{G_{12}}^*(u) > \sqrt{\frac{b}{2}})$ by adding the skewness term into the integral

Lemma 4.3.1 *Approximation of* $P(\max_{\frac{n_0}{n} \leq t \leq \frac{n_1}{n}} \hat{G_{12}}^*(u) > \sqrt{\frac{b}{2}})$ *with skewness correction*

$$P(\max_{\frac{n_0}{n} \leq t \leq \frac{n_1}{n}} \hat{G_{12}}^*(u) > \sqrt{\frac{b}{2}}) \approx \sqrt{\frac{b}{2}}\phi(\sqrt{\frac{b}{2}}) \int_{\frac{n_0}{n}}^{\frac{n_1}{n}} S_{G_{12}}(nx) g_{12}(x) \upsilon(b_0\sqrt{2g_{12}(x)})dx$$

where

$$S_{G_{12}}(t) = \frac{exp((1/2)(b-\theta_b(t))^2+(1/6)\gamma(t)\theta_b(t)^3)}{\sqrt{1+\gamma(t)\theta_b(t)}}$$

with $\gamma(t) = \boldsymbol{E}(G_{12}(t)^3)$ *and* $\theta_b(t) = (-1+\sqrt{1+2\gamma(t)b})/\gamma(t)$.

Here as we have already derived the skewness correction method for approximating $P(\max_{\frac{n_0}{n}\leq t\leq \frac{n_1}{n}} \hat{G_{12}}^*(u) > \sqrt{\frac{b}{2}})$, skewness correction for $P(\max_{\frac{n_0}{n}\leq t\leq \frac{n_1}{n}} \hat{G_{21}}^*(u) > \sqrt{\frac{b}{2}})$ can also be carried out in the same way by simply replacing the third-moment term $\gamma(t)$ and the other terms likewise.

Lemma 4.3.2 *Approximation of* $P(\max_{\frac{n_0}{n}\leq t\leq \frac{n_1}{n}} \hat{G_{21}}^*(u) > \sqrt{\frac{b}{2}})$ *with skewness correction*

$$P(\max_{\frac{n_0}{n}\leq t\leq \frac{n_1}{n}} \hat{G_{12}}^*(u) > \sqrt{\tfrac{b}{2}}) \approx \sqrt{\tfrac{b}{2}}\phi(\sqrt{\tfrac{b}{2}})\int_{\frac{n_0}{n}}^{\frac{n_1}{n}} S_{G_{21}}(nx)g_{21}(x)\upsilon(b_0\sqrt{2g_{21}(x)})dx$$

where

$$S_{G_{21}}(t) = \frac{exp((1/2)(b-\theta_b(t))^2+(1/6)\gamma(t)\theta_b(t)^3)}{\sqrt{1+\gamma(t)\theta_b(t)}}$$

with $\gamma(t) = \boldsymbol{E}(G_{21}(t)^3)$ *and* $\theta_b(t) = (-1+\sqrt{1+2\gamma(t)b})/\gamma(t)$.

5 Numerical Analysis

5.1 Critical Value

Numerical analysis is conducted to test the validity of the new test statistic. For a given p-value p, the critical value c is chosen to satisfy the condition

$$P(\max_{n_0\leq t\leq n_1} Z_G(t) > c) = p$$

The true critical value could be replaced by the permutation critical value when the simulations times are sufficiently large. If the new test statistic performs well, then the estimated asymptotic critical value would not be much different from the permutation critical value.

The table below shows both the true and the estimated critical value with the given p-value of 0.05. 1000 samples are generated from the ten-dimensional Gaussian distribution $N(0, I_{10})$.

Permutation Critical Value	Asymptotic Critical Value	Asymptotic Critical Value with Skewness Correction	t_1	t_2
14.36	15.05	14.65	26	975
14.81	15.05	15.03	26	975
14.44	15.05	14.80	26	975
13.99	14.45	14.15	51	950
13.47	14.45	13.77	51	950
13.67	14.45	13.91	51	950

From the table above, we could found out that the asymptotic critical value could give a rough estimate of the true critical value and that skewness correction could improve the validity of the original estimation to some extent, though not significantly.

6 Application to Real Life Data

6.1 The Character-Event Matrix for *A Dream in Red Mansions*

6.1.1 Description of the Data

As is known to all, *A Dream in Red Mansions* is among one of the most famous novels in the history of Chinese literature, which is characterized by tortuous plots and vivid character images. Here I checked up the character-event matrix and performed change point detection on it.

The Character-Event Matrix is a matrix with 475 rows and 374 columns. Each row represents an event happened and is arranged according to time; each column represents a character in the novel. Every element in the matrix only takes value 0 or 1. For example, the (i, j)th element in the matrix takes value 1 if the j th character occurs in the i th event and takes value 0 if the j th character did not occur in the i th event.

Then the matrix could be treated as a sequence of 475 observations sampled in a 374-dimensional space with each observation representing an event. If the i th dimension of an observation takes value in 1 then the i th character will occur in the event corresponding to that observation.

6.1.2 Application Result

I conducted change point detection for the matrix through two different graphs constructed upon it, the **1-NNG** and the **3-NNG** and the corresponding test statistic $Z_G(t)$ is also computed and its maximal value is compared with the theoretically approximated p-value.

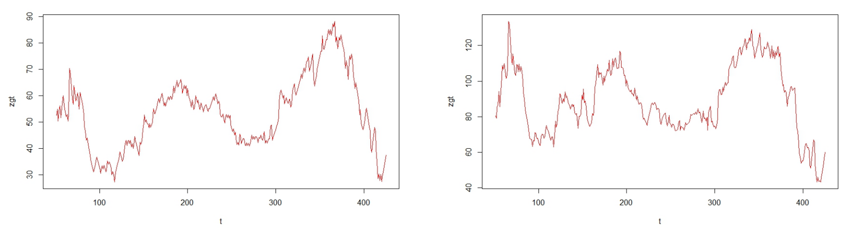

Figure 4: Figure of $Z_G(t)$: Left:**1-NNG**; Right:**3-NNG**

In both cases, the constraints for the starting point n_0 is chosen as 50, while n_1 is chose as 425. The critical value b obtained from p-value approximation such that $P(\max_{\frac{50}{475} \leq u \leq \frac{425}{475}} Z_G^*(t) > b) = 0.01$ is 37.1. However, in the **1-NNG** case, the $\max_{50 \leq t \leq 425} Z_G(t)$ is 88.3 and in the **3-NNG** case, it becomes even greater and reaches 133.7. In both cases, we could reach the conclusion that a significant change point has occurred in the character-event data.

Moreover, from the figures above, in both cases, $Z_G(t)$ has reached to three peaks across the sequence of events. In the **1-NNG** case, $Z_G(t)$ reaches to a peak at the 66th event, the 192th event and the 352th event; while in the **3-NNG** case, it reaches the peak at the 66th event, the 193th event and the 367the event. The three peaks happened at around the 16th chapter, the 50th chapter and the 82th chapter, and reasonable answers could be found out to account for these changes.

It is acknowledged that in the 17th chapter *Literary Talent Is Tested by Composing Inscription in Grand View Garden; Those Losing Their Way at Happy Red Court Explore a Secluded Retreat*, the plot has come to a dramatic turn. In the 50th chapter, *In Reed Snow Cottage Girls Vie in Composing a Collective*

Poem; In Warm Scented Arbour Fine Lantern Riddles Are Made, a lot of characters had gathered together, making the observations corresponding to that chapter significantly deviated from the others. It is also widely assumed that the author for *A Dream in Red Mansions* has switched from *Xueqin Cao* to *E Gao* after the 80th chapter, which could also account for the third peak, namely, the change at around 350 to 360th events.

7 Appendix

7.1 Appendix A Proof for the Convergence of $G_{12}(t)$ and $G_{21}(t)$

7.1.1 $G_{12}(t)$ and $G_{21}(t)$ under the Bootstrap null distribution

First denote $E_B(G_{12}(t))$ and $E_B(G_{21}(t))$ as the expectations and $Var_B(G_{12}(t))$ and $Var_B(G_{21}(t))$ as the variances under the bootstrap null distribution

$$E_B(G_{12}(t)) = E_B(G_{21}(t)) = K * \frac{t*(n-t)}{n}$$

$$\begin{aligned} Var_B(G_{12}(t)) = n * K * \frac{t*(n-t)*(n^2-n*t+t^2)}{n^4} \\ -(2*n*K^2 - \textstyle\sum_{i,j=1}^n a_{ij}^+ a_{ji}^+) * \frac{t^2*(n-t)^2}{n^4} \\ +n*K*(K-1)*\frac{t^3*(n-t)}{n^4} \\ +(\textstyle\sum_{i,j,l=1}^n a_{ij}^+ a_{lj}^+ - K*n) * \frac{t*(n-t)^3}{n^4} \end{aligned}$$

$$\begin{aligned} Var_B(G_{21}(t)) = n * K * \frac{t*(n-t)*(n^2-n*t+t^2)}{n^4} \\ -(2*n*K^2 - \textstyle\sum_{i,j=1}^n a_{ij}^+ a_{ji}^+) * \frac{t^2*(n-t)^2}{n^4} \\ +n*K*(K-1)*\frac{t*(n-t)^3}{n^4} \\ +(\textstyle\sum_{i,j,l=1}^n a_{ij}^+ a_{lj}^+ - K*n) * \frac{t^3*(n-t)}{n^4} \end{aligned}$$

7.1.2 The Relation between the Permutation Null Distribution and the Bootstrap Null Distribution

Let

$$W_1 = \frac{G_{12}(t)-E(G_{12}(t))}{\sqrt{Var(G_{12}(t))}} \qquad W_2 = \frac{G_{21}(t)-E(G_{21}(t))}{\sqrt{Var(G_{21}(t))}}$$
$$W_{1B} = \frac{G_{12}(t)-E_B(G_{12}(t))}{\sqrt{Var_B(G_{12}(t))}} \qquad W_{2B} = \frac{G_{21}(t)-E_B(G_{21}(t))}{\sqrt{Var_B(G_{21}(t))}}$$

Under some algebraic computation, we have

$$W_1 = \frac{\sqrt{Var_B(G_{12}(t))}}{\sqrt{Var(G_{12}(t))}} W_{1B} - \frac{K * \frac{t*(n-t)}{n*(n-1)}}{\sqrt{Var(G_{12}(t))}}$$
$$W_2 = \frac{\sqrt{Var_B(G_{21}(t))}}{\sqrt{Var(G_{21}(t))}} W_{2B} - \frac{K * \frac{t*(n-t)}{n*(n-1)}}{\sqrt{Var(G_{21}(t))}}$$

7.1.3 The Limiting distribution of (W_{1B}, W_{2B})

In this section, we will show that, under some restrictive conditions, (W_{1B}, W_{2B}) will converge to a two-dimensional Gaussian distribution as n grow larger.
First denote two functions about an edge e

$$I_e = I(\text{e connects a node sampled before t to a node sampled after t})$$
$$\tilde{I}_e = I(\text{e connects a node sampled after t to a node sampled before t})$$

In order to prove that (W_{1B}, W_{2B}) converge to a two-dimensional Gaussian distribution, we only need to show that $W_B = a * W_{1B} + b * W_{2B}$ converges to a Gaussian distribution for any pair of (a, b) such that the variance of W_B is larger than 0.

Given the definition of I_e and $\tilde{I}_e$, we can write W_B in another representation

$$W_B = \sum_{e \in G} \left(a * \frac{I_e - \frac{t*(n-t)}{n^2}}{\sqrt{Var_B(G_{12}(t))}} + b * \frac{\tilde{I}_e - \frac{t*(n-t)}{n^2}}{\sqrt{Var_B(G_{21}(t))}} \right)$$

Also let

$$\zeta_e = a * \frac{I_e - \frac{t*(n-t)}{n^2}}{\sqrt{Var_B(G_{12}(t))}} + b * \frac{\tilde{I}_e - \frac{t*(n-t)}{n^2}}{\sqrt{Var_B(G_{21}(t))}}$$

For any edge e in the graph, let

$$K_e = \left\{ e^{'} : e^{'} shares \quad a \quad node \quad with \quad e \right\}$$
$$N_e = \{nodes \quad in \quad K_e\}$$
$$L_e = \left\{ e^{"} : e^{"} has \quad a \quad node \quad in \quad N_e \right\}$$

Assumption 7.1.1 *For each $i \in \mathcal{J}$, there exits $\mathcal{S}_i \subset \mathcal{T}_i \subset \mathcal{J}$ such that ξ_i is independent of $\xi_{S_i^c}$ and ξ_{S_i} is independent of $\xi_{T_i^c}$*

Theorem 7.1.1 *Under* ***Assumption 7.1.1****, we have*

$$\sup_{h \in Lip(i)} |\boldsymbol{E}h(W) - \boldsymbol{E}h(Z)| < \delta$$

where $Lip(1) = \{h : \mathbb{R} \to \mathbb{R}\}$*,* $Z \sim N(0,1)$*, and*

$$\delta = 2 \sum_{i \in \mathcal{J}} (\boldsymbol{E}|\xi_i \eta_i \theta_i| + |\boldsymbol{E}(\xi_i \eta_i)| \boldsymbol{E}|\theta_i|) + \sum_{i \in \mathcal{J}} \boldsymbol{E}|\xi_i \eta_i^2|$$

with $\eta_i = \sum_{j \in \mathcal{S}_i} \xi_j$ *and* $\theta_i = \sum_{j \in \mathcal{T}_i} \xi_j$*, where* S_i *and* T_i *are as defined in* ***Assumption 6.1.1***

It is clear that e is independent of K_e^c and K_e is also independent of L_e^c and these two types of sets satisfy **Assumption 6.1.1**. Let $\eta_e = \sum_{e' \in K_e} \zeta_{e'}$ and $\theta_e = \sum_{e" \in L_e} \zeta_{e"}$

According to the **Theorem 6.1.1**,

$$\sup_{h \in Lip(1)} \quad E|h(\frac{W_B}{\sqrt{Var(W_B)}}) - h(Z)| < \delta \qquad Z \sim N(0,1)$$

where

$$\delta = \frac{1}{\sqrt{(Var(W_b))^3}} * (2 \sum_{e \in G} (E_B|\zeta_e \eta_e \theta_e| + |E_B(\zeta_e \eta_e)|(E_B|\theta_e|)) + \sum_{e \in G} E_B|\zeta_e \eta_e|^2)$$

Also let $m = max(|a|, |b|)$ and $\sigma = \sqrt{min(Var_B(G_{12}(t)), Var_B(G_{21}(t)))}$, then for any edge $e \in G$,

$$\begin{aligned} |\zeta_e| &\leq |a * \frac{I_e - \frac{t*(n-t)}{n^2}}{\sqrt{Var_B(G_{12}(t))}}| + |b * \frac{\tilde{I}_e - \frac{t*(n-t)}{n^2}}{\sqrt{Var_B(G_{21}(t))}}| \\ &\leq \frac{|a|}{\sqrt{Var_B(G_{12}(t))}} + \frac{|b|}{\sqrt{Var_B(G_{21}(t))}} \\ &\leq \frac{2m}{\sigma} \end{aligned}$$

Consequently $|\eta_e| \leq |K_e| * \frac{2m}{\sigma}$ and $|\theta_e| \leq |L_e| * \frac{2m}{\sigma}$ and

$$\begin{aligned} &\delta \leq \\ &\frac{1}{\sqrt{(Var(W_b))^3}} \left[2 \sum_{e \in G} (E_B|\zeta_e||\eta_e||\theta_e| + (E_B|\zeta_e||\eta_e|)(E_B|\theta_e|)) + \sum_{e \in G} E_B(|\zeta_e||\eta_e|^2) \right] \\ &\leq \frac{1}{\sqrt{(Var(W_b))^3}} \left[2 \sum_{e \in G} \left((\frac{2m}{\sigma})^3 |K_e||L_e| + (\frac{2m}{\sigma})^2 |K_e| * (\frac{2m}{\sigma})|L_e| \right) \right] \\ &\leq \frac{1}{\sqrt{(Var(W_b))^3}} * \frac{8m^3}{\sigma^3} \left(6 \sum_{e \in G} |K_e||L_e| + \sum_{e \in G} |K_e|^2 \right) \\ &\leq \frac{56m^3}{\sqrt{(Var(W_b))^3}} * \frac{\sum_{e \in G} |K_e||L_e|}{\sigma^3} \end{aligned}$$

Since $Var(W_B) = a^2 + b^2 + 2ab * corr_B(G_{12}(t), G_{21}(t))$ is of constant order, the order of δ only depends on the orders of $\sum_{e \in G} |K_e||L_e|$ and σ. For simplicity, denote $p_n = \frac{t}{n}$ and $q_n = \frac{n-t}{n}$ then

$$
\begin{aligned}
& Var_B(G_{12}(t)) \\
& = nKp_nq_n(1 - p_nq_n) - (2nK^2 - \textstyle\sum_{i,j=1}^{n} a_{ij}^{+} a_{ji}^{+})p_n^2 q_n^2 + nK(K-1)p_n^3 q_n + \\
& (\textstyle\sum_{i,j,l=1}^{n} a_{ij}^{+} a_{lj}^{+} - nK)p_n q_n^3 \\
& \geq nKp_nq_n(1 - p_nq_n) - 2nK^2 p_n^2 q_n^2 + nK^3 p_n^3 q_n - nKp_n^3 q_n + (nK^2 - nK)p_n q_n^3 \\
& (\text{since } \textstyle\sum_{i,j,l=1}^{n} a_{ij}^{+} a_{lj}^{+} \geq nK^2) \\
& \geq n(Kp_nq_n)(1 + Kp_n + Kq_n - p_nq_n - Kp_nq_n - p_n^2 - q_n^2)
\end{aligned}
$$

Because for any pair of p_n and q_n, $(1 + Kp_n + Kq_n - p_nq_n - Kp_nq_n - p_n^2 - q_n^2)$ is always larger than 0, so $Var_B(G_{21}(t))$ is at least of order $O(n)$, and it is also the same with $Var_B(G_{21}(t))$, then $\sigma = \sqrt{min(Var_B(G_{12}(t)), Var_B(G_{21}(t)))} \geq O(n^{0.5})$. Consequently, as long as the order of $\sum_{e \in G} |K_e||L_e|$ is smaller than $o(n^{1.5})$, δ will surely converge to 0 and (W_{1B}, W_{2B}) will converge to a two-dimensional Gaussian distribution.

7.1.4 The distribution of (W_1, W_2)

First recall that

$$
\begin{aligned}
W_1 &= \frac{\sqrt{Var_B(G_{12}(t))}}{\sqrt{Var(G_{12}(t))}} W_{1B} - \frac{K * \frac{t*(n-t)}{n*(n-1)}}{\sqrt{Var(G_{12}(t))}} \\
W_2 &= \frac{\sqrt{Var_B(G_{21}(t))}}{\sqrt{Var(G_{21}(t))}} W_{2B} - \frac{K * \frac{t*(n-t)}{n*(n-1)}}{\sqrt{Var(G_{21}(t))}}
\end{aligned}
$$

Since $Var_B(G_{12}(t))$ and $Var(G_{12}(t))$ are of the same order, both $\frac{\sqrt{Var_B(G_{12}(t))}}{\sqrt{Var(G_{12}(t))}}$ and $\frac{\sqrt{Var_B(G_{21}(t))}}{\sqrt{Var(G_{21}(t))}}$ will converge to a constant number. Moreover, since $\frac{K * \frac{t*(n-t)}{n*(n-1)}}{\sqrt{Var(G_{12}(t))}}$ and $\frac{K * \frac{t*(n-t)}{n*(n-1)}}{\sqrt{Var(G_{21}(t))}}$ will converge to 0, W_1 and W_2 could be treated as a linear transformation of W_{1B} and W_{2B}, from which we could reach the conclusion that (W_1, W_2) will also converge to a two-dimensional Gaussian distribution, and now the proof is fully done.

7.2 Appendix B Derivation for p-value approximation

Here approximate $P(\max_{n_0 \le t \le n_1} \hat{G_{12}}(t) > b)$ by the Gaussian process that it converges to, $P(\max_{n_0 \le t \le n_1} \hat{G_{12}}^*(\frac{t}{n}) > b)$

$$
\begin{aligned}
& P(\max_{n_0 \le t \le n_1} \hat{G_{12}}^*(\tfrac{t}{n}) > b) \\
= & \sum_{n_0 \le t \le n_1} \int_0^\infty \frac{1}{\sqrt{2\pi}} e^{-\frac{(x+b)^2}{2}} P(\max_{n_0 \le s} \hat{G_{12}}^*(\tfrac{s}{n}) < b | \hat{G_{12}}^*(\tfrac{t}{n}) = x+b) dx \\
= & \frac{\phi(b)}{b} \sum_{n_0 \le t \le n_1} \int_0^\infty e^{-x} e^{-\frac{x^2}{2b^2}} P(b(\hat{G_{12}}^*(\tfrac{s}{n}) - \hat{G_{12}}^*(\tfrac{t}{n})) < -x | \hat{G_{12}}^*(\tfrac{t}{n}) = b + \tfrac{x}{b}) dx
\end{aligned}
$$

If $x \sim o(b^2), \frac{x^2}{b^2}$ is negligible to x and $\frac{x}{b}$ is negligible to b, then

$$
\begin{aligned}
& P(\max_{n_0 \le t \le n_1} \hat{G_{12}}^*(\tfrac{t}{n}) > b) \\
= & \frac{\phi(b)}{b} \sum_{n_0 \le t \le n_1} \int_0^\infty e^{-x} P(b(\hat{G_{12}}^*(\tfrac{s}{n}) - \hat{G_{12}}^*(\tfrac{t}{n})) < -x | \hat{G_{12}}^*(\tfrac{t}{n}) = b) dx
\end{aligned}
$$

First consider the conditional distribution of $b(\hat{G_{12}}^*(u) - \hat{G_{12}}^*(v)) | \hat{G_{12}}^*(v) = b$

$b(\hat{G_{12}}^*(u) - \hat{G_{12}}^*(v)) | \hat{G_{12}}^*(v) = b \sim N((\rho_G^*(u,v) - 1)b^2, (1 - \rho_G^*(u,v)^2)b^2)$

and by second order Taylor Expansion, we have

$$
\begin{aligned}
\rho_G^*(u,v) &= 1 + f'_{v,-}(0)(u-v) + \tfrac{1}{2} f''_{v,-}(0)(u-v)^2 \\
\rho_G^*(u,v)^2 &= 1 + 2f'_{v,-}(0)(u-v) + [f'_{v,-}(0)^2 + f''_{v,-}(0)](u-v)^2
\end{aligned}
$$

After some algebraic calculation, it can be easily shown that $f'_{v,-}(0) = \frac{1}{2v(1-v)}$. Neglecting the second order terms, we can approximate the distribution as $b(\hat{G_{12}}^*(u) - \hat{G_{12}}^*(v)) | \hat{G_{12}}^*(v) = b \sim N(-f'_{v,-}(0)|v-u|b^2, 2f'_{v,-}(0)|v-u|b^2)$ by the Taylor expansion of $\rho_G^*(u,v)$.

Let $W_m^{(t)}$ be a random work with $W_1^{(t)} \sim N(\frac{1}{n} f'_{v,-}(0)b^2, 2\frac{1}{n} f'_{v,-}(0)b^2)$ then we can further the approximation as

$$
\begin{aligned}
& P(\max_{n_0 \le s < t} b(\hat{G_{12}}^*(u) - \hat{G_{12}}^*(v)) < -x | \hat{G_{12}}^*(v) = b) \\
& \sim P(\max_{n_0 \le s < t} -W_{t-s}^{(t)} < -x) \sim P(\min_{t \ge 1} W_m^{(t)} > x)
\end{aligned}
$$

First recall the definition for $g_{12}(x)$

$$g_{12}(x) = \lim_{u \nearrow x} \frac{\partial \rho_{g_{12}}(u,x)}{\partial u} = f'_{x,-}(0) = \frac{1}{2x(1-x)}$$

Also let $x_0 = \lim_{n\to\infty} \frac{n_0}{n}$, $x_1 = \lim_{n\to\infty} \frac{n_1}{n}$

Theorem 6.2.1 *If a random walk W_m satisfies $W_1 \sim N(\mu, \sigma)$,then*

$$\int_0^\infty e^{\frac{-2\mu x}{\sigma}} P(\min_{m\geq 1} W_m > x)dx = \mu \mathrm{v}(\tfrac{2\mu}{\sigma})$$

From **Theorem 7.2.1**,we could finally derive the approximation:

$$\begin{aligned}
&P(\max_{n_0 \leq t \leq n_1} \hat{G_{12}}^*(\tfrac{t}{n}) > b) \\
&\sim \lim_{n\to\infty} \frac{\phi(b)}{b} \sum\nolimits_{n_0\leq t\leq n_1} \frac{b^2}{n} g_{12}(\tfrac{t}{n}) v(\tfrac{b}{\sqrt{n}} \sqrt{2g_{12}(\tfrac{t}{n})})) \\
&\sim b\phi(b) \int_{x_0}^{x_1} g_{12}(x) v(\tfrac{b}{\sqrt{n}} \sqrt{2g_{12}(x)}) dx
\end{aligned}$$

7.3 Appendix C Analytic Representation for the third moments

$$\mathrm{E}[(\sum_{i=1}^{t}\sum_{j=t+1}^{n} A_{ij}^{+})(\sum_{i=t+1}^{n}\sum_{j=1}^{t} A_{ij}^{+})^2]$$

$$
\begin{aligned}
= \quad & (nt-t) && \cdot \mathrm{E}\, A_{ij}^{+2} A_{ji}^{+} \\
& + t(n-t)(n-t-1) && \cdot \mathrm{E}\, A_{ij}^{+2} A_{ki}^{+} \\
& + t(t-1)(n-t) && \cdot \mathrm{E}\, A_{ij}^{+2} A_{jk}^{+} \\
& + t(t-1)(n-t)(n-t-1) && \cdot \mathrm{E}\, A_{ij}^{+2} A_{ke}^{+} \\
& + 2t(n-t)(n-t-1) && \cdot \mathrm{E}\, A_{ij}^{+} A_{ik}^{+} A_{ji}^{+} \\
& + 2t(n-t)(n-t-1)(t-1) && \cdot \mathrm{E}\, A_{ij}^{+} A_{ik}^{+} A_{je}^{+} \\
& + t(n-t)(n-t-1)(n-t-2) && \cdot \mathrm{E}\, A_{ij}^{+} A_{ik}^{+} A_{li}^{+} \\
& + t(n-t)(n-t-1)(n-t-2)(t-1) && \cdot \mathrm{E}\, A_{ij}^{+} A_{ik}^{+} A_{lm}^{+} \\
& + 2t(n-t)(t-1) && \cdot \mathrm{E}\, A_{ij}^{+} A_{kj}^{+} A_{ji}^{+} \\
& + t(t-1)(n-t)(t-2) && \cdot \mathrm{E}\, A_{ij}^{+} A_{kj}^{+} A_{je}^{+} \\
& + 2t(t-1)(n-t)(n-t-1) && \cdot \mathrm{E}\, A_{ij}^{+} A_{kj}^{+} A_{li}^{+} \\
& + t(t-1)(n-t)(n-t-1)(t-2) && \cdot \mathrm{E}\, A_{ij}^{+} A_{kj}^{+} A_{lm}^{+} \\
& + 2t(t-1)(n-t)(n-t-1) && \cdot \mathrm{E}\, A_{ij}^{+} A_{ke}^{+} A_{ji}^{+} \\
& + 2t(t-1)(n-t)(n-t-1) && \cdot \mathrm{E}\, A_{ij}^{+} A_{ke}^{+} A_{jk}^{+} \\
& + 2t(t-1)(n-t)(n-t-1)(t-2) && \cdot \mathrm{E}\, A_{ij}^{+} A_{ke}^{+} A_{jm}^{+} \\
& + 2t(t-1)(n-t)(n-t-1)(n-t-2) && \cdot \mathrm{E}\, A_{ij}^{+} A_{ke}^{+} A_{mi}^{+} \\
& + t(t-1)(n-t)(n-t-1)(n-t-2)(t-2) && \cdot \mathrm{E}\, A_{ij}^{+} A_{ke}^{+} A_{mn}^{+}
\end{aligned}
$$

$$
\begin{aligned}
& \mathrm{E}[(\sum_{i=1}^{t}\sum_{j=t+1}^{n} A_{ij}^{+})^3] \\
= \quad & t(n-t) && \cdot \mathrm{E}\, A_{ij}^{+3} \\
& + 3t(n-t)(n-t-1) && \cdot \mathrm{E}\, A_{ij}^{+2} A_{ik}^{+} \\
& + 3t(n-t)(t-1) && \cdot \mathrm{E}\, A_{ij}^{+2} A_{kj}^{+} \\
& + 3t(t-1)(n-t)(n-t-1) && \cdot \mathrm{E}\, A_{ij}^{+2} A_{ke}^{+} \\
& + t(n-t)(n-t-1)(n-t-2) && \cdot \mathrm{E}\, A_{ij}^{+} A_{il}^{+} A_{ik}^{+} \\
& + 3t(t-1)(n-t)(n-t-1)(n-t-2) && \cdot \mathrm{E}\, A_{ij}^{+} A_{ik}^{+} A_{ne}^{+} \\
& + 6t(t-1)(n-t)(n-t-1) && \cdot \mathrm{E}\, A_{ij}^{+} A_{ik}^{+} A_{nk}^{+} \\
& + t(t-1)(t-2)(n-t) && \cdot \mathrm{E}\, A_{il}^{+} A_{je}^{+} A_{ke}^{+} \\
& + 3t(t-1)(t-2)(n-t)(n-t-1) && \cdot \mathrm{E}\, A_{ie}^{+} A_{je}^{+} A_{km}^{+} \\
& + t(t-1)(t-2)(n-t)(n-t-1)(n-t-2) && \cdot \mathrm{E}\, A_{ij}^{+} A_{ke}^{+} A_{mn}^{+}
\end{aligned}
$$

References

[1] A new graph-based two-sample test for multivariate and object data, *JASA*, H.Chen, J.H.Friedman, 2016

[2] Detecting simultaneous change-points in multiple sequences, *Biometrika*, N.Zhang, D.Siegmund, 2010

[3] Change-point problems, *IMS*, E.G.Carlstein, H.G.Muller, D.Siegmund, 1994

[4] Tail approximations for maixma of random fields, *de Gruyter*, D.Siegmund, 1992

[5] The maximum of a function of a Markov Chain and application to linkage analysis, *Adv. in Appl.Probab*, I.P.Tu, D.Siegmund, 1999

[6] Detection of the disorder in multi-dimensional random-processes, *Doklady Akademii Nauk SSSR*, L.J.Vostrikova, 1981